Zachodniopomorski Uniwersytet Technologiczny
w Szczecinie

Marcin Pluciński

Zastosowanie lokalnych modeli regresyjnych (mini-modeli) w przetwarzaniu informacji niepewnych

SZCZECIN 2019

Recenzenci
Zbigniew Pietrzykowski
Zenon Zwierzewicz

Opracowanie redakcyjne
Anna Ciciak

Projekt okładki
Marcin Pluciński

Wydano za zgodą
Rektora Zachodniopomorskiego Uniwersytetu Technologicznego
w Szczecinie

ISBN 978-83-7663-287-2

Wydawnictwo Uczelniane Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie
70-311 Szczecin, al. Piastów 50, tel. 91 449 47 60, e-mail: wydawnictwo@zut.edu.pl
Druk PPH-ZAPOL

Spis treści

1. Wstęp 5

2. Mini-modele 9

2.1. Mini-modele: liniowe i nieliniowe modele lokalne 9

2.2. Mini-modele z bazą hiper-elipsoidalną 13

2.3. Eksperymenty 15

3. Interwały 21

3.1. Pojęcia podstawowe 21

3.2. Wielowymiarowa arytmetyka interwałowa RDM 22

3.3. Odległość pomiędzy interwałami 29

3.4. Interwałowa wersja metody najmniejszych kwadratów 30

3.4.1. Model z jednym wejściem 31

3.4.2. Model z wieloma wejściami 35

3.5. Modelowanie lokalne dla danych interwałowych 38

3.5.1. Dokładność modelu 39

3.6. Eksperymenty 41

3.7. Wnioski 52

4. Liczby rozmyte 53

4.1. Pojęcia podstawowe 53

4.2. Wielowymiarowa arytmetyka liczb rozmytych oparta na notacji RDM 55

4.2.1. Arytmetyka liczb rozmytych w postaci RDM 58

4.2.2. Własności arytmetyki liczb rozmytych w postaci RDM 61

4.3. Odległość pomiędzy liczbami rozmytymi 63

4.4. Rozmyta wersja metody najmniejszych kwadratów 65

4.4.1. Model z jednym wejściem 66

4.4.2. Model z wieloma wejściami 71

4.5. Modelowanie lokalne dla danych rozmytych 75

4.5.1. Dokładność modelu 76

4.6. Eksperymenty 76

4.7. Wnioski 92

5. Liczby Z 97
5.1. Pojęcia podstawowe 97
5.2. Arytmetyka liczb Z i Z^+ 99
5.3. Odległość pomiędzy liczbami Z 105
5.4. Metoda najmniejszych kwadratów dla liczb Z 106
5.4.1. Model z jednym wejściem 106
5.4.2. Model z wieloma wejściami 114
5.5. Modelowanie lokalne dla danych w postaci liczb Z 121
5.5.1. Dokładność modelu 122
5.6. Eksperymenty 123
5.7. Wnioski 146

6. Zakończenie 151

Literatura 153

Skorowidz 163

Summary 165

1. Wstęp

Książka poświęcona jest jednemu z głównych rodzajów indukcyjnego uczenia się, jakim jest aproksymacja funkcji. Inaczej mówiąc, będziemy poszukiwali modeli realizujących odwzorowanie przykładów z pewnej dziedziny na przeciwdziedzinę. Aproksymata tworzona będzie na podstawie danych (próbek), z których każda opisywana będzie zestawem cech (atrybutów, wejść) oraz informacją o zadanej wartości odpowiedzi. Proces uczenia się, z jakim będziemy mieli do czynienia, będzie zatem procesem nadzorowanym.

Uczące się aproksymatory funkcji możemy różnicować według różnych kryteriów, takich jak: dopuszczalne tryby uczenia się, postać informacji trenującej, czy ich wewnętrzna struktura. Struktura, czyli sposób reprezentacji aproksymowanej funkcji, określa sposób jej modyfikowania (uczenia) na podstawie dostępnych przykładów [13]. Biorąc pod uwagę kryterium reprezentacji, modele możemy podzielić na dwie podstawowe kategorie.

1. Aproksymatory parametryczne, dla których odpowiedź wyznaczana jest na podstawie zestawu parametrów modyfikowanych w procesie uczenia, według pewnej ustalonej z góry formuły. Aproksymatory tego typu tworzą zwykle jeden model dla całej dziedziny danych, stąd nazwiemy go modelem globalnym.
2. Aproksymatory pamięciowe, przechowujące dostarczone im przykłady trenujące i wyznaczające odpowiedź dla podanego zapytania na podstawie znanych wartości funkcji docelowej dla najbardziej do niego podobnych zapamiętanych próbek. Odpowiedź aproksymatora będzie tu zatem wyznaczana jako odpowiedź modelu lokalnego, tworzonego dla sąsiedztwa punktu wejściowego.

Uczenie aproksymatorów funkcji z wykorzystaniem tzw. metod pamięciowych może być często atrakcyjnym podejściem w porównaniu z technikami tworzącymi modele globalne, bazujące na regresji parametrycznej. W pewnych przypadkach (np. dla niewielkiej liczby danych uczących) tworzenie modeli globalnych może być trudne lub nawet niemożliwe i wtedy metody pamięciowe stają się jedynymi z dostępnych metod realizujących zadanie aproksymacji.

Metody pamięciowe są dobrze zbadane i opisane w literaturze przedmiotu. Najważniejszą z nich jest na pewno metoda k najbliższych sąsiadów (ang. *k nearest neighbours method – kNN method*) opisana dobrze w pracach [13, 44, 72]. Mimo swojej popularności, metoda ta jest cały czas przedmiotem badań [58, 60] i udoskonaleń. Stosowana jest w pro-

stej, podstawowej formie (opisanej dokładniej w rozdziale 2), jak i np. z uwzględnieniem wag przypisanych do próbek [7, 13]. Do metod pamięciowych zaliczyć możemy także regresję lokalną [44], probabilistyczne sieci neuronowe (ang. *probabilistic neural networks*), czy uogólnione sieci regresyjne (ang. *generalised regression networks*) [103, 136].

Metody bazujące na lokalnej regresji (czasem nazywanej też regresją ruchomą), zamiast jednego, wspólnego dla całej dziedziny modelu globalnego, tworzą wiele modeli lokalnych, których parametry zmieniają się w zależności od położenia punktu wejściowego w dziedzinie danych. Modele lokalne zwykle są proste (często jest to regresja liniowa), dzięki czemu udaje się powiązać prostotę tworzenia modelu z dobrym dopasowaniem do danych, jakie z kolei daje regresja nieliniowa.

Koncepcja regresji lokalnej przedstawiona została w latach siedemdziesiątych i osiemdziesiątych XX wieku, między innymi przez W. Clevelanda [14, 15, 16]. Autor nazwał swoją metodę lokalnie ważoną regresją wielomianową (ang. *locally weighted polynomial regression*). Podobne rozwiązania można znaleźć przykładowo w publikacjach J. Friedmana i innych [30, 32, 49]. W cytowanych pracach odpowiedź modelu wyznaczana była na podstawie wielomianu niskiego rzędu dopasowanego do próbek uczących znajdujących się w sąsiedztwie punktu zapytania. W metodach opisanych przez Clevelanda, współczynniki wielomianu wyznaczane były za pomocą ważonej metody najmniejszych kwadratów, przypisującej większą wagę punktom znajdującym się bliżej wejścia modelu.

Największą zaletą modelowania opartego na regresji lokalnej jest brak konieczności budowania jednego globalnego modelu dopasowanego do wszystkich danych. Otrzymujemy tu elastyczne narzędzie, które może być wykorzystane do modelowania złożonych procesów, dla których nie istnieją (bądź nie są zupełnie znane) modele teoretyczne. W połączeniu z prostotą metody otrzymujemy więc bardzo atrakcyjne narzędzie do modelowania złożonych problemów. Z kolei wadą modelowania tego typu jest brak modelu reprezentowanego przez czytelną matematyczną formułę. Aby opisać model, musimy znać dane uczące i sposób tworzenia regresji lokalnej dla podanego punktu zapytania. Metody pamięciowe mogą być złożone obliczeniowo w przypadku, kiedy ilość danych uczących jest bardzo duża [13, 44]. Proste, zachłanne poszukiwanie najbliższych sąsiadów punktu wejściowego może być wtedy bardzo czasochłonne i wymagające bardziej zaawansowanych technik, jak np. k-d drzewa [12, 37].

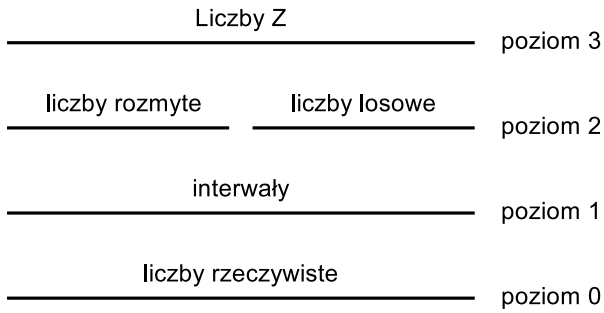
Nieco inne podejście do modelowania lokalnego zastosowano w metodzie znanej jako regresja odcinkowa (ang. *segmented regression, piecewise regression*). Dziedzina zmiennej niezależnej (w przypadku danych z jedną zmienną wejściową) jest tu dzielona na interwały, po czym dla każdego z nich wyznaczana jest lokalna regresja liniowa. Metoda może być zastosowana także dla danych z wieloma zmiennymi wejściowymi, przy czym przestrzeń wejść jest wtedy dzielona w bardziej skomplikowany sposób (np. na simpleksy lub hiper-prostokąty). Regresja odcinkowa jest szczególnie użyteczna, kiedy związek mię-

dzy wejściem a wyjściem opisany jest zależnościami, których postać zależy od położenia w dziedzinie danych [34, 63, 75, 76, 80].

Z jeszcze innym podejściem do regresji lokalnej mamy do czynienia w przypadku regresji jądrowej (ang. *kernel regression*). Jest to technika znana ze statystyki, za pomocą której można szacować oczekiwaną wartość zmiennej losowej. Aby tego dokonać, poszukuje się nieliniowej zależności pomiędzy losową zmienną niezależną i zależną. Najbardziej znanym przykładem regresji jądrowej jest technika opisana przez parę autorów – E. Nadaraya i S. Watsona (ang. *Nadaraya-Watson kernel regression*) [10, 77, 137]. Zaproponowali oni, aby poszukiwaną zależność wyznaczać jako lokalną średnią ważoną danych sąsiednich, wykorzystując jądro jako funkcję wagi. Inne przykłady zastosowań regresji jądrowej opisane są w pracach [51, 57, 124].

Dane wykorzystywane do tworzenia modeli nie zawsze muszą być pewne. Dostępność dokładnych, pewnych informacji jest sytuacją komfortową, jednak często mogą być one podane w sposób nieprecyzyjny, niekompletny lub zniekształcony [145]. Człowiek posiada zadziwiającą umiejętność wykorzystywania tego typu informacji do skutecznego podejmowania decyzji, stąd naturalne dążenie do tworzenia metod, które także będą umiały je wykorzystać.

Niepewność informacji posiada określoną hierarchię przedstawioną na rys. 1.1 [1].



Rysunek 1.1. Hierarchia liczb niepewnych [1]

Im wyższy poziom niepewności, tym mniejsza jest precyzja informacji opisywanej przez liczby określonego typu. Dla każdego z nich opracowana jest specjalna arytmetyka, umożliwiająca przetwarzanie danych, przy czym warte zauważenia jest, że arytmetyka liczb każdego poziomu wykorzystuje arytmetyki liczb poziomów niższych. Przykładowo: arytmetyka liczb rozmytych wykorzystuje arytmetykę interwałową, a arytmetyka liczb Z wykorzystuje arytmetykę liczb rozmytych i losowych.

Oprócz typów niepewności wymienionych na rys. 1.1 istnieją także inne, np.: liczby rozmyte typu II [21, 43, 69], liczby szare (ang. *grey numbers*) [64, 68, 138], liczby w formie zbioru możliwych wartości (ang. *set-valued numbers*) [1, 67]. Liczby tego typu nie będą omawiane i wykorzystywane w dalszej części książki.

Ludzie przetwarzając informacje, czy podejmując decyzje, często robią to z wykorzystaniem tzw. granul informacyjnych (ang. *information granules*). Posiadamy umiejęt-

ność ich tworzenia, przetwarzania, czy wykorzystywania w procesie wnioskowania. Przetwarzaniem danych tego typu zajmuje się teoria obliczeń granularnych (ang. *granular computing*) [31, 84] i jest ona uogólnieniem wielu istniejących, zbadanych i opisanych teorii takich jak: klasyczna teoria zbiorów, teoria zbiorów rozmytych, teoria zbiorów przybliżonych, czy arytmetyka interwałowa [25, 74, 82, 86, 142]. Niezwykle istotne jest, aby modele matematyczne przetwarzające granule informacyjne, posiadały zdolność radzenia sobie z danymi różnego typu (często mieszanego) [48, 67]. Przykładowo, model utworzony na podstawie danych dokładnych powinien być w stanie przetwarzać granule informacyjne. Dodatkowo, powinna także istnieć możliwość tworzenia modeli na podstawie danych niepewnych [33, 35, 42, 47, 144].

Monografia poświęcona jest przede wszystkim modelom bazującym na mini-modelach, które należą do metod pamięciowych i są przykładem zastosowania regresji lokalnej. Koncepcja mini-modeli po raz pierwszy została zaproponowana przez prof. A. Piegata i opisana w pracach [96, 97]. W kolejnych rozdziałach książki podjęto próbę zaadaptowania metod – które oryginalnie opracowane zostały dla danych rzeczywistych – do danych niepewnych (interwałów, liczb rozmytych i liczb Z). Opracowane zostały metody tworzenia modeli lokalnych na podstawie nieprecyzyjnej informacji, a także sposoby wyznaczania odpowiedzi takich modeli.

Rozprawa ma następujący układ. W rozdziale drugim zaprezentowana została koncepcja mini-modeli liniowych i nieliniowych wykorzystujących bazę hiper-sferyczną i hiper-elipsoidalną. Rozdział kończą wyniki eksperymentów, w których zbadano jakość działania i dokładność modeli.

Rozdział trzeci poświęcony jest pierwszemu analizowanemu typowi niepewności, czyli interwałom. W tej części pracy omówiono pojęcia podstawowe i arytmetykę interwałów, a także zaprezentowano interwałową wersję metody najmniejszych kwadratów, która wykorzystywana jest przy tworzeniu mini-modeli. Rozdział ten, podobnie jak poprzednie, kończy prezentacja wyników badań eksperymentalnych.

W rozdziale czwartym przedstawiono najważniejsze informacje o liczbach rozmytych i ich arytmetyce. Najważniejszą jego częścią jest omówienie sposobu modelowania lokalnego dla danych rozmytych, bazującego na rozmytej wersji metody najmniejszych kwadratów.

W przedostatnim, piątym rozdziale scharakteryzowano liczby Z i Z^+ oraz ich arytmetykę. Również i tu opisana została metoda najmniejszych kwadratów dla danych w postaci liczb Z oraz wykorzystujące ją metody modelowania lokalnego. Zarówno czwarty, jak i piąty rozdział, kończą wyniki badań, których głównym zadaniem było zademonstrowanie wiarygodności i dobrej jakości uzyskiwanych wyników.

Ostatnią częścią rozprawy jest zakończenie, w którym posumowano korzyści wynikające z zaproponowanych rozwiązań i omówiono zakres dalszych prac.

2. Mini-modele

2.1. Mini-modele: liniowe i nieliniowe modele lokalne

Koncepcja mini-modeli jest podobna do metody k najbliższych sąsiadów (kNN). Podczas obliczania wyjścia modelu dla zadanego punktu wejściowego $\mathbf{x}^*$, branych jest pod uwagę tylko k najbliższych (w sensie założonej metryki) próbek sąsiednich. W dalszej części pracy, do wyznaczania odległości stosowano metrykę euklidesową.

W klasycznej (prostej) wersji metody kNN, wyjście modelu jest obliczane jako średnia arytmetyczna lub średnia ważona wartości zadanych z analizowanych próbek. Przy wyznaczaniu średniej ważonej, wartości wag przypisanych do próbek zwykle zależą od odległości $\delta(\mathbf{x}^*;\mathbf{x})$ pomiędzy wejściem $\mathbf{x}^*$ i analizowanymi sąsiadami $\mathbf{x}$. Przykładowo:

$$w_{x^*;x} = \frac{1}{1 + m \cdot (\delta(\mathbf{x}^*;\mathbf{x})/k)^2}, \tag{2.1}$$

gdzie: parametr m jest dobierany empirycznie.

Każdy mini-model jest lokalną regresją, a odpowiedź dla wejścia $\mathbf{x}^*$ modelu globalnego jest wyznaczana na podstawie modelu lokalnego tworzonego dla k najbliższych sąsiadów. Mini-modele są tworzone w czasie wyznaczania wyjścia modelu (oznacza to, że podobnie jak w innych metodach pamięciowych, w czasie uczenia próbki są jedynie zapamiętywane, a lokalny mini-model jest tworzony podczas obliczania wyjścia dla zadanego punktu wejściowego $\mathbf{x}^*$) [105, 107].

W najprostszym przypadku można zastosować mini-modele liniowe. Wyjście będzie wtedy obliczane na podstawie liniowej regresji:

$$f(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{x}^*, \tag{2.2}$$

gdzie: $\mathbf{w}$ – wektor współczynników liniowego mini-modelu, wyznaczony jest dla k najbliższych sąsiadów.

Wektor $\mathbf{w}$ możemy wyznaczyć za pomocą metody najmniejszych kwadratów [59]. Dla k sąsiadów możemy utworzyć macierz danych wejściowych $\mathbf{X}$ i wektor zadanych wyjść

$\mathbf{Y}$:

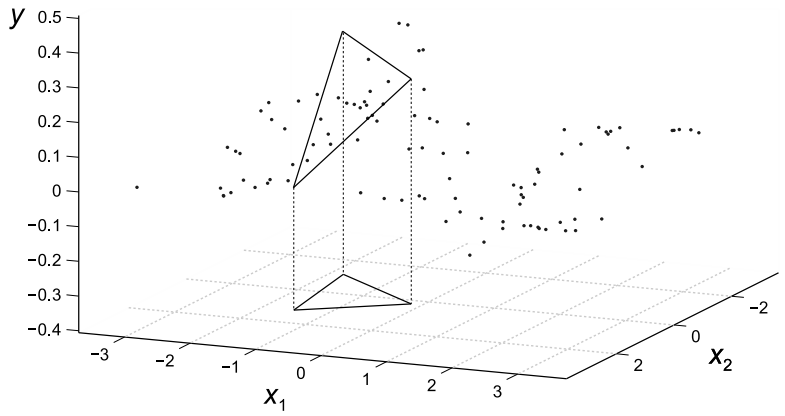
$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{N1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1k} & x_{2k} & \dots & x_{Nk} \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}, \tag{2.3}$$

gdzie: $i = 1 \dots N$ – numer wejścia, $j = 1 \dots k$ – numer próbki.

Wektor parametrów modelu, minimalizujący średni błąd kwadratowy, możemy wyznaczyć z zależności:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \cdot \mathbf{Y}. \tag{2.4}$$

W artykułach A. Piegata i in. [96, 97] opisane są mini-modele tworzone dla wycinków przestrzeni wejściowej o kształcie trójkątnym (dla 2-wymiarowej przestrzeni wejściowej) lub simpleksowym (w wielowymiarowej przestrzeni wejściowej). Taki wycinek będzie dalej nazywany bazą mini-modelu. Przykłady mini-modeli tego typu pokazane zostały na rys. 2.1 i 2.2.

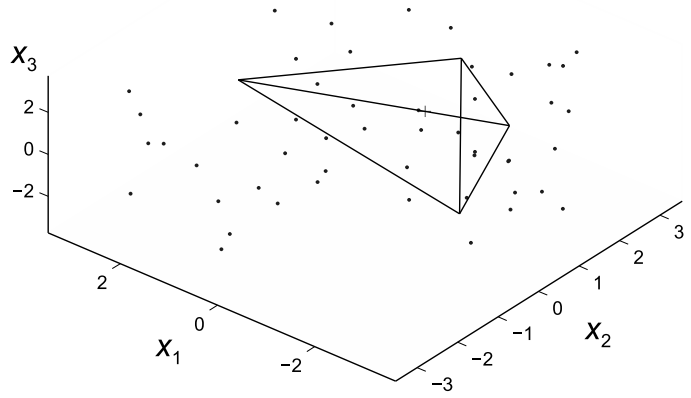


Rysunek 2.1. Przykładowy mini-model i jego baza w 2-wymiarowej przestrzeni wejściowej[1]

Przykłady zastosowania mini-modeli z bazą simpleksową, można także znaleźć w pracach M. Pietrzykowskiego [98, 99, 100, 101, 102]. Dobór parametrów w mini-modelach tego typu jest często kłopotliwy, ze względu na nieliniowy charakter funkcji celu, zmieniającej skokowo swoją wartość wraz ze zmianą położenia bazy.

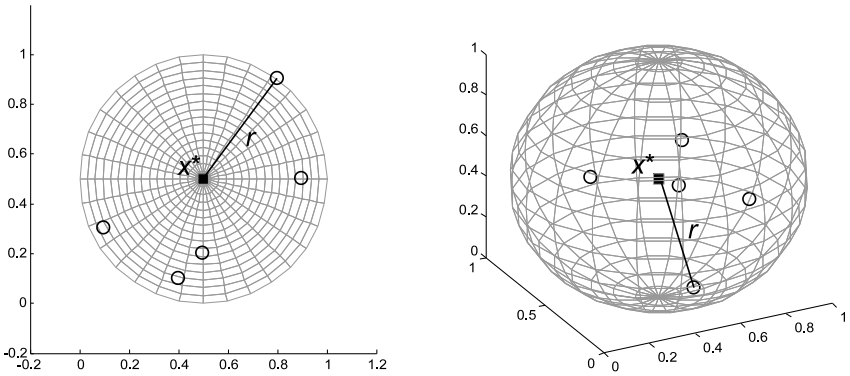
Przedstawione w dalszej części rozprawy mini-modele będą miały zawsze bazę kołową w 2-wymiarowej przestrzeni wejściowej, sferyczną w przestrzeni 3-wymiarowej lub hiper-sferyczną w przestrzeni o większej wymiarowości.

[1] Źródło rysunku i wszystkich kolejnych w monografii: opracowanie własne.



Rysunek 2.2. Przykładowa baza mini-modelu w 3-wymiarowej przestrzeni wejściowej

Baza mini-modelu ma środek w zadanym punkcie wejściowym $\mathbf{x}^*$, a jej promień jest zdefiniowany jako odległość między punktem $\mathbf{x}^*$ a najbardziej odległym punktem spośród k sąsiadów (rys. 2.3).



Rysunek 2.3. Baza mini-modelu dla 2- i 3-wymiarowej przestrzeni wejściowej

Mini-modele nieliniowe mają lepsze możliwości dopasowywania się do próbek. Wyjście takiego modelu będzie sumą mini-modelu liniowego oraz dodatkowego nieliniowego składnika:

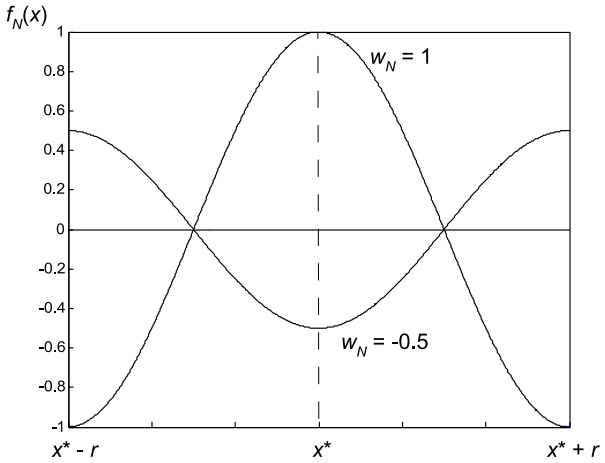
$$f(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{x}^* + f_N(\mathbf{x}^*). \tag{2.5}$$

Ponieważ mini-model jest zwykle tworzony dla niewielkiej liczby k najbliższych sąsiadów, nieliniowy składnik f_N powinien mieć możliwość zmiany kształtu dzięki tak małej ilości współczynników jak to tylko możliwe (ponieważ $n+1$ współczynników musi być dobranych w wektorze $\mathbf{w}$).

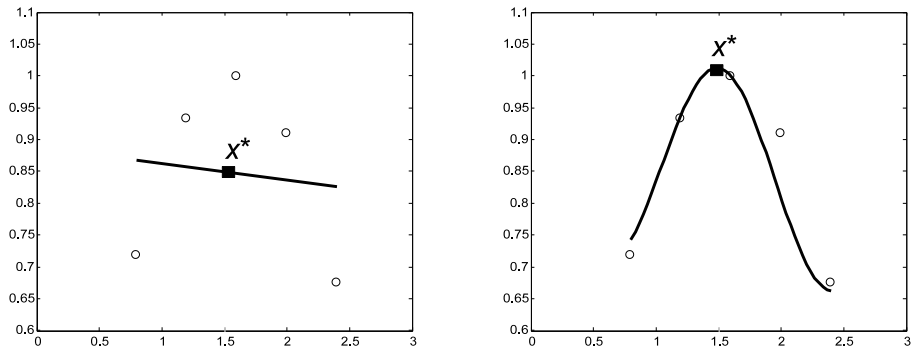
Spośród wielu zbadanych, bardzo korzystne własności miała funkcja:

$$f_N(\mathbf{x}) = w_N \cdot \sin\left[\frac{\pi}{2} - ||\mathbf{x} - \mathbf{x}^*|| \frac{\pi}{r}\right], \tag{2.6}$$

gdzie: r to promień bazy mini-modelu. W funkcji utworzonej w ten sposób, występuje tylko jeden współczynnik w_N do dobrania (rys. 2.4). W procesie strojenia należy dobrać taką wartość w_N, aby zapewnić najlepsze dopasowanie mini-modelu do k sąsiadów. Przykładowe liniowe i nieliniowe mini-modele w przestrzeniach wejściowych 1- i 2-wymiarowych pokazane zostały na rys. 2.5 i 2.6.

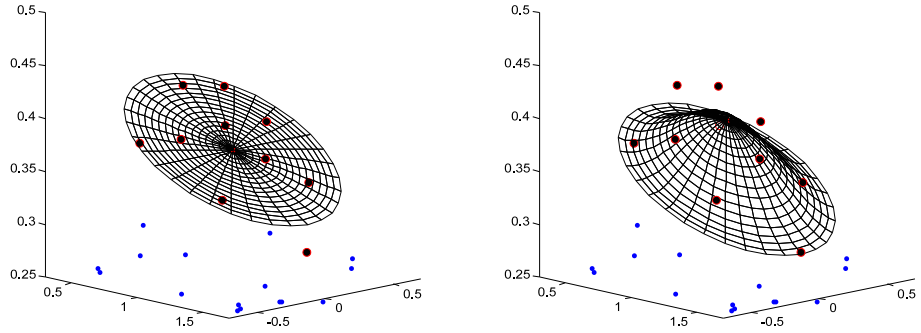


Rysunek 2.4. Przykładowy kształt nieliniowego składnika mini-modelu dla $w_N = 1$ i $w_N = -0.5$



Rysunek 2.5. Przykłady liniowych i nieliniowych mini-modeli w 1-wymiarowej przestrzeni wejściowej

Jednym z najważniejszych parametrów mini-modelu jest liczba sąsiadów k, brana pod uwagę w obliczeniach. Liczba ta może być stała, jednak może być także zmienna, zależna od położenia punktu w przestrzeni wejściowej. Popularną techniką doboru parametru k jest zastosowanie walidacji krzyżowej (np. metody minus jednego elementu) albo korzystanie z dwóch różnych zbiorów danych: danych uczących, które są zapamiętywane podczas tworzenia modelu oraz danych walidujących, służących do szacowania rzeczywistego błędu modelu. Najlepsza wartość k to taka, która zapewnia najmniejszy błąd walidacji lub kroswalidacji. Najmniejszy błąd danych walidujących gwarantuje najmniejszy błąd rzeczywisty modelu i tym samym najlepszą generalizację [13, 44, 107].



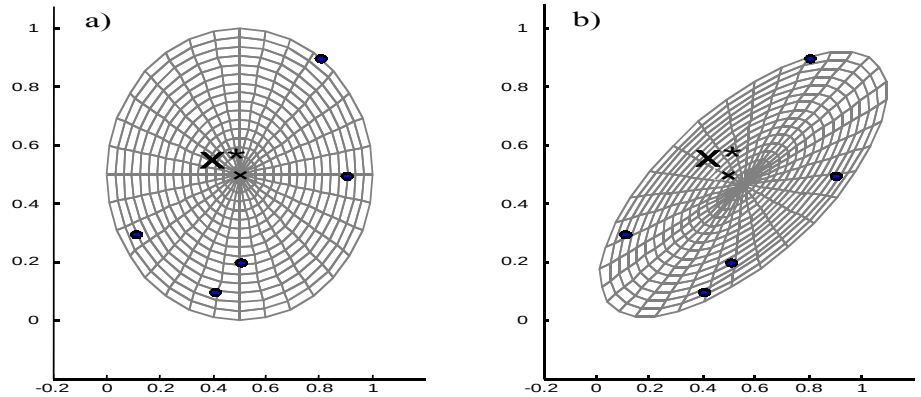
Rysunek 2.6. Przykłady liniowych i nieliniowych mini-modeli w 2-wymiarowej przestrzeni wejściowej

Funkcja aproksymująca oparta na mini-modelach posiada wiele użytecznych własności [105]. Zapewnia dobrą dokładność modelu i korzystne własności ekstrapolacyjne. Uczenie funkcji jest szybkie i niezłożone obliczeniowo.

2.2. Mini-modele z bazą hiper-elipsoidalną

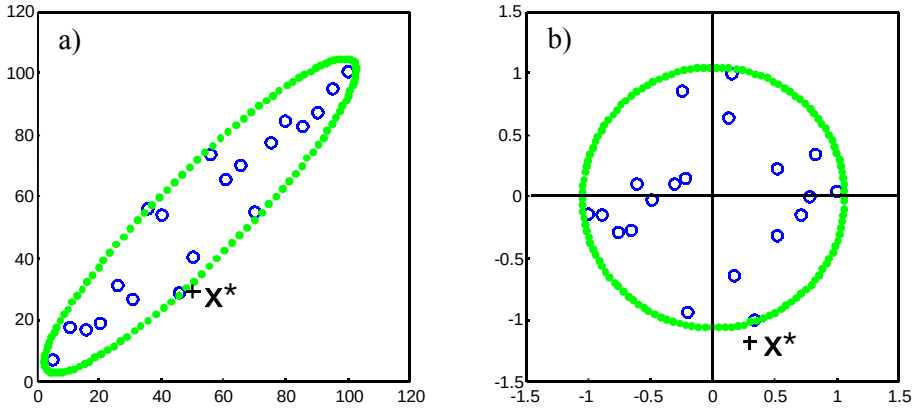
Bardzo często dane nie są rozłożone równomiernie w przestrzeni wejściowej. Sytuacja taka może mieć miejsce w przypadku przestrzeni wejściowych o dużej wymiarowości. Dla takich danych, jakość modelu można poprawić stosując mini-modele o bazie innej niż hiper-sferyczna. Mini-modele z bazą simpleksową [96, 97] są bardziej elastyczne w dopasowywaniu się do danych, jednak bardzo trudno je stosować w przestrzeniach wejściowych o większej wymiarowości [99, 101].

Przykładem mini-modelu, który jest bardziej elastyczny w dopasowaniu do danych, a jednocześnie nadal łatwy w strojeniu, jest mini-model z bazą hiper-elipsoidalną [106]. Przykład takiej bazy mini-modelu w 2-wymiarowej przestrzeni wejściowej pokazano na rys. 2.7.



Rysunek 2.7. Przykład kołowej i eliptycznej bazy mini-modelu w 2-wymiarowej przestrzeni wejściowej

Podczas tworzenia mini-modelu z hiper-elipsoidalną bazą, dane lokalne, które są wybrane, aby utworzyć jego bazę (k najbliższych sąsiadów) podlegają transformacji PCA, a następnie są normalizowane. Proces transformacji danych zilustrowano na rys. 2.8.



Rysunek 2.8. Przykładowe dane przed (a) i po (b) transformacji

Po transformacji, próbki rozłożone są równomiernie w całej przestrzeni wejściowej (rys. 2.8b), a ich wartości należą do znormalizowanego przedziału $[-1, 1]$. Tak przygotowane nowe dane lokalne mogą być wykorzystane do utworzenia nieliniowego hiper-sferycznego mini-modelu z wykorzystaniem równań podobnych do (2.5) i (2.6). Mini-model posiada bazę hiper-sferyczną w nowej (po transformacji) przestrzeni wejść, jednak w przestrzeni oryginalnej jego baza posiada kształt hiper-elipsoidalny i tym samym może lepiej dopasować się do danych (rys. 2.8a i 2.9). Należy tu podkreślić, że mini-model o bardziej złożonym kształcie bazy jest tworzony z taką samą liczbą współczynników jak jego odpowiednik z bazą hiper-sferyczną, a jego strojenie przebiega w podobny sposób (konieczne jest jedynie wstępne przeprowadzenie transformacji PCA i normalizacji, co nie jest kosztowne obliczeniowo, gdyż wykonywane jest dla niewielkiej liczby próbek k).

Jest niewielka różnica w obliczaniu wyjścia modelu opartego na mini-modelach hiper-elipsoidalnych. Środek bazy mini-modelu nie pokrywa się teraz z punktem wejściowym $\mathbf{x}^*$. Leży on w początku nowego (po transformacji) układu współrzędnych, co zaprezentowano na rys. 2.8.

Kompletny algorytm obliczania wyjścia dla punktu wejściowego $\mathbf{x}^*$ może być opisany w następujących krokach.

1. Znajdź k najbliższych sąsiadów punktu wejściowego $\mathbf{x}^*$.
2. Przeprowadź transformację PCA i normalizację wybranych próbek.
3. Przeprowadź transformację punktu wejściowego $\mathbf{x}^*$ do nowego układu współrzędnych $\mathbf{x}^* \rightarrow \mathbf{x}^*_{PCA}$.

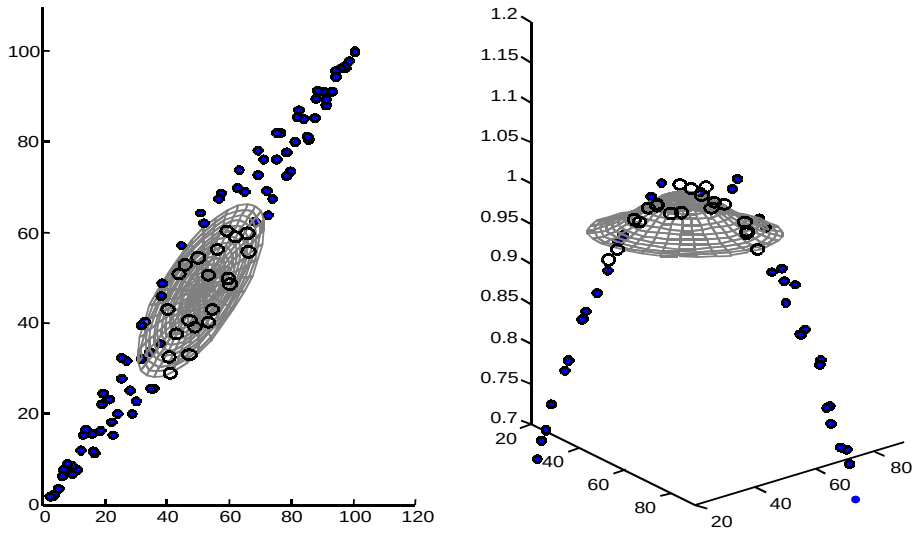
4. Dobierz parametry mini-modelu i oblicz jego odpowiedź dla punktu $\mathbf{x}^*_{PCA}$ zgodnie ze wzorami:

$$f(\mathbf{x}^*_{PCA}) = \mathbf{w}^T \cdot \mathbf{x}^*_{PCA} + f_N(\mathbf{x}^*_{PCA}), \tag{2.7}$$

$$f_N(\mathbf{x}_{PCA}) = w_N \cdot \sin\left[\frac{\pi}{2} - \mathbf{x}^*_{PCA} \cdot \frac{\pi}{r}\right], \tag{2.8}$$

gdzie: $r = 1$ w nowym układzie współrzędnych (po transformacji i normalizacji).

Przykład mini-modelu z bazą elipsoidalną utworzonego dla próbek w 2-wymiarowej przestrzeni wejściowej pokazano na rys. 2.9.

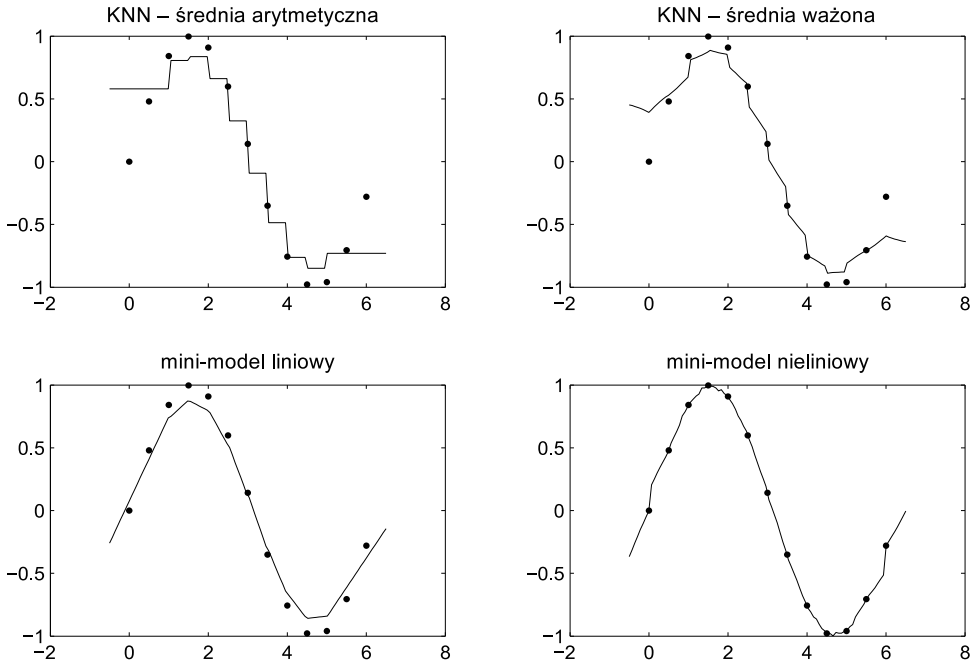


Rysunek 2.9. Przykład mini-modelu z bazą elipsoidalną utworzonego dla próbek w 2-wymiarowej przestrzeni wejściowej

2.3. Eksperymenty

Dla lepszego zobrazowania działania mini-modeli, wstępne eksperymenty przeprowadzono dla danych o 1- i 2-wymiarowej przestrzeni wejściowej [105]. Na rys. 2.10 pokazano dane uczące i charakterystyki modeli utworzonych z wykorzystaniem liniowych i nielinowych mini-modeli. Dla porównania, przedstawiono także charakterystyki modeli utworzonych za pomocą metody k najbliższych sąsiadów (kNN).

Modele bazujące na mini-modelach mają bardzo korzystne własności ekstrapolacyjne, co przedstawione zostało na rys. 2.11. Tak jak wcześniej, na wykresach pokazane są dane uczące oraz charakterystyki modeli utworzonych z wykorzystaniem metod kNN i mini-modeli. Przede wszystkim, należy zwrócić uwagę na kształt charakterystyk w przedziałach, w których brak jest próbek (luki informacyjne) oraz poza zakresem do którego



Rysunek 2.10. Charakterystyki modeli utworzonych metodami kNN oraz z wykorzystaniem mini-modeli, dla danych należących do 1-wymiarowej przestrzeni wejściowej

należą. Modele oparte na mini-modelach dają wyjścia, które są zgodne ze zdrowym rozsądkiem, a dodatkowo ich charakterystyki są bardziej gładkie.

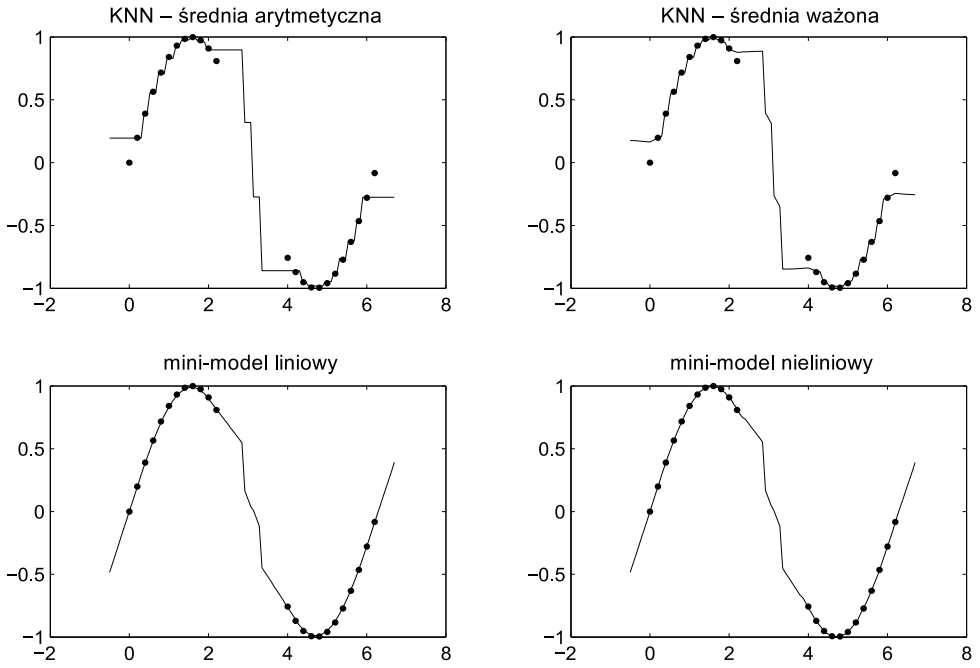
Na rys. 2.12 pokazano dane uczące i powierzchnie charakterystyk modeli utworzonych w 2-wymiarowej przestrzeni wejściowej. Oba modele oparte na mini-modelach (rys. 2.10 i 2.12) charakteryzują się lepszą dokładnością niż modele utworzone za pomocą metod kNN. Wartości błędów modeli podane zostały w tabeli 2.1.

W kolejnych eksperymentach wyznaczany był błąd rzeczywisty aproksymatorów funkcji opartych na metodach kNN i mini-modelach, (tab. 2.1). Badania przeprowadzone były dla danych utworzonych przez autora oraz danych pochodzących z popularnych repozytoriów sieciowych[2]. Ponieważ atrybuty wejściowe próbek należały do różnych zakresów, przed rozpoczęciem obliczeń dane były zawsze normalizowane (do przedziału [0,1]).

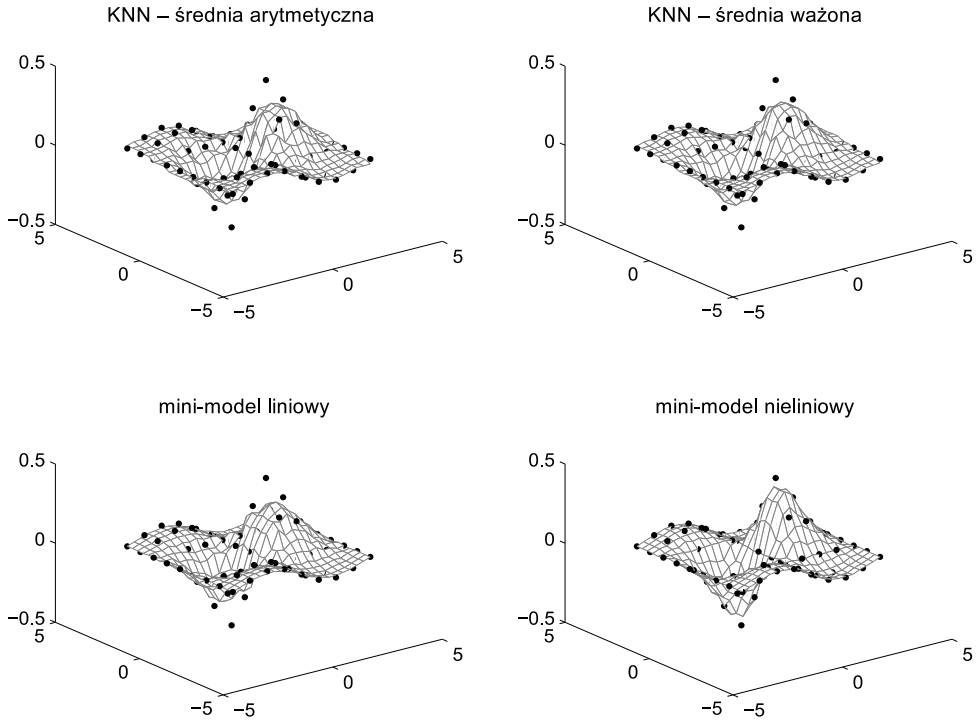
Błąd rzeczywisty wyznaczany był za pomocą metody minus jednego elementu. Błąd aproksymatora opartego na mini-modelach porównany został z błędem aproksymatora kNN, przy czym we wszystkich przypadkach zaprezentowano wyniki obliczeń dla optymalnej liczby próbek sąsiednich (czyli gwarantującej najmniejszą wartość błędu). Dodatkowo, dla porównania, pokazany został także błąd modelu wyznaczony dla uogólnionej sieci regresyjnej (GRN), w której szerokość neuronów również dobrana była optymalnie.

Szczególnie interesujące wyniki uzyskano dla danych z pliku `elusage` (rys. 2.13). Model oparty na mini-modelach z elipsoidalną bazą zapewnił tu najmniejszy błąd rzeczy-

[2] `http://archive.ics.uci.edu/ml/` – The UCI Machine Learning Repository: zbiór danych i generatorów wykorzystywanych do empirycznych badań algorytmów uczenia maszynowego.



Rysunek 2.11. Charakterystyki modeli utworzonych dla danych z luką informacyjną

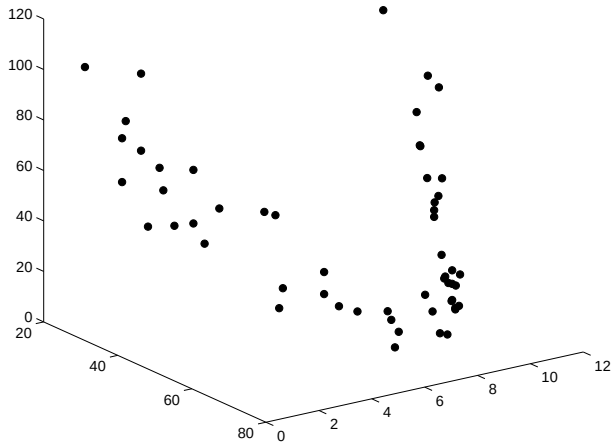


Rysunek 2.12. Charakterystyki modeli utworzonych metodami kNN oraz z wykorzystaniem mini-modeli, dla danych należących do 2-wymiarowej przestrzeni wejściowej

Tabela 2.1. Błąd rzeczywisty aproksymatorów funkcji

dane	liczba wejść	kNN – śr. arytmet.	kNN – śr. ważona	mini-model liniowy	mini-model nieliniowy	mini-model z bazą elipsoidalną	GRN
$\sin(x)$ – krok próbkowania 0.5 (rys. 2.10)	1	0.166	0.161	0.095	0.086	0.086	0.142
$\sin(x)$ – krok próbkowania 0.2	1	0.031	0.030	0.013	0.008	0.008	0.025
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – krok próbkowania 0.25 (rys. 2.12)	2	0.0281	0.0277	0.0228	0.0259	0.0237	0.0254
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – krok próbkowania 0.1	2	0.0058	0.0058	0.0049	0.0055	0.0054	0.0052
`bodyfat`	14	2.236	2.211	0.472	0.475	0.476	2.671
`cpu`	6	29.507	28.937	27.305	27.826	26.779	28.771
`diabetes_numeric`	2	0.474	0.481	0.475	0.487	0.479	0.471
`housing`	13	2.771	2.685	2.311	2.293	2.346	2.506
`elusage`	2	8.885	8.984	8.648	9.154	8.539	9.323

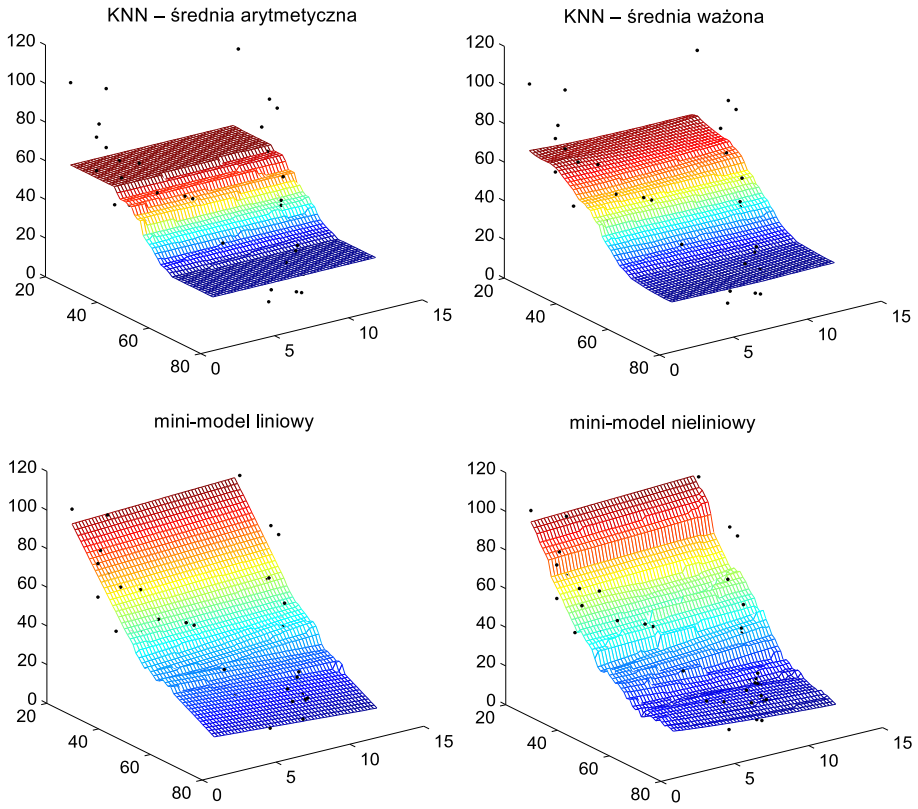
wisty. Przestrzeń wejściowa dla tych danych jest 2-wymiarowa, stąd można je zwizualizować (rys. 2.14 i 2.15). Z danych przedstawionych na rysunkach wynika, że próbki nie są rozłożone równomiernie w całej przestrzeni wejść, układając się w formie pasa. W takiej sytuacji mini-modele z bazą hiper-elipsoidalną powinny zagwarantować lepszą dokładność modelu globalnego niż modele z bazą hiper-sferyczną.



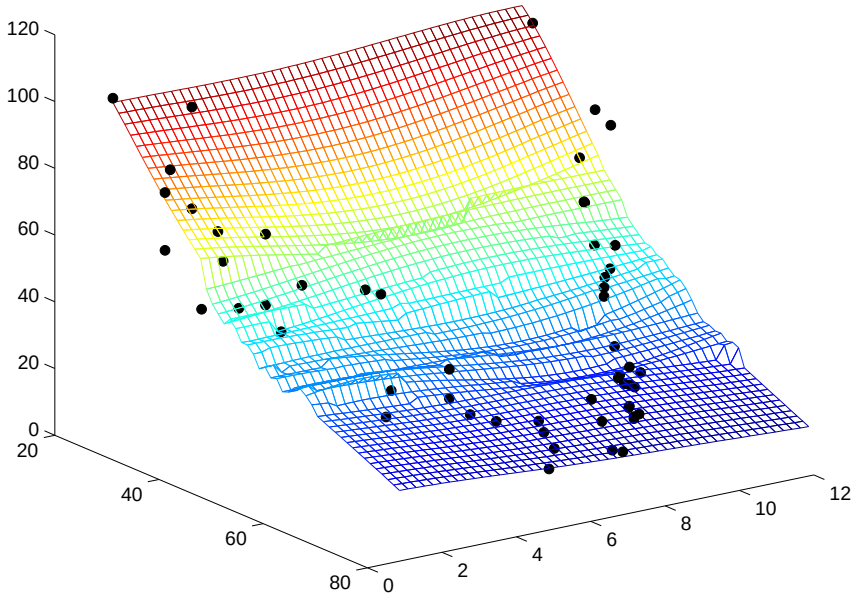
Rysunek 2.13. Położenie próbek w danych `elusage`

Podsumowując wyniki eksperymentów, modele utworzone na bazie mini-modeli wykazały się dobrą dokładnością (tab. 2.1). Dokładność ta jest szczególnie wysoka dla danych niezaszumionych. W przypadku danych z szumem, modele bazujące na mini-modelach posiadały dokładność porównywalną lub nieznacznie mniejszą od modeli opartych na metodach kNN.

Uczenie aproksymatorów jest bardzo szybkie – tak naprawdę wystarczy zapamiętać dane uczące, a lokalny mini-model jest tworzony podczas obliczania wyjścia dla zadanego



Rysunek 2.14. Charakterystyki modeli utworzonych metodami kNN oraz z wykorzystaniem mini-modeli dla danych `elusage`



Rysunek 2.15. Charakterystyka modelu utworzonego z wykorzystaniem mini-modeli z eliptyczną bazą dla danych `elusage`

punktu wejściowego $\mathbf{x}^*$. Tworzenie mini-modeli nie jest złożone obliczeniowo, ponieważ są one budowane na podstawie niewielkiej ilości próbek (k najbliższych sąsiadów). Dla mini-modelu liniowego poszukujemy liniowej regresji. W przypadku mini-modelu nieliniowego postępujemy podobnie, a dodatkowo wyznaczamy jeszcze jeden współczynnik występujący przy składowej nieliniowej.

Modele oparte na mini-modelach mają bardzo korzystne własności ekstrapolacyjne. Wynikają one z faktu, że każdy nimi-model modeluje nie tylko wartości zadane próbek uczących, ale także trend zmiany wartości istniejący w sąsiedztwie punktu wejściowego $\mathbf{x}^*$. Wykorzystanie informacji o trendzie umożliwia lepsze zachowanie modelu w tych obszarach przestrzeni wejść, dla których brak jest danych uczących (luki informacyjne i obszary spoza dziedziny wyznaczonej przez dane). Sytuacje takie są bardzo charakterystyczne dla danych wielowymiarowych, dla których ilość danych jest niewielka. W przypadkach nierównomiernego pokrycia przestrzeni wejść danymi uczącymi, dobrą alternatywą może być zastosowanie modeli bazujących na mini-modelach z bazą hiper-elipsoidalną.

Pewnym mankamentem mini-modeli, w porównaniu z technikami kNN, jest konieczność brania pod uwagę większej liczby próbek k podczas tworzenia modelu. W metodzie kNN można w skrajnym przypadku przyjąć $k = 1$, podczas gdy minimalna liczba k sąsiadów wykorzystanych do utworzenia mini-modelu liniowego wynosi $n+1$, gdzie n to rozmiar przestrzeni wejściowej.

3. Interwały

3.1. Pojęcia podstawowe

Z najczęstszym typem niepewności mamy do czynienia w sytuacji, w której zamiast dokładnej wartości x znamy dla niej jedynie dolne $\underline{x}$ i górne $\overline{x}$ ograniczenie. Nie wiemy jakie wartości interwału $[\underline{x}, \overline{x}]$ są bardziej lub mniej możliwe czy prawdopodobne. Niepewność tego typu będziemy nazywali niepewnością interwałową, a wartości, liczbami interwałowymi lub prościej interwałami.

Formalnie, interwał X będziemy definiowali jako zbiór liczb rzeczywistych [45, 73, 74]:

$$X = [\underline{x}, \overline{x}] = \{x \in \mathbb{R} : \ \underline{x} \leqslant x \leqslant \overline{x}\}. \tag{3.1}$$

Zbiór wszystkich interwałów oznaczymy jako $\mathbb{IR}$. Liczby rzeczywiste również mogą być opisywane jako interwały w postaci $X = [x, x]$, dla których dolne i górne ograniczenie ma tę samą wartość. Będziemy je nazywać interwałami zdegenerowanymi. W takim ujęciu zbiór interwałów możemy traktować jako rozszerzenie zbioru liczb rzeczywistych: $\mathbb{R} \in \mathbb{IR}$ [40].

Opracowana została arytmetyka interwałowa, definiująca sposób realizacji podstawowych operacji arytmetycznych [73, 74]. Dla dwóch interwałów $X = [\underline{x}, \overline{x}]$ i $Y = [\underline{y}, \overline{y}]$ poszczególne działania wykonywane są następująco:

- dodawanie:

$$X + Y = [\underline{x} + \underline{y}, \overline{x} + \overline{y}], \tag{3.2}$$

- odejmowanie:

$$X - Y = [\underline{x} - \overline{y}, \overline{x} - \underline{y}], \tag{3.3}$$

- mnożenie:

$$X \cdot Y = [\min(Z), \max(Z)], \text{ gdzie} : Z = \{\underline{x} \cdot \underline{y}, \underline{x} \cdot \overline{y}, \overline{x} \cdot \underline{y}, \overline{x} \cdot \overline{y}\}, \tag{3.4}$$

- dzielenie:

$$X/Y = X \cdot \frac{1}{Y}, \text{ gdzie} : \frac{1}{Y} = [1/\overline{y}, 1/\underline{y}], \text{ przy czym } 0 \notin Y. \tag{3.5}$$

3.2. Wielowymiarowa arytmetyka interwałowa RDM

Za pomocą zaprezentowanej powyżej arytmetyki interwałowej (znanej jako arytmetyka Moore'a) można poprawnie realizować różnorodne operacje matematyczne, choć w sposób uproszczony, przede wszystkim bez uwzględnienia zależności jakie mogą występować pomiędzy poszczególnymi operandami. Podejście takie może powodować wiele paradoksów, które opisane zostały w literaturze [27, 119]. Najważniejsze mankamenty arytmetyki opracowanej przez R. Moore'a wyliczono poniżej [92, 94]:

a) z każdą kolejną operacją rośnie szerokość interwału wynikowego;
b) brak możliwości uwzględnienia zależności występujących pomiędzy operandami;
c) kłopotliwe rozwiązywanie nawet najprostszych równań interwałowych.

Problem c) można zilustrować przykładem równania z jedną niewiadomą:

$$\begin{aligned} &[\underline{a}, \overline{a}] + [\underline{x}, \overline{x}] = [\underline{c}, \overline{c}], \\ &[1, 3] + [\underline{x}, \overline{x}] = [3, 4]. \end{aligned} \tag{3.6}$$

Równanie takie można spróbować rozwiązać w następujący sposób:

$$\begin{aligned} 1 + \underline{x} &= 3, \qquad & \underline{x} &= 2, \\ 3 + \overline{x} &= 4, \qquad & \overline{x} &= 1. \end{aligned}$$

Ostatecznie otrzymujemy rozwiązanie, które nie ma sensu, ponieważ: $\underline{x} > \overline{x}$. Równanie (3.6) można także rozwiązać w inny sposób:

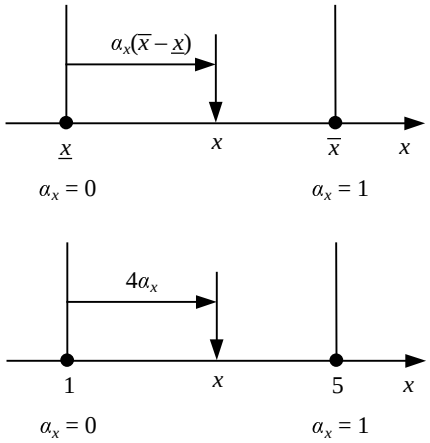
$$[\underline{x}, \overline{x}] = [3, 4] - [1, 3] = [0, 3].$$

Jednak takie rozwiązanie nie spełnia równania (3.6) po podstawieniu:

$$[1, 3] + [0, 3] \neq [3, 4].$$

Dobrą alternatywą dla arytmetyki Moore'a może być być opis interwałów z wykorzystaniem notacji RDM [92, 94]. Autorem tej koncepcji jest prof. Andrzej Piegat

[87, 88]. Wielowymiarowa arytmetyka RDM wprowadza dodatkową zmienną wewnętrzną $\alpha \in [0,1]$, która jest względną miarą odległości (ang. *relative-distance-measure – RDM*) od dolnego ograniczenia interwału [87, 88, 89], rys. 3.1.



Rysunek 3.1. Przykładowa zmienna wewnętrzna α, będąca względną miarą odległości od dolnego ograniczenia interwału

W notacji RDM interwał X można opisać jako:

$$X = [\underline{x}, \overline{x}] = \underline{x} + \alpha_x(\underline{x} - \overline{x}), \quad \alpha_x \in [0,1]. \tag{3.7}$$

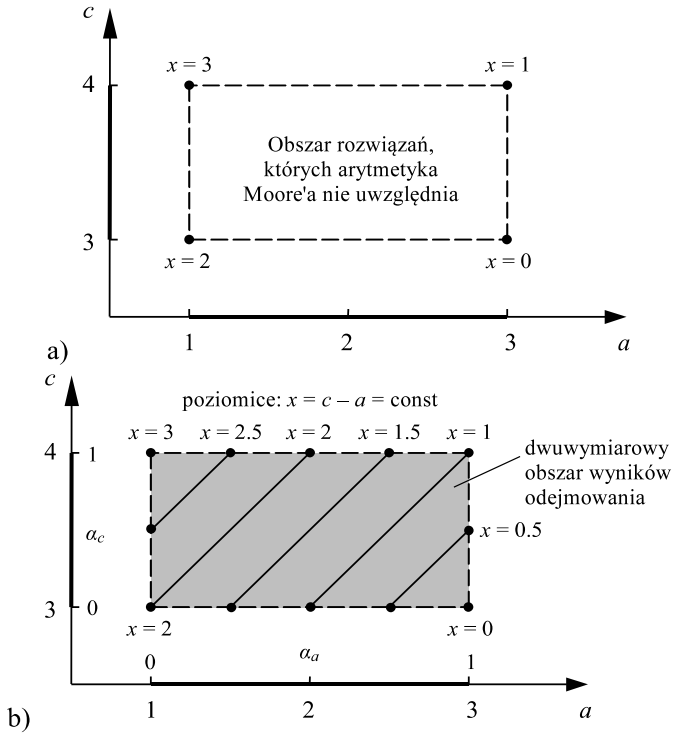
Przykładowo, interwał $A = [1,4]$ w notacji RDM będzie miał postać:

$$A = 1 + 3\alpha_A, \quad \alpha_A \in [0,1].$$

Celem wprowadzenia do obliczeń dodatkowej zmiennej RDM nie jest niepotrzebna parametryzacja interwału, ale zdefiniowanie lokalnej współrzędnej, umożliwiającej uwzględnienie w obliczeniach jego wnętrza. Możemy zauważyć, że w arytmetyce interwałowej Moore'a, w obliczeniach biorą udział tylko wartości brzegowe interwałów. Wartości wewnętrzne nie są uwzględniane.

Rozpatrzmy odejmowanie $A - C = X = [3,4] - [1,3] = [0,3]$ realizowane zgodnie z arytmetyką Moore'a. Wynik zaprezentowany został na rys. 3.2a. Jeśli wykonamy dodawanie $X + C = [0,3] + [1,3]$, wówczas otrzymamy w wyniku interwał $[1,6]$, a nie $A = [3,4]$. Dlaczego? Ponieważ interwał $[0,3]$ nie jest właściwym wynikiem, a jedynie reprezentacją jego rozpiętości, co pokazano na rys. 3.2b. Wartości minimalne i maksymalne funkcji matematycznych nie zawsze muszą się znaleźć na brzegach dziedziny, mogą także leżeć w jej wnętrzu i w takich przypadkach, nie zostaną wyznaczone przez arytmetykę Moore'a.

Dzięki zmiennej α, arytmetyka RDM wprowadza do interwału lokalną współrzędną i pozwala na uwzględnienie w obliczeniach także jego wnętrza. Jest to szczególnie istotne



Rysunek 3.2. Wynik odejmowania interwałów $C - A = X = [3, 4] - [1, 3]$: a) arytmetyka Moore'a, b) arytmetyka RDM

w bardziej skomplikowanych obliczeniach, dla których największe lub najmniejsze wartości wyniku nie są efektem działań na wartościach brzegowych operandów.

Arytmetyka RDM posiada prawie te same własności co klasyczna arytmetyka liczb rzeczywistych [62, 92, 94] Załóżmy, że A,B,C to interwały. Najważniejsze własności arytmetyki RDM przedstawiono poniżej.

1. $A + B = B + A$, $AB = BA$ – prawo przemienności dodawania i mnożenia.
2. $A+(B+C) = (A+B)+C$, $A(BC) = (AB)C$ – prawo łączności dodawania i mnożenia.
3. Dla każdego interwału A należącego do $\mathbb{IR}$ istnieje interwał $-A$ należący do $\mathbb{IR}$ taki, że $A + (-A) = (-A) + A = 0$. $-A$ jest liczbą przeciwną do A.
4. Dla każdego interwału A należącego do $\mathbb{IR}$, $0 \notin A$, istnieje interwał $A^{-1} = 1/A$ należący do $\mathbb{IR}$ taki, że $AA^{-1} = A(1/A) = 1$. A^{-1} jest liczbą odwrotną do A.
5. $A(B + C) = (AB) + (AC)$, $(B + C)A = (BA) + (CA)$ – prawo rozdzielności mnożenia względem dodawania.
6. $A + C = B + C \Rightarrow A = B$.
7. $CA = CB \Rightarrow A = B$.

W przypadku arytmetyki Moore'a, prawa 3, 4, 5 i 7 nie są spełnione [74], czego najważniejszą konsekwencją jest niemożliwość przekształcania wzorów. Przykładowo, w równaniu $A + X = C$ przeniesienie A na prawą stronę, aby otrzymać wyrażenie $X = C - A$, nie jest dozwolone, ponieważ prawo nr 3 nie jest spełnione. Ponieważ nie jest możliwe

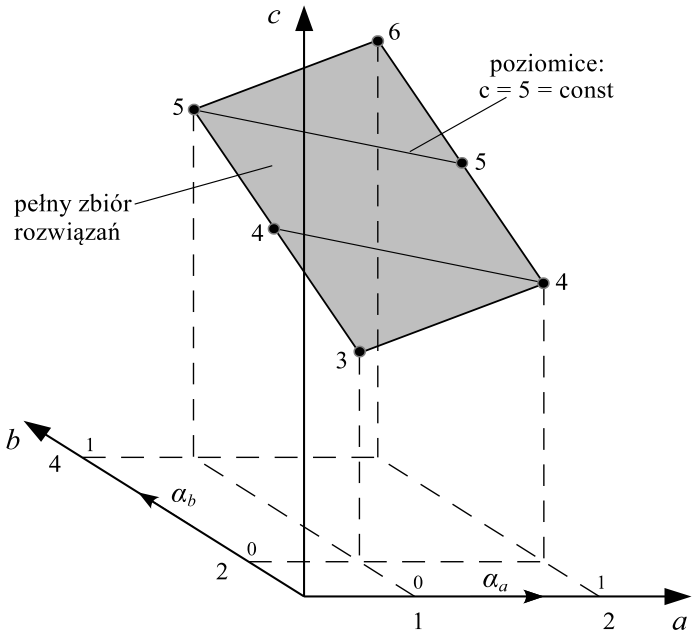
przekształcanie wzorów, wiele bardziej skomplikowanych problemów matematycznych nie może być skutecznie rozwiązanych.

Udowodnienie prawdziwości powyższych praw dla arytmetyki RDM jest proste. Wystarczy w miejsce powyższych wzorów wstawić interwały A,B,C zapisane w formie RDM. Do najważniejszych korzyści jakie daje arytmetyka RDM można zaliczyć:

a) możliwość rozwiązywania złożonych problemów (przede wszystkim równań), dzięki możliwości skutecznego przekształcania wzorów;
b) prawie wszystkie prawa arytmetyki liczb rzeczywistych zachowane są dla arytmetyki RDM;
c) wynik wyznaczony z wykorzystaniem arytmetyki RDM jest kompletnym, wielowymiarowym rozwiązaniem, z którego możemy wyznaczyć różne uproszczone reprezentacje, jak: rozkład kardynalności, rozpiętość wartości (rozwiązanie jakie daje arytmetyka Moore'a), czy środek ciężkości.

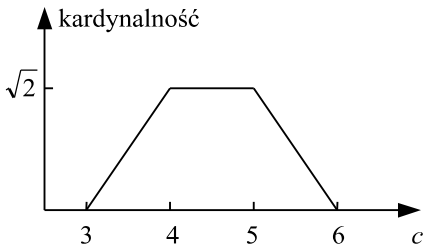
Ostatnią korzyść z powyższej listy można wyjaśnić za pomocą przykładu dodawania interwałów $A + B = C$, gdzie: $A = [1, 2]$, $B = [2, 4]$. W arytmetyce RDM interwały będą opisane z wykorzystaniem pomocniczych zmiennych RDM [92, 94].

$$\begin{aligned} &A: \ a = 1 + \alpha_a, \quad B: \ b = 2 + 2\alpha_b, \\ &c = a + b = 3 + \alpha_a + 2\alpha_b, \quad \alpha_a, \alpha_b \in [0, 1]. \end{aligned} \tag{3.8}$$



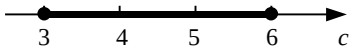
Rysunek 3.3. Kompletny, trójwymiarowy wynik $C: \ c = a + b = 3 + \alpha_a + 2\alpha_b, \ \alpha_a, \alpha_b \in [0, 1]$, dodawania interwałów $A + B = [1, 2] + [2, 4]$, otrzymany z wykorzystaniem arytmetyki RDM

Kompletny zbiór rozwiązań C nie jest jednowymiarowym interwałem, tylko trójwymiarową granulą opisaną równaniem (3.8). Na rys. 3.3, widoczne są linie stałych wartości wyniku dodawania ($c = a + b = \text{const}$). Długości tych linii (odcinków) są miarą kardynalności poszczególnych wartości rozwiązań (na przykład dla $c = 5 = \text{const}$). Rozkład kardynalności dla rozwiązań $c = \text{const}$ pokazano na rys. 3.4. Jest on dwuwymiarową reprezentacją kompletnego rozwiązania $c = a + b = 3 + \alpha_a + 2\alpha_b,\ \alpha_a, \alpha_b \in [0, 1]$.



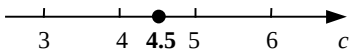
Rysunek 3.4. Rozkład kardynalności jako dwuwymiarowa reprezentacja kompletnego, trójwymiarowego rozwiązania $C:\ c = a + b = 3 + \alpha_a + 2\alpha_b,\ \alpha_a, \alpha_b \in [0, 1]$

Kolejnym przykładem reprezentacji pełnego zbioru rozwiązań z rys. 3.3 może być rozpiętość: $[\min(c), \max(c)] = [\underline{c}, \overline{c}] = [3, 6]$, pokazana na rys. 3.5. Rozpiętość $[\underline{c}, \overline{c}] = [3, 6]$ można wyznaczyć poprzez zwykłe badanie funkcji kompletnego zbioru rozwiązań $C:\ c = a + b = 3 + \alpha_a + 2\alpha_b,\ \alpha_a, \alpha_b \in [0, 1]$. Łatwo można wyznaczyć, że $\min(c) = \underline{c} = 3$ jest osiągane dla $\alpha_a = \alpha_b = 0$ i $\max(c) = \overline{c} = 6$ dla $\alpha_a = \alpha_b = 1$.



Rysunek 3.5. Rozpiętość $[3, 6]$ pełnego rozwiązania $C:\ c = a+b = 3+\alpha_a+2\alpha_b,\ \alpha_a, \alpha_b \in [0, 1]$, będąca rozwiązaniem Moore'a i jednowymiarową reprezentacją dodawania interwałów $A + B = C$

Ostatnią i najprostszą formą reprezentacji pełnego zbioru rozwiązań może być jego środek ciężkości, zaprezentowany na rys. 3.6.



Rysunek 3.6. Środek ciężkości pełnego zbioru rozwiązań $C:\ c = a + b = 3 + \alpha_a + 2\alpha_b,\ \alpha_a, \alpha_b \in [0, 1]$ dodawania interwałów $A + B$

Jedną z największych zalet arytmetyki RDM, jest możliwość uwzględnienia w obliczeniach zależności występujących między interwałami. W problemach rzeczywistych występują one bardzo często. Zależności takie można uwzględniać za pomocą zmiennych RDM [109].

Przykładowo, w klasycznej arytmetyce interwałowej Moore'a, w wyniku odejmowania dwóch identycznych interwałów mamy:

$$Y = X - X = [\underline{x}, \overline{x}] - [\underline{x}, \overline{x}] = [\underline{x} - \overline{x}, \overline{x} - \underline{x}]. \tag{3.9}$$

Jeśli $A = [1, 4]$:

$$A - A = [1, 4] - [1, 4] = [-3, 3].$$

W arytmetyce RDM, możemy wprowadzić dwie pomocnicze zmienne RDM: α_{x1}, $\alpha_{x2} \in [0, 1]$:

$$Y = X - X = (\underline{x} + \alpha_{x1}(\overline{x} - \underline{x})) - (\underline{x} + \alpha_{x2}(\overline{x} - \underline{x})) = (\alpha_{x1} - \alpha_{x2})(\overline{x} - \underline{x}). \tag{3.10}$$

Wynik jest funkcją dwóch zmiennych RDM α_{x1} i α_{x2}: $Y = f(\alpha_{x1}, \alpha_{x2})$. Możemy wyznaczyć z niego prostsze reprezentacje opisane powyżej, między innymi dolne i górne ograniczenie interwału będącego rozpiętością rozwiązania. Funkcja przyjmuje najmniejszą wartość dla $\alpha_{x1} = 0$ i $\alpha_{x2} = 1$ – w ten sposób wyznaczamy dolne ograniczenie rozpiętości, które jest równe $\underline{x} - \overline{x}$. Funkcja przyjmuje największą wartość dla $\alpha_{x1} = 1$ i $\alpha_{x2} = 0$ – w ten sposób wyznaczamy górne ograniczenie rozpiętości, które jest równe $\overline{x} - \underline{x}$.

Jeśli jednak wiemy, że we wzorze (3.9) występuje ten sam interwał X, możemy założyć że $\alpha_{x1} = \alpha_{x2} = \alpha_x$. Wówczas:

$$Y = X - X = (\underline{x} + \alpha_x(\overline{x} - \underline{x})) - (\underline{x} + \alpha_x(\overline{x} - \underline{x})) = 0, \tag{3.11}$$

co jest zgodne ze zdrowym rozsądkiem, ale nieosiągalne dla klasycznej arytmetyki Moore'a.

Aby rozwiązać równanie (3.6) można użyć 2 zmiennych RDM: α_a and α_c [92]. Wówczas interwał $[\underline{a}, \overline{a}] = [1, 3]$ przyjmuje postać: $a = 1 + 2\alpha_a$, $\alpha_a \in [0, 1]$, a interwał $[\underline{c}, \overline{c}] = [3, 4]$ można zapisać jako: $c = 3 + \alpha_c$, $\alpha_c \in [0, 1]$. Równanie (3.6) można teraz rozwiązać z wykorzystaniem reguł arytmetyki RDM:

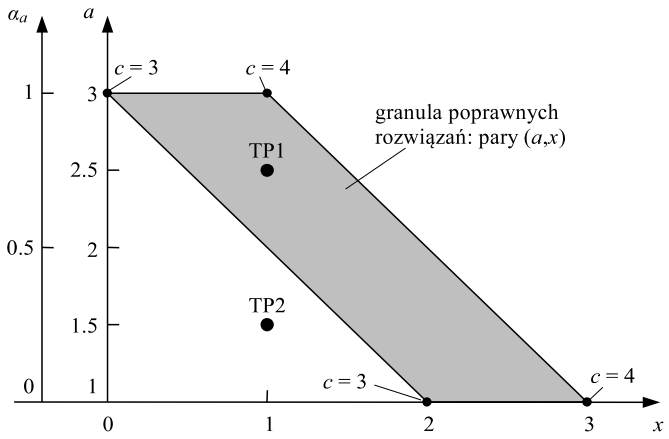
$$(1 + 2\alpha_a) + x = 3 + \alpha_c,$$

i ostatecznie:

$$x = 2 - 2\alpha_a + \alpha_c, \quad \alpha_a \in [0, 1],\ \alpha_c \in [0, 1]. \tag{3.12}$$

Należy zauważyć, że rozwiązanie x zależy od 2 zmiennych: α_a i α_c. Wynika z tego, że rozwiązanie nie jest jednowymiarowe, jak sugeruje arytmetyka Moore'a, ale dwuwy-

miarowe. Rozwiązanie pokazane jest na rys. 3.7, a rozwiązania dla brzegowych wartości zmiennych a i c dodatkowo pokazano w tabeli 3.1.



Rysunek 3.7. Zbiór rozwiązań równania $[1,3] + [\underline{x}, \overline{x}] = [3,4]$ składający się z par (a, x) spełniających zależność (3.12). TP1, TP2 – punkty testowe

Tabela 3.1. Wartości zmiennych a i c dla brzegowych wartości zmiennych RDM α_a i α_c

α_a	0	0	1	1
a	1	1	3	3
α_c	0	1	0	1
c	3	4	3	4
x	2	3	0	1

Powyższy przykład pokazuje, że w ogólnym przypadku nie jest możliwe znalezienie jednowymiarowego rozwiązania $[\underline{x}, \overline{x}]$ dla równania interwałowego. Poprawność wyniku można zbadać za pomocą punktów testowych. Załóżmy, że dane są 2 takie punkty pokazane na rys. 3.7: TP1($a = 2.5$, $x = 1$) i TP2($a = 1.5$, $x = 1$). Po wstawieniu punktu TP2(1.5,1) do równania $a + x = c$ otrzymujemy:

$$1.5 + 1 = 2.5 \notin [3,4],$$

tak więc punkt nie spełnia równania interwałowego $[1,3] + [\underline{x}, \overline{x}] = [3,4]$. Po wstawieniu współrzędnych punktu TP1(2.5,1) widzimy, że jest on prawidłowym rozwiązaniem równania:

$$2.5 + 1 = 3.5 \in [3,4].$$

W równaniach interwałowych, wszystkie znane interwały tworzą hiper-prostokątną granulę informacyjną. W przypadku równania (3.6) jest to dwuwymiarowa granula $[1,3] \times$

$[3, 4]$. Rozwiązaniem równania jest także dwuwymiarowa granula $[\underline{x} = 3 - a, \overline{x} = 4 - a]$, pokazana na rys. 3.7. Wielowymiarowe, nieregularne, nie hiper-prostokątne granule rozwiązań nie mogą być uproszczone do jednowymiarowego interwału $[\underline{x}, \overline{x}]$. Interwał taki może być jedynie informacją o rozpiętości wielowymiarowego rozwiązania – uproszczoną informacją o rozwiązaniu pełnym. W przypadku równania (3.6) interwał ten ma postać: $[\underline{x}, \overline{x}] = [0, 3]$.

3.3. Odległość pomiędzy interwałami

Przy wyznaczaniu lokalnych modeli regresji, pojęcie odległości pomiędzy danymi wykorzystywane jest do dwóch zadań:

- po pierwsze, musimy mieć narzędzie do wyznaczania najbliższych sąsiadów;
- po drugie, określenie odległości niezbędne będzie przy opracowaniu zależności dla interwałowej wersji metody najmniejszych kwadratów (podrozdział 3.4).

W literaturze opisano kilka sposobów wyznaczania odległości pomiędzy interwałami. Można wykorzystać metrykę Hausdorffa, zaadaptowaną do przestrzeni $\mathbb{IR}$ przez R. Moore'a [40, 74]. Dla interwałów $A = [\underline{a}, \overline{a}]$ i $B = [\underline{b}, \overline{b}]$ odległość można wyznaczyć jako:

$$d_H(A, B) = \max(|\underline{a} - \underline{b}|, |\overline{a} - \overline{b}|). \tag{3.13}$$

Wybieramy tu zatem większą z odległości pomiędzy dolnymi i górnymi ograniczeniami interwałów.

Inny sposób, bazujący na zależnościach podanych w artykule [38], definiuje odległość z wykorzystaniem dwóch pomocniczych parametrów: $p \in [1, \infty)$ i $q \in [0, 1]$, jako:

$$d_{p,q}(A, B) = \sqrt[p]{(1 - q) \cdot |\underline{a} - \underline{b}|^p + q \cdot |\overline{a} - \overline{b}|^p}. \tag{3.14}$$

Parametr q charakteryzuje wagi przypisane do brzegów interwału. Jeśli nie ma powodu do wyróżnienia żadnego z nich (a tak jest najczęściej) powinien on przyjmować wartość $q = 0.5$.

W dalszej części monografii wykorzystana będzie metryka (3.14) z parametrem $p = 2$:

$$d_2(A, B) = d_{p=2,q=0.5}(A, B) = \sqrt{0.5 \cdot (\underline{a} - \underline{b})^2 + 0.5 \cdot (\overline{a} - \overline{b})^2}. \tag{3.15}$$

Jest to spowodowane większą czułością metryki d_2 od metryki d_H na zmiany położenia obu brzegów interwałów. Przykładowo:

$$d_H([1,2],[3,6]) = \max(2,4) = 4,$$
$$d_H([1,2],[2,6]) = \max(1,4) = 4,$$
$$d_2([1,2],[3,6]) = \sqrt{0.5 \cdot 2^2 + 0.5 \cdot 4^2} = \sqrt{10},$$
$$d_2([1,2],[2,6]) = \sqrt{0.5 \cdot 1^2 + 0.5 \cdot 4^2} = \sqrt{8.5}.$$

Dodatkowo, z wykorzystaniem metryki d_2 można dużo skuteczniej opracować zależności dla interwałowej wersji metody najmniejszych kwadratów.

Należy nadmienić, że zarówno metryka d_H, jak i d_2 spełniają dla dowolnych interwałów A, B i C niezbędne warunki:

- nieujemności: $s(A,B) \geqslant 0$,
- identyczności: $d(A,B) = 0 \iff A = B$,
- symetrii: $d(A,B) = d(B,A)$,
- nierówności trójkąta: $d(A,B) \leqslant d(A,C) + d(C,B)$,

stąd pary $(\mathbb{IR}, d_H)$ i $(\mathbb{IR}, d_2)$ są przestrzeniami metrycznymi [38, 74].

3.4. Interwałowa wersja metody najmniejszych kwadratów

Zadanie interwałowej regresji liniowej może być rozwiązane na wiele różnych sposobów. W jednym z podejść, poszukuje się regresji, która stara się pokryć całkowicie lub w części obserwowane dane [39, 40, 41, 146]. Dla sposobu tego bardzo ważne są prace prof. S. Sharego, dotyczące między innymi rozwiązywania równań interwałowych [120, 121, 122, 123]. Opisano w nich metodykę wyznaczania rozwiązania oraz wyróżniono kilka typów możliwych rozwiązań. Mankamentem tego typu podejścia jest przede wszystkim duża wrażliwość na występowanie w danych punktów odstających (ang. *outliers*) [39, 146] i tym samym możliwa spora szerokość estymowanych wartości. Cechuje się ono także dużą złożonością obliczeniową.

Zupełnie inny sposób wyznaczania regresji opisany jest w pracach [11, 22]. Wychodząc z założenia, że dla każdej danej w postaci interwału rozkład prawdopodobieństwa wystąpienia możliwych wartości ma charakter jednostajny, autorzy zaproponowali poszukiwanie regresji dla środków interwałów. Podejście to ulepszone zostało w artykule [65]. Autorzy poszukiwali w nim dwóch osobnych modeli regresji: jeden tworzony był dla środków interwałów, a drugi modelował szerokość estymaty. Efektem takiego rozwiązania była zmienna szerokość modelowanych wartości, co w skrajnych przypadkach (szczególnie w przypadku ekstrapolacji) mogło nawet skutkować zamianą dolnego i górnego brzegu estymowanego interwału wynikowego. Aby nie dopuścić do takiej sytuacji, ci sami autorzy

opisali kolejną modyfikację swojego rozwiązania [66], nakładając na wyznaczaną regresję dodatkowe warunki.

W rozwiązaniu zaprezentowanym poniżej, zastosowano uproszczony model z parametrami rzeczywistymi wyznaczany dla środków interwałów, a szerokość estymowanego interwału wynikowego wyznaczana jest jako uśredniona szerokość interwałów opisujących wyjście zadane. Tego typu podejście jest proste obliczeniowo i daje dobre, wiarygodne rozwiązania.

3.4.1. Model z jednym wejściem

Interwałowa wersja metody najmniejszych kwadratów zostanie zaprezentowana w pierwszej kolejności dla modelu z jednym wejściem. Załóżmy, że posiadamy dane w postaci par (X_i, Y_i), $i = 1 \dots K$, przy czym zarówno wejście $X_i = [\underline{X_i}, \overline{X_i}]$, jak i wyjście $Y_i = [\underline{Y_i}, \overline{Y_i}]$ ma postać interwału. W przypadku, w którym jakaś wartość określona jest w postaci liczby rzeczywistej, można zastąpić ją interwałem zdegenerowanym, posiadającym taką samą wartość dolnego i górnego ograniczenia. Jednym z ważniejszych założeń zaproponowanej poniżej metody jest jej skuteczna praca zarówno z danymi w postaci interwałów, jak i liczb rzeczywistych.

Będziemy poszukiwali liniowego modelu aproksymującego w postaci:

$$\hat{Y} = w_1 \cdot x_s + w_0 + R, \tag{3.16}$$

gdzie: $x_s = 0.5 \cdot (\underline{X} + \overline{X})$ oznacza środek interwału X, będącego wejściem modelu, R – interwał w postaci $[-r, r]$, w_1 i w_0 – rzeczywiste współczynniki modelu.

Będziemy chcieli, aby model dopasowany był do danych w taki sposób, aby minimalizować kryterium jakości oparte na metryce d_2 opisanej zależnością (3.15):

$$Q(w_1, w_0, r) = \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) \longrightarrow \min, \tag{3.17}$$

gdzie: $\hat{Y}_i = w_1 \cdot x_{si} + w_0 + R$.

Twierdzenie 3.1. Dla modelu aproksymującego w postaci (3.16) współczynniki minimalizujące kryterium jakości (3.17) można obliczyć z zależności:

$$w_1 = \frac{K \cdot \sum_{i=1}^{K} x_{si} \cdot \frac{\underline{Y_i} + \overline{Y_i}}{2} - \sum_{i=1}^{K} x_{si} \cdot \sum_{i=1}^{K} \frac{\underline{Y_i} + \overline{Y_i}}{2}}{K \cdot \sum_{i=1}^{K} x_{si}^2 - \left(\sum_{i=1}^{K} x_{si}\right)^2}, \tag{3.18}$$

$$w_0 = \frac{1}{K} \cdot \left(\sum_{i=1}^{K} \frac{\underline{Y_i} + \overline{Y_i}}{2} - w_1 \cdot \sum_{i=1}^{K} x_{si} \right),$$

$$r = \frac{1}{K} \cdot \sum_{i=1}^{K} \frac{\overline{Y_i} - \underline{Y_i}}{2}.$$

Dowód

Kryterium jakości (3.17) można inaczej zapisać w postaci:

$$\begin{aligned} Q(w_1, w_0, r) &= \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) = \sum_{i=1}^{K} d_2^2(w_1 \cdot x_{si} + w_0 + R, Y_i) \\ &= \sum_{i=1}^{K} \frac{1}{2} \left(w_1 \cdot x_{si} + w_0 - r - \underline{Y_i} \right)^2 + \frac{1}{2} \left(w_1 \cdot x_{si} + w_0 + r - \overline{Y_i} \right)^2. \end{aligned}$$

Aby wyznaczyć minimum, liczymy pochodne cząstkowe po parametrach modelu i przyrównujemy je do zera.

$$\begin{aligned} \frac{\partial Q(w_1, w_0, r)}{\partial w_1} &= \sum_{i=1}^{K} \left(w_1 \cdot x_{si} + w_0 - r - \underline{Y_i} \right) \cdot x_{si} + \left(w_1 \cdot x_{si} + w_0 + r - \overline{Y_i} \right) \cdot x_{si} \\ &= \sum_{i=1}^{K} w_1 x_{si}^2 + w_0 x_{si} - r x_{si} - \underline{Y_i} x_{si} + w_1 x_{si}^2 + w_0 x_{si} + r x_{si} - \overline{Y_i} x_{si} \\ &= \sum_{i=1}^{K} 2 w_1 x_{si}^2 + 2 w_0 x_{si} - x_{si}(\underline{Y_i} + \overline{Y_i}) = 0 \\ \frac{\partial Q(w_1, w_0, r)}{\partial w_0} &= \sum_{i=1}^{K} \left(w_1 \cdot x_{si} + w_0 - r - \underline{Y_i} \right) + \left(w_1 \cdot x_{si} + w_0 + r - \overline{Y_i} \right) \\ &= \sum_{i=1}^{K} 2 w_1 x_{si} + 2 w_0 - (\underline{Y_i} + \overline{Y_i}) = 0 \\ \frac{\partial Q(w_1, w_0, r)}{\partial r} &= \sum_{i=1}^{K} \left(w_1 \cdot x_{si} + w_0 - r - \underline{Y_i} \right) \cdot (-1) + \left(w_1 \cdot x_{si} + w_0 + r - \overline{Y_i} \right) \\ &= \sum_{i=1}^{K} 2r + \underline{Y_i} - \overline{Y_i} = 0 \end{aligned}$$

Po przekształceniach otrzymujemy ostatecznie układ 3 równań z trzema niewiadomymi parametrami:

$$\begin{aligned} w_1 \sum_{i=1}^{K} x_{si}^2 + w_0 \sum_{i=1}^{K} x_{si} &= \sum_{i=1}^{K} x_{si} \frac{\underline{Y_i} + \overline{Y_i}}{2}, \\ w_1 \sum_{i=1}^{K} x_{si} + w_0 K &= \sum_{i=1}^{K} \frac{\underline{Y_i} + \overline{Y_i}}{2}, \\ 2Kr &= \overline{Y_i} - \underline{Y_i}, \end{aligned}$$

po rozwiązaniu którego otrzymujemy współczynniki opisane zależnościami (3.18).

Jak dotąd, rozpatrzony został jedynie warunek konieczny na istnienie ekstremum kryterium jakości (3.17). Należy dalej zbadać, czy znalezione rozwiązanie faktycznie określa jego minimum. Warunkiem dostatecznym, aby funkcja wielu zmiennych posiadała minimum, jest dodatnia określoność Hessianu funkcji w wyznaczonym punkcie.

Kryterium jakości (3.17) w skrócie oznaczymy jako Q. Wówczas:

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 Q}{\partial w_1^2} & \dfrac{\partial^2 Q}{\partial w_1 \partial w_0} & \dfrac{\partial^2 Q}{\partial w_1 \partial r} \\ \dfrac{\partial^2 Q}{\partial w_0 \partial w_1} & \dfrac{\partial^2 Q}{\partial w_0^2} & \dfrac{\partial^2 Q}{\partial w_0 \partial r} \\ \dfrac{\partial^2 Q}{\partial r \partial w_1} & \dfrac{\partial^2 Q}{\partial r \partial w_0} & \dfrac{\partial^2 Q}{\partial r^2} \end{bmatrix} = \begin{bmatrix} 2\sum_{i=1}^{K} x_{si}^2 & 2\sum_{i=1}^{K} x_{si} & 0 \\ 2\sum_{i=1}^{K} x_{si} & 2 \cdot K & 0 \\ 0 & 0 & 2 \cdot K \end{bmatrix}. \tag{3.19}$$

Badamy dalej wyznaczniki macierzy.

$$\Delta_1 = 2\sum_{i=1}^{K} x_{si}^2 > 0$$

$$\begin{aligned} \Delta_2 &= 2K \cdot 2\sum_{i=1}^{K} x_{si}^2 - \left(2\sum_{i=1}^{K} x_{si}\right)^2 = 4\left[K\sum_{i=1}^{K} x_{si}^2 - \left(\sum_{i=1}^{K} x_{si}\right)^2\right] \\ &= 4\left[K\sum_{i=1}^{K} x_{si}^2 - \left(K \cdot \frac{\sum_{i=1}^{K} x_{si}}{K}\right)^2\right] = 4\left[K\sum_{i=1}^{K} x_{si}^2 - K^2 \cdot \overline{x}_s^2\right] \\ &= 4K\left[\sum_{i=1}^{K} x_{si}^2 - K \cdot \overline{x}_s^2\right] = 4K \cdot \sum_{i=1}^{K} \left(x_{si} - \overline{x}_s\right)^2 > 0 \end{aligned}$$

$$\Delta_3 = \Delta_2 \cdot 2K > 0$$

W powyższych wzorach $\overline{x}_s$ oznacza średnią arytmetyczną wartości x_{si}, $i = 1 \dots K$. Przy określaniu znaku wyznacznika Δ_2 skorzystano z zależności:

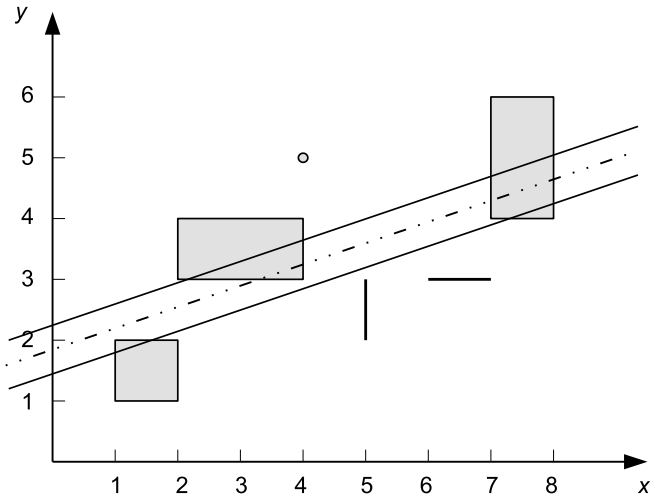
$$\begin{aligned}\sum_{i=1}^{K}(x_{si}-\overline{x}_s) &= \sum_{i=1}^{K}\left(x_{si}^2+\overline{x}_s^2-2x_{si}\overline{x}_s\right) = \sum_{i=1}^{K}x_{si}^2+K\overline{x}_s^2-2\overline{x}_s\sum_{i=1}^{K}x_{si}\\ &= \sum_{i=1}^{K}x_{si}^2+K\overline{x}_s^2-2\overline{x}_s\cdot K\frac{\sum_{i=1}^{K}x_{si}}{K} = \sum_{i=1}^{K}x_{si}^2+K\overline{x}_s^2-2K\overline{x}_s^2\\ &= \sum_{i=1}^{K}x_{si}^2-K\overline{x}_s^2.\end{aligned}$$

Ponieważ Hessian jest dodatnio określony, kryterium (3.17) w punkcie opisanym równaniami (3.18) posiada minimum.

□

Przykład 3.1.

Dla danych z tabeli zaprezentowanych na rys. 3.8 należy wyznaczyć model aproksymujący w postaci (3.16), minimalizujący kryterium jakości (3.17).



X	Y
[1,2]	[1,2]
[2,4]	[3,4]
5	[2,3]
4	5
[6,7]	3
[7,8]	[4,6]

Rysunek 3.8. Widok danych wykorzystanych w przykładzie 3.1 i wyznaczone rozwiązanie

Parametry wyznaczone za pomocą zależności (3.18) mają wartość:

$$w_1 \approx 0.366, \qquad w_0 \approx 1.879, \qquad r \approx 0.417.$$

Ponieważ poszukiwany model daje odpowiedź w postaci interwału, pokazana na rys. 3.8 charakterystyka ma postać pasa. Jego szerokość odpowiada średniej szerokości interwałów definiujących zadane odpowiedzi modelu.

3.4.2. Model z wieloma wejściami

Załóżmy, że zmienna zależna y zależy tym razem od N zmiennych objaśniających x_i, $i = 1 \dots N$. Będziemy poszukiwali liniowego modelu aproksymującego w postaci:

$$\hat{Y} = \mathbf{w}^T \cdot \mathbf{x_s} + R, \tag{3.20}$$

gdzie: $\mathbf{x_s} = [1, x_{1s}, \dots x_{Ns}]^T$ oznacza wektor środków interwałów X_i, będących wejściami modelu, R – interwał w postaci $[-r, r]$, $\mathbf{w} = [w_0, w_1, \dots w_N]^T$ – wektor rzeczywistych współczynników modelu.

Model utworzony zostanie na podstawie K danych pomiarowych, przy czym zarówno dane wejściowe (zmienne objaśniające) X_{ij}, jak i wyjściowe (zadane wartości zmiennej zależnej) Y_i mają postać interwałów bądź mogą być do takiej postaci sprowadzone ($i = 1 \dots N$, $j = 1 \dots K$). Dane opisane będą macierzą $\mathbf{X}$ i wektorem $\mathbf{Y}$:

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{N1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1K} & X_{2K} & \dots & X_{NK} \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_K \end{bmatrix}. \tag{3.21}$$

Będziemy chcieli, aby model dopasowany był do danych w taki sposób, aby minimalizować kryterium jakości oparte na metryce d_2 opisanej zależnością (3.15):

$$Q(\mathbf{w}, r) = \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) \longrightarrow \min, \tag{3.22}$$

gdzie: $\hat{Y}_i = \mathbf{w}^T \cdot \mathbf{x_{si}} + R$. Oznaczmy przez $\mathbf{X_s}$ macierz środków interwałów będących elementami macierzy $\mathbf{X}$.

Twierdzenie 3.2. Dla modelu aproksymującego w postaci (3.20) współczynniki minimalizujące kryterium jakości (3.22) można obliczyć z zależności:

$$\mathbf{w} = (\mathbf{X_s}^T \mathbf{X_s})^{-1} \mathbf{X_s}^T \cdot \frac{\underline{\mathbf{Y}} + \overline{\mathbf{Y}}}{2}, \tag{3.23}$$

$$r = \frac{1}{K} \cdot \sum_{i=1}^{K} \frac{\overline{Y_i} - \underline{Y_i}}{2}.$$

Dowód

Kryterium jakości (3.22) można inaczej zapisać w postaci:

$$Q(\mathbf{w}, r) = \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) = \sum_{i=1}^{K} d_2^2(\mathbf{w}^T \cdot \mathbf{x_{si}} + R, Y_i)$$
$$= \sum_{i=1}^{K} \frac{1}{2}\left(\mathbf{w}^T \cdot \mathbf{x_{si}} - r - \underline{Y_i}\right)^2 + \frac{1}{2}\left(\mathbf{w}^T \cdot \mathbf{x_{si}} + r - \overline{Y_i}\right)^2.$$

W zapisie macierzowym będzie ono miało postać:

$$Q(\mathbf{w}, r) = \frac{1}{2}(\mathbf{X_s w} - \mathbf{r} - \underline{\mathbf{Y}})^T(\mathbf{X_s w} - \mathbf{r} - \underline{\mathbf{Y}}) + \frac{1}{2}(\mathbf{X_s w} + \mathbf{r} - \overline{\mathbf{Y}})^T(\mathbf{X_s w} + \mathbf{r} - \overline{\mathbf{Y}}), \tag{3.24}$$

gdzie: $\underline{\mathbf{Y}}$, $\overline{\mathbf{Y}}$ – wektory dolnych i górnych ograniczeń wektora interwałów $\mathbf{Y}$, $\mathbf{r}$ – kolumnowy wektor K takich samych wartości r.

Po wymnożeniu i uproszczeniu wyrażenia otrzymujemy:

$$\begin{aligned} Q(\mathbf{w}, r) = \frac{1}{2} \; & (\mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w} - \mathbf{r}^T \mathbf{X_s w} - \underline{\mathbf{Y}}^T \mathbf{X_s w} - \mathbf{w}^T \mathbf{X_s}^T \mathbf{r} + \mathbf{r}^T \mathbf{r} + \underline{\mathbf{Y}}^T \mathbf{r} \\ & - \mathbf{w}^T \mathbf{X_s}^T \underline{\mathbf{Y}} + \mathbf{r}^T \underline{\mathbf{Y}} + \underline{\mathbf{Y}}^T \underline{\mathbf{Y}} + \mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w} + \mathbf{r}^T \mathbf{X_s w} - \overline{\mathbf{Y}}^T \mathbf{X_s w} \\ & + \mathbf{w}^T \mathbf{X_s}^T \mathbf{r} + \mathbf{r}^T \mathbf{r} - \overline{\mathbf{Y}}^T \mathbf{r} - \mathbf{w}^T \mathbf{X_s}^T \overline{\mathbf{Y}} - \mathbf{r}^T \overline{\mathbf{Y}} + \overline{\mathbf{Y}}^T \overline{\mathbf{Y}}) \\ = \frac{1}{2} \; & [2\mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w} - (\underline{\mathbf{Y}}^T + \overline{\mathbf{Y}}^T)\mathbf{X_s w} + 2\mathbf{r}^T \mathbf{r} - (\overline{\mathbf{Y}}^T - \underline{\mathbf{Y}}^T)\mathbf{r} \\ & - \mathbf{X_s}^T \mathbf{w}^T(\underline{\mathbf{Y}} + \overline{\mathbf{Y}}) - \mathbf{r}^T(\overline{\mathbf{Y}} - \underline{\mathbf{Y}}) + \underline{\mathbf{Y}}^T \underline{\mathbf{Y}} + \overline{\mathbf{Y}}^T \overline{\mathbf{Y}}] \\ = \frac{1}{2} \; & [2\mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w} - 2(\underline{\mathbf{Y}}^T + \overline{\mathbf{Y}}^T)\mathbf{X_s w} + 2\mathbf{r}^T \mathbf{r} - 2(\overline{\mathbf{Y}}^T - \underline{\mathbf{Y}}^T)\mathbf{r} + \underline{\mathbf{Y}}^T \underline{\mathbf{Y}} + \overline{\mathbf{Y}}^T \overline{\mathbf{Y}}]. \end{aligned}$$

Aby wyznaczyć minimum, liczymy pochodne cząstkowe po parametrach modelu i przyrównujemy je do zera.

$$\begin{aligned} \frac{\partial Q(\mathbf{w}, r)}{\partial \mathbf{w}} &= \frac{\partial \mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w}}{\partial \mathbf{w}} - \frac{\partial (\underline{\mathbf{Y}}^T + \overline{\mathbf{Y}}^T)\mathbf{X_s w}}{\partial \mathbf{w}} \\ &= 2\mathbf{X_s}^T \mathbf{X_s w} - [(\underline{\mathbf{Y}}^T + \overline{\mathbf{Y}}^T)\mathbf{X_s}]^T \\ &= 2\mathbf{X_s}^T \mathbf{X_s w} - \mathbf{X_s}^T(\underline{\mathbf{Y}} + \overline{\mathbf{Y}}) = 0 \end{aligned}$$

Możemy teraz wyznaczyć wektor parametrów:

$$\mathbf{w} = (\mathbf{X_s}^T \mathbf{X_s})^{-1} \mathbf{X_s}^T \cdot \frac{\underline{\mathbf{Y}} + \overline{\mathbf{Y}}}{2}.$$

Kryterium jakości można inaczej zapisać w postaci:

$$Q(\mathbf{w}, r) = \frac{1}{2}\,[2\mathbf{w}^T {\mathbf{X_s}}^T \mathbf{X_s}\mathbf{w} - 2(\underline{\mathbf{Y}}^T + \overline{\mathbf{Y}}^T)\mathbf{X}\mathbf{w} + 2Kr^2 - 2\sum_{i=1}^{K}(\overline{Y_i} - \underline{Y_i}) \cdot r + \underline{\mathbf{Y}}^T\underline{\mathbf{Y}} + \overline{\mathbf{Y}}^T\overline{\mathbf{Y}}].$$

Liczymy kolejną pochodną cząstkową i przyrównujemy ją do zera.

$$\frac{\partial Q(\mathbf{w}, r)}{\partial r} = 2Kr - \sum_{i=1}^{K}(\overline{Y_i} - \underline{Y_i}) = 0$$

Otrzymujemy ostatecznie zależność na parametr r:

$$r = \frac{1}{K} \cdot \sum_{i=1}^{K} \frac{\overline{Y_i} - \underline{Y_i}}{2}.$$

Badamy dalej warunek dostateczny. Aby kryterium (3.22) miało w punkcie (3.23) minimum, jego Hessian w tym punkcie musi być dodatnio określony. Podobnie jak wcześniej, kryterium (3.22) oznaczymy w skrócie jako Q. Hessian będzie miał wtedy postać:

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 Q}{\partial \mathbf{w} \partial \mathbf{w}^T} & \dfrac{\partial^2 Q}{\partial \mathbf{w} \partial r} \\ \dfrac{\partial^2 Q}{\partial r \partial \mathbf{w}} & \dfrac{\partial^2 Q}{\partial r^2} \end{bmatrix} = \begin{bmatrix} 2{\mathbf{X_s}}^T \mathbf{X_s} & 0 \\ 0 & 2K \end{bmatrix}. \tag{3.25}$$

Obliczamy wyznaczniki macierzy.

$$\Delta_1 = 2{\mathbf{X_s}}^T \mathbf{X_s}$$

$$\Delta_2 = 4K{\mathbf{X_s}}^T \mathbf{X_s}$$

Macierz ${\mathbf{X_s}}^T \mathbf{X_s}$ jest dodatnio określona, jeśli tylko jest nieosobliwa, w związku z tym warunek dodatniej określoności Hessianu jest spełniony.

□

Przykład 3.2.

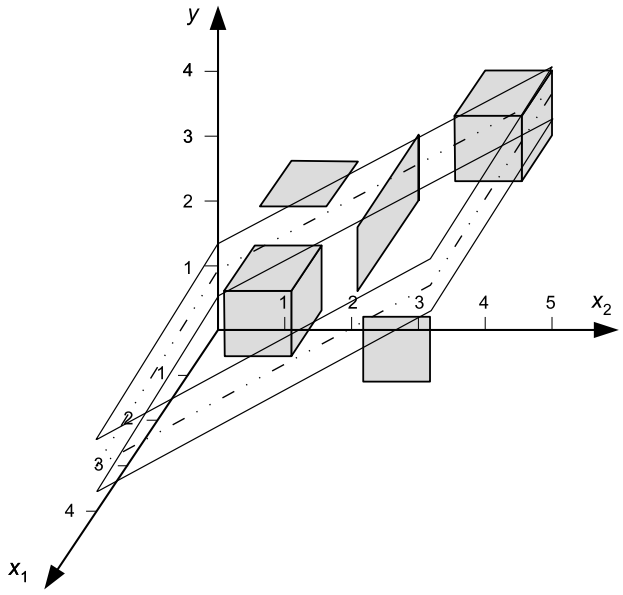
Dla danych z tabeli należy wyznaczyć model aproksymujący w postaci (3.20), minimalizujący kryterium jakości (3.22).

X_1	X_2	Y
[1,2]	[1,2]	[0,1]
4	[4,5]	[2,3]
[2,3]	[2,3]	4
[0,2]	3	[2,3]
[0,1]	[4,5]	[3,4]

Parametry wyznaczone za pomocą zależności (3.23) mają wartość:

$$\mathbf{w}^T \approx [0.9546, -0.0571, 0.5481], \qquad r \approx 0.417.$$

Ponieważ poszukiwany model daje odpowiedź w postaci interwału, pokazana na rys. 3.9 charakterystyka ma postać wycinka trójwymiarowej przestrzeni pomiędzy dwoma płaszczyznami. Jego szerokość odpowiada średniej szerokości interwałów definiujących zadane odpowiedzi modelu i jest równa $2 \cdot r$.



Rysunek 3.9. Widok danych wykorzystanych w przykładzie 3.2 i wyznaczone rozwiązanie

3.5. Modelowanie lokalne dla danych interwałowych

Sposób modelowania lokalnego, opisany w podrozdziale 2.1, będzie wymagał modyfikacji, ponieważ zarówno punkt wejściowy $\mathbf{x}^*$, jak i dane uczące będą interwałami.

W przypadku, gdy któraś z analizowanych wartości będzie liczbą rzeczywistą, zastąpimy ją interwałem zdegenerowanym, posiadającym taką samą wartość dolnego i górnego ograniczenia [108].

Aby wyznaczyć najbliższych sąsiadów punktu wejściowego $\mathbf{x}^*$, będziemy badać odległość między wielowymiarowymi danymi interwałowymi. Możemy ją obliczyć jako:

$$D(\mathbf{x}, \mathbf{x}^*) = \sqrt{\sum_{i=1}^{N} d_2^2(x_i, x_i^*)}, \tag{3.26}$$

gdzie: $\mathbf{x}, \mathbf{x}^*$ – wektory interwałów, N – rozmiar wektora danych, $d_2(x_i, x_i^*)$ – odległość między interwałami wyznaczona z zależności (3.15).

Modelowanie lokalne dla interwałów z wykorzystaniem metody kNN, może być zrealizowane identycznie jak dla danych rzeczywistych biorąc za podstawę zasady arytmetyki interwałowej, wzory (3.2) – (3.5). Można w ten sposób wyznaczyć średnią arytmetyczną bądź średnią ważoną z wyjść zadanych dla próbek będących sąsiadami punktu wejściowego $\mathbf{x}^*$.

W przypadku mini-modelu liniowego, będziemy poszukiwali regresji:

$$f(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{x_s}^* + R, \tag{3.27}$$

w sposób opisany w podrozdziale 3.4. Dla mini-modelu nieliniowego, regresja będzie miała postać:

$$f(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{x_s}^* + R + f_N(\mathbf{x_s}^*), \tag{3.28}$$

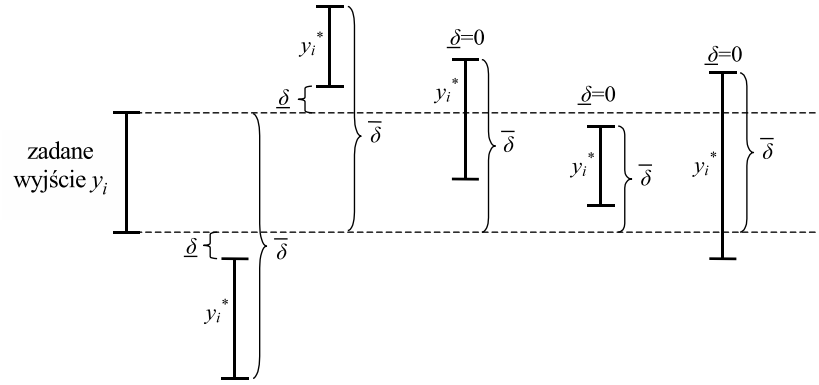
przy czym współczynnik składnika nieliniowego wyznaczany będzie dla środków interwałów danych uczących i punktu wejściowego $\mathbf{x}^*$.

3.5.1. Dokładność modelu

Załóżmy, że mamy do dyspozycji dwa znormalizowane, rozłączne zbiory danych. Pierwszy z nich będzie zawierał dane uczące (generujące modele lokalne), a każda próbka składać się będzie z wektora interwałów wejściowych $\mathbf{x}_j$ i zadanej interwałowej wartości wyjściowej y_j, $j = 1 \ldots L$. Drugi zbiór danych zawierał będzie dane walidujące i podobnie, każda próbka składać się będzie z wektora interwałów wejściowych $\mathbf{x}_i$ i zadanej interwałowej wartości wyjściowej y_i, $i = 1 \ldots M$.

Miarą dokładności modelu może być jego średni błąd bezwzględny, wyznaczony dla danych walidujących:

$$e_{MAE} = \frac{1}{M} \cdot \sum_{i=1}^{M} d_2(y_i, y_i^*), \tag{3.29}$$



Rysunek 3.10. Przykłady wyznaczania dolnego i górnego ograniczenia błędu modelu

gdzie: y_i – to wyjście zadane dla próbki i, a y_i^* – to wyjście wyznaczone przez model. Wyznaczony w ten sposób błąd będzie liczbą rzeczywistą.

Alternatywnie, możemy wziąć pod uwagę, że dane na których pracujemy są interwałami, stąd dokładność modelu również może być określona jako interwał [104, 108]. Różnicę interwałów y_i i y_i^* możemy obliczyć ze wzoru (3.3) jako:

$$d_i = y_i - y_i^* = [\underline{y_i} - \overline{y_i^*}, \overline{y_i} - \underline{y_i^*}] = [\underline{d_i}, \overline{d_i}]. \tag{3.30}$$

Błąd modelu możemy obliczyć jako interwał:

$$\delta_i = [\underline{\delta_i}, \overline{\delta_i}], \tag{3.31}$$

w którym dolne ograniczenie będzie wyznaczone za pomocą zależności:

$$\underline{\delta_i} = \begin{cases} 0 & \text{if } y_i \cap y_i^* \neq \emptyset \\ \min\{|\underline{d_i}|, |\overline{d_i}|\} & \text{otherwise,} \end{cases} \tag{3.32}$$

a górne jako:

$$\overline{\delta_i} = \max\{|\underline{d_i}|, |\overline{d_i}|\}. \tag{3.33}$$

Na rysunku 3.10 zilustrowano sposób wyznaczania dolnego i górnego ograniczenia błędu modelu.

Ostatecznie możemy wyznaczyć interwałowy, średni błąd bezwzględny danych walidujących jako:

$$e_{\mathrm{IMAE}} = \left[\underline{e_{\mathrm{IMAE}}}, \overline{e_{\mathrm{IMAE}}}\right], \tag{3.34}$$

gdzie:

$$\underline{e_{\mathrm{IMAE}}} = \frac{1}{M}\sum_{i=1}^{M}\underline{\delta_i} \quad \text{i} \quad \overline{e_{\mathrm{MAE}}} = \frac{1}{M}\sum_{i=1}^{M}\overline{\delta_i}. \tag{3.35}$$

Dodatkowo, możemy jeszcze wprowadzić pojęcie dokładności modelu w postaci:

$$q = \frac{1}{1+e_{\mathrm{MAE}}} = [\underline{q}, \overline{q}] = \left[\frac{1}{1+\overline{e_{\mathrm{MAE}}}}, \frac{1}{1+\underline{e_{\mathrm{MAE}}}}\right]. \tag{3.36}$$

Dokładność zdefiniowana w ten sposób przyjmuje wartości od 0 (dla błędu modelu dążącego do nieskończoności) do 1 (kiedy błąd spada do 0). Z definicji pojęcia wynika, że zawsze spełniona jest zależność: $\underline{q} \leqslant \overline{q}$.

Jeżeli nie jesteśmy w stanie wyodrębnić z posiadanych danych rozłącznej części próbek do walidacji jakości modelu, możemy zastosować kroswalidację (walidację krzyżową). Wyznaczanie błędu kroswalidacji i dokładności modelu realizowane jest podobnie jak dla danych walidujących na podstawie wzorów: (3.29) – (3.36). W dalszej części książki, do walidacji krzyżowej wykorzystana zostanie metoda minus jednego elementu. W metodzie tej z danych uczących usuwana jest jedna próbka, a reszta wykorzystywana jest do utworzenia modelu. Następnie, określa się błąd modelowania dla usuniętej próbki, po czym zabieg powtarza się dla wszystkich kolejnych. Wyznaczone błędy ostatecznie się uśrednia.

3.6. Eksperymenty

Badania nr 1 – dane jednowejściowe

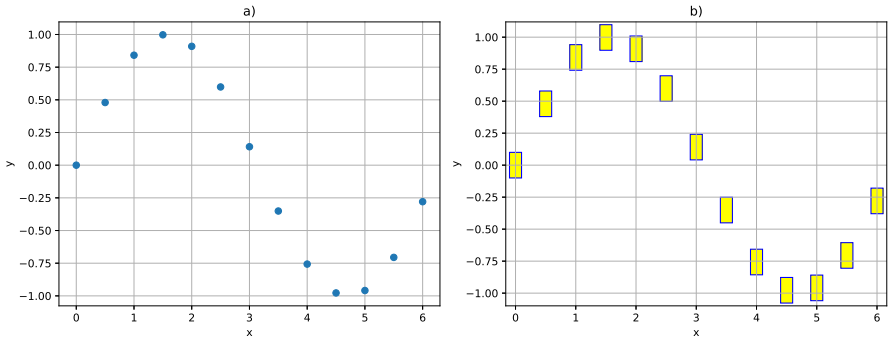
W pierwszym eksperymencie wykorzystane zostaną proste dane jednowejściowe. Jego celem będzie przede wszystkim zbadanie poprawności metod opisanych we wcześniejszych podrozdziałach.

Dane opisują sinusoidę próbkowaną co 0.5 dla wejścia $x \in [0, 2\pi]$ i składać się będą z 13 próbek pokazanych na rys. 3.11a. Dane poddane zostały interwałowemu rozszerzeniu o wartości $\alpha = 0.1$, przy czym rozszerzeniu podlegał zarówno atrybut wejściowy, jak i wyjściowy danych. Dane po rozszerzeniu pokazane zostały na rys. 3.11b.

W trakcie rozszerzania interwałowego, każda rzeczywista wartość x jest zastępowana interwałem $[\underline{x}, \overline{x}]$, w sposób opisany wzorem:

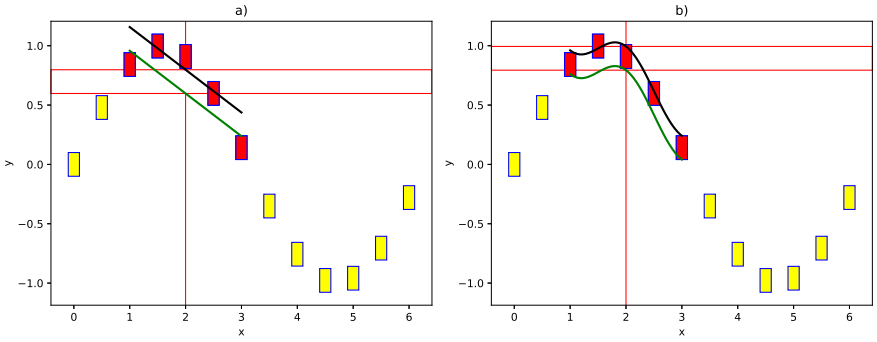
$$[\underline{x}, \overline{x}] = [x - \alpha, x + \alpha]. \tag{3.37}$$

Na rysunkach od 3.12 do 3.15 pokazano wykresy przykładowych interwałowych mini-modeli liniowych (a) i nieliniowych (b), wyznaczonych przy założeniu, że liczba uwzględnianych sąsiadów wynosi $k = 5$. Na każdym wykresie jest też zaznaczony interwał



Rysunek 3.11. Dane wykorzystane w badaniach nr 1 przed rozszerzeniem (a) i po rozszerzeniu interwałowym o wartości $\alpha = 0.1$ (b)

odpowiedzi modelu. Szerokość wszystkich interwałów wyjściowych jest równa dwukrotnej wartości rozszerzenia interwałowego $\alpha = 0.1$.



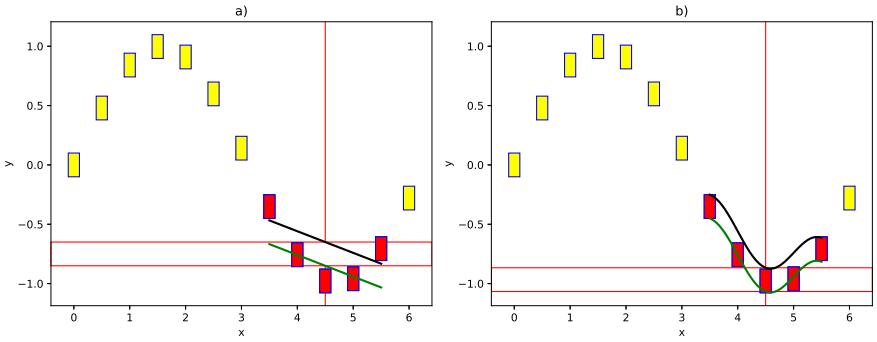
Rysunek 3.12. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla rzeczywistego punktu wejściowego $x^* = 2$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [0.598, 0.798]$ i nieliniowych $y \approx [0.794, 0.994]$

Na rysunku 3.16 pokazane zostały interwałowe charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Podobnie jak powyżej, liczba uwzględnianych sąsiadów wynosiła $k = 5$. W tabeli 3.2 podany jest dodatkowo średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej na podstawie wzorów: (3.29) i (3.34).

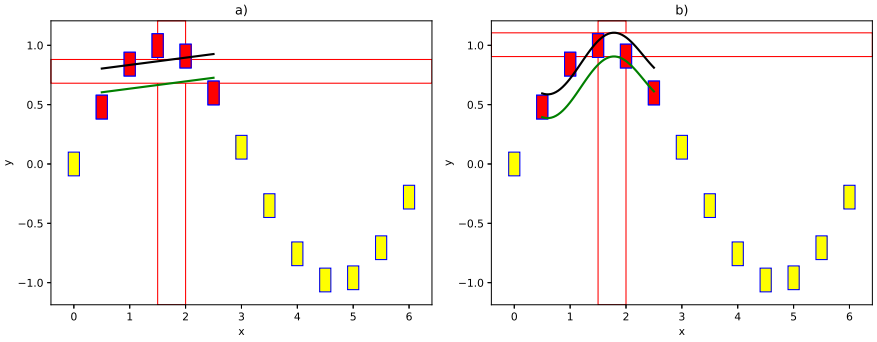
Na rysunku 3.16 uwagę zwraca najlepsze dopasowanie do danych, widoczne dla charakterystyki modelu bazującego na mini-modelach nieliniowych. Ma to również przełożenie na zdecydowanie najmniejszy błąd e_{MAE}.

Badania nr 2 – zaszumiona sinusoida

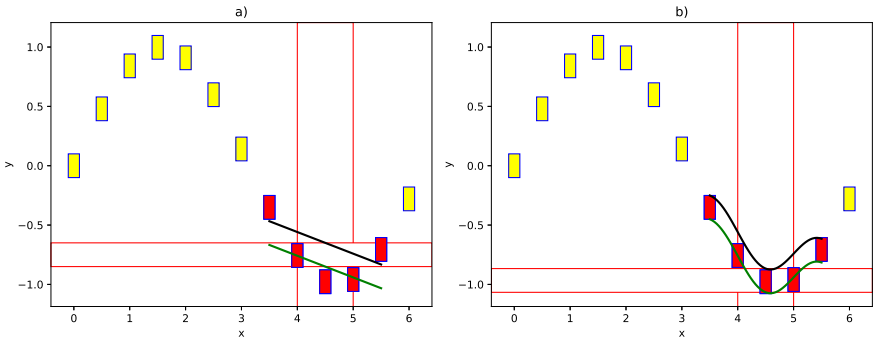
W drugim eksperymencie również wykorzystane zostaną dane jednowejściowe. Dane opisują równomiernie próbkowaną sinusoidę dla wejścia $x \in [0, 2\pi]$ i składać się będą ze 100 próbek, przy czym zarówno atrybut wejściowy, jak i wyjściowy danych, zaszumiony został losowymi wartościami z przedziału $[0, 0.2]$. Dane przedstawiono na rys. 3.17a. Próbki



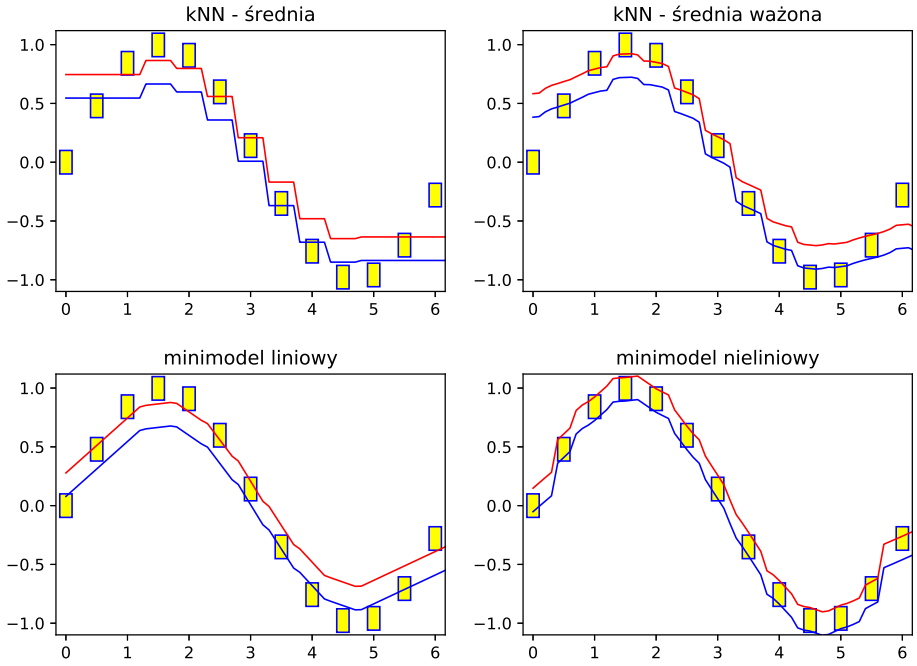
Rysunek 3.13. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla rzeczywistego punktu wejściowego $x^* = 4.5$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [-0.850, -0.650]$ i nieliniowych $y \approx [-1.067, -0.867]$



Rysunek 3.14. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla interwałowego punktu wejściowego $x^* = [1.5, 2]$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [0.681, 0.881]$ i nieliniowych $y \approx [0.905, 1.105]$



Rysunek 3.15. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla interwałowego punktu wejściowego $x^* = [4, 5]$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [-0.850, -0.650]$ i nieliniowych $y \approx [-1.067, -0.867]$

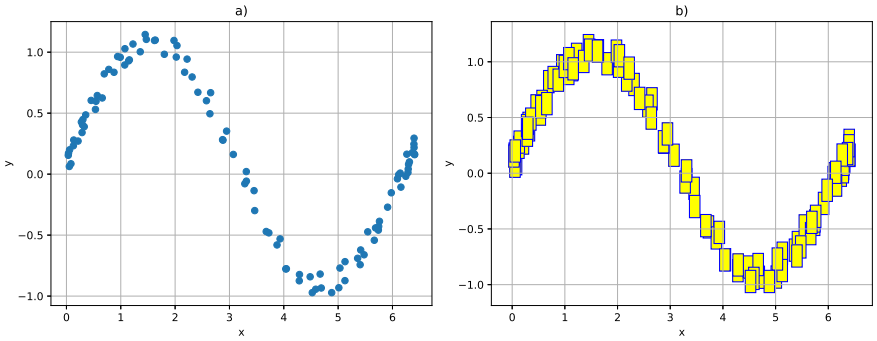


Rysunek 3.16. Interwałowe charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych

Tabela 3.2. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 1

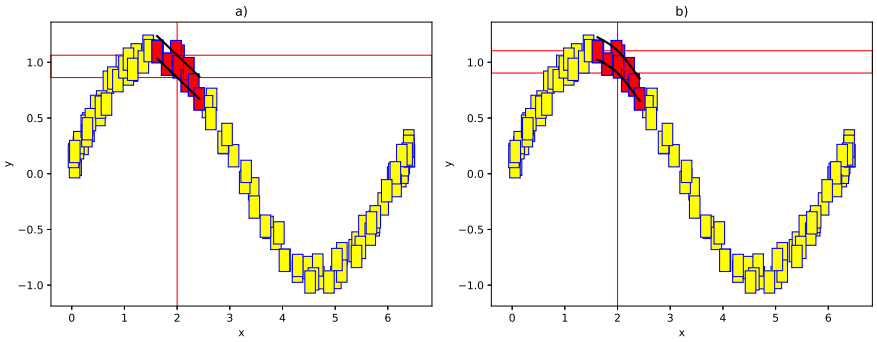
	e_{IMAE}	e_{MAE}
kNN – średnia arytmetyczna	[0.152,0.527]	0.327
kNN – średnia ważona	[0.126,0.494]	0.294
mini-model liniowy	[0.143,0.505]	0.305
mini-model nieliniowy	[0.061,0.349]	0.149

poddane zostały interwałowemu rozszerzeniu o wartości $\alpha = 0.1$, przy czym rozszerzeniu podlegał zarówno atrybut wejściowy, jak i wyjściowy danych. Dane po rozszerzeniu pokazane są na rys. 3.17b.



Rysunek 3.17. Dane wykorzystane w badaniach nr 2 przed rozszerzeniem (a) i po rozszerzeniu interwałowym o wartości $\alpha = 0.1$ (b)

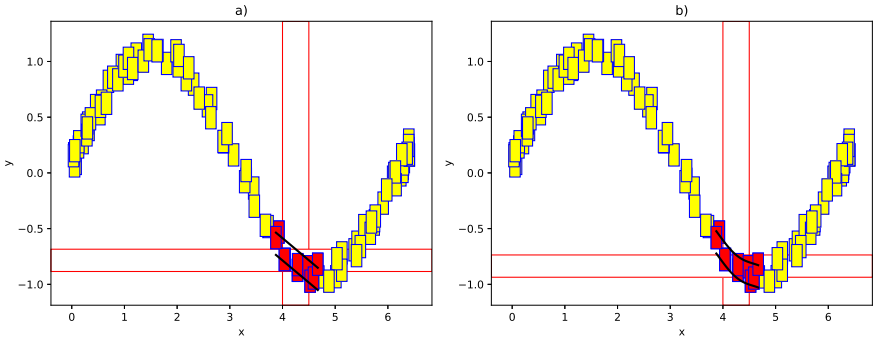
Na rysunkach 3.18 i 3.19 pokazano wykresy przykładowych interwałowych mini-modeli liniowych (a) i nieliniowych (b), wyznaczonych przy założeniu, że liczba uwzględnianych sąsiadów wynosi $k = 10$. Na każdym wykresie zaznaczono także interwał odpowiedzi modelu. Szerokość wszystkich interwałów wyjściowych jest równa dwukrotnej wartości rozszerzenia interwałowego $\alpha = 0.1$.



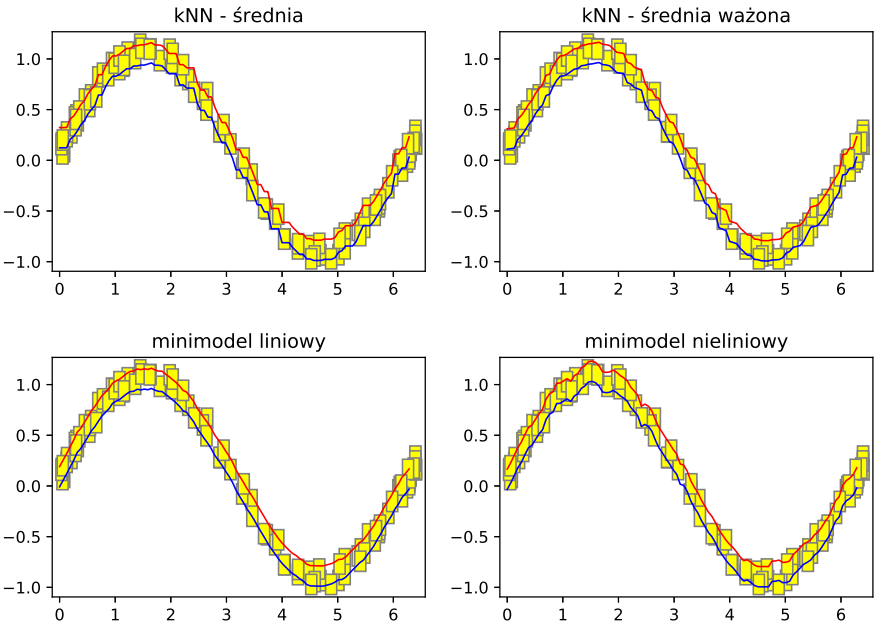
Rysunek 3.18. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla rzeczywistego punktu wejściowego $x^* = 2$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [0.863, 1.063]$ i nieliniowych $y \approx [0.904, 1.104]$

Na rysunku 3.20 pokazane są interwałowe charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Podobnie jak powyżej, liczba uwzględnianych sąsiadów wynosiła $k = 10$. W tabeli 3.3 podany jest dodatkowo średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej na podstawie wzorów: (3.29) i (3.34).

Na rysunku 3.21 zobrazowano jak zmienia się średni błąd bezwzględny modelu w zależności od przyjętej liczby najbliższych sąsiadów k. Wykres 3.21a sporządzono dla ważonej metody kNN, a wykres 3.21b dla modelu bazującego na mini-modelach liniowych.



Rysunek 3.19. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla interwałowego punktu wejściowego $x^* = [4.0, 4.5]$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [-0.885, -0.685]$ i nieliniowych $y \approx [-0.936, -0.736]$

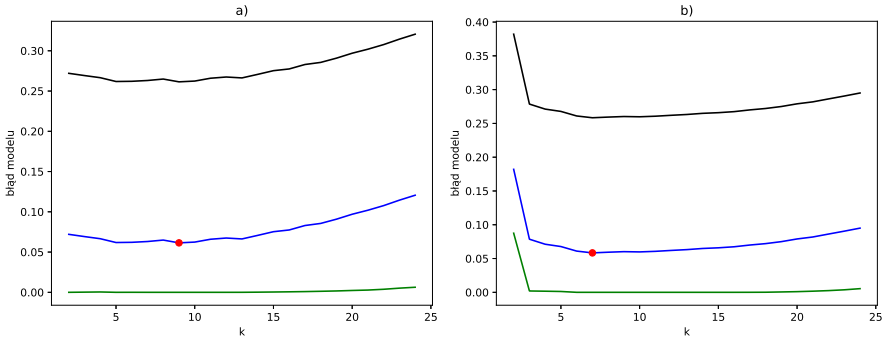


Rysunek 3.20. Interwałowe charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych

Tabela 3.3. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 2

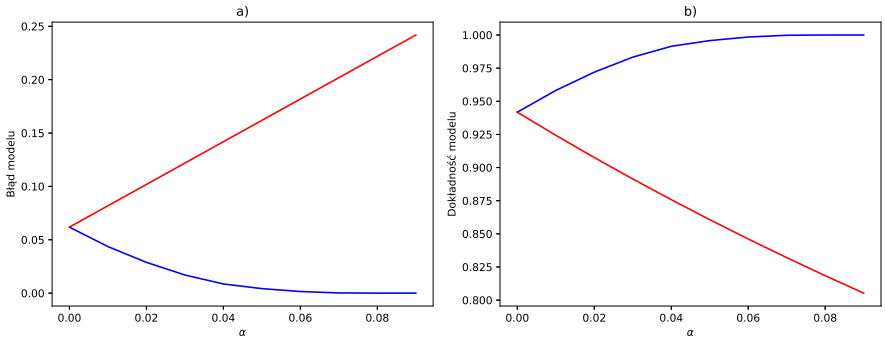
	e_{IMAE}	e_{MAE}
kNN – średnia arytmetyczna	[0.0,0.265]	0.065
kNN – średnia ważona	[0.0,0.262]	0.062
mini-model liniowy	[0.0,0.259]	0.059
mini-model nieliniowy	[0.0,0.261]	0.061

Z wykresów można odczytać optymalną (gwarantującą najmniejszy błąd rzeczywisty modelu) wartość $k = 9$ dla ważonej metody kNN i $k = 7$ dla mini-modeli liniowych. Na wykresach, dolna i górna linia to ograniczenia interwałowego, średniego błędu bezwzględnego e_{IMAE}, a linia środkowa to średni błąd bezwzględny e_{MAE}.



Rysunek 3.21. Średni błąd bezwzględny modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla ważonej metody kNN (a) i dla modelu bazującego na mini-modelach liniowych (b)

Na rysunku 3.22 pokazano jak zmienia się interwałowy, średni błąd bezwzględny i dokładność modelu wraz ze zmianą wielkości rozszerzenia interwałowego α od 0.0 do 0.1.



Rysunek 3.22. Interwałowy, średni błąd bezwzględny i dokładność modelu w zależności od wielkości rozszerzenia interwałowego α

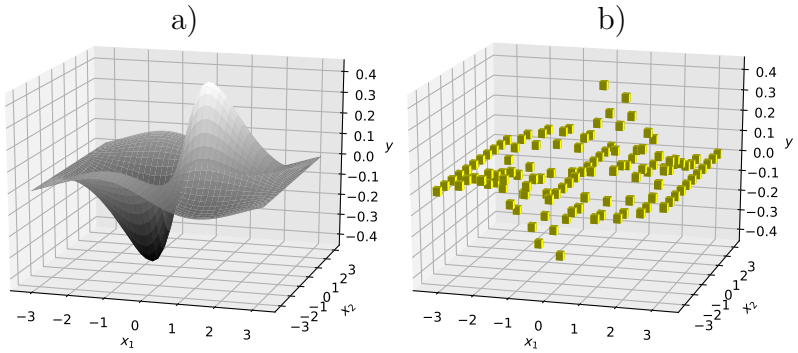
Badania nr 3 – dane 2-wejściowe

W eksperymencie nr 3 wykorzystane zostały dane 2-wejściowe. Umożliwiły one zbadanie, jak metoda zachowuje się dla danych o większej wymiarowości, przy jednoczesnej możliwości wizualizacji wyników. Dane wygenerowane zostały na podstawie funkcji:

$$y = \frac{\sin(x_1) \cdot \cos(x_2)}{x_1^2 + x_2^2 + 1}, \tag{3.38}$$

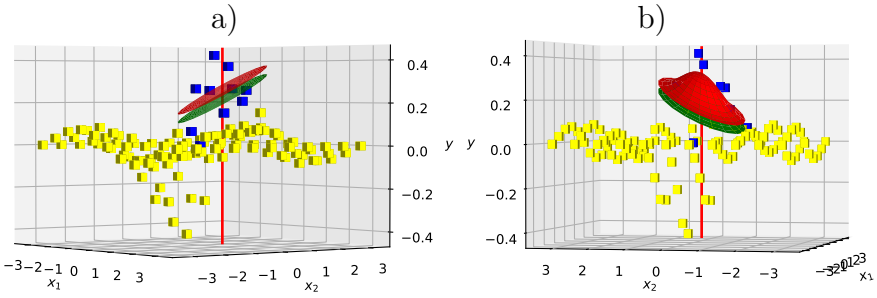
przy czym oba wejścia zmieniały się w przedziale $x_1, x_2 \in [-\pi, \pi]$ i próbkowane były co $\pi/5$. Dało do ostatecznie 121 próbek. Wykres wykorzystanej funkcji pokazany został na

rys. 3.23a. Dane poddane zostały interwałowemu rozszerzeniu o wartości $\alpha_x = 0.1$ dla zmiennych x_1 i x_2 oraz $\alpha_y = 0.02$ dla zmiennej wyjściowej y. Różne wartości rozszerzenia wynikały z różnych zakresów wartości jakie przyjmowały zmienne. Dane po rozszerzeniu pokazano na rys. 3.23b.



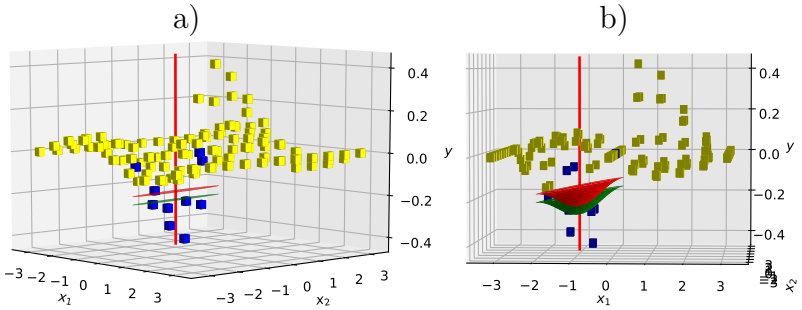
Rysunek 3.23. Wykres funkcji opisanej równaniem (3.38) (a) i widok danych po rozszerzeniu interwałowym (b)

Na rysunkach 3.24 i 3.25 przedstawiono wykresy przykładowych interwałowych mini- -modeli liniowych (a) i nieliniowych (b), wyznaczonych przy założeniu, że liczba uwzględnianych sąsiadów wynosi $k = 10$. Nie jest to wartość optymalna, jednak pozwala w poglądowy sposób pokazać kształt wynikowych mini-modeli. Na każdym wykresie zaznaczona została także linia określająca punkt zapytania oraz wyróżniono próbki sąsiednie.

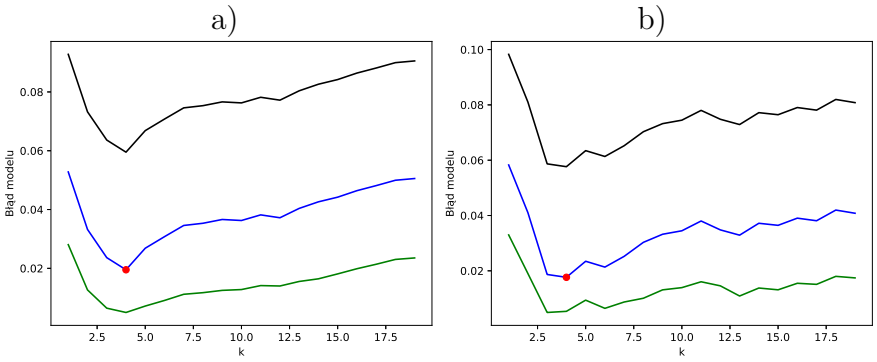


Rysunek 3.24. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla punktu wejściowego o współrzędnych $\mathbf{x}^* = (x_1^*, x_2^*) = (1.0, 0.0)$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [0.221, 0.261]$ i nieliniowych $y \approx [0.287, 0.327]$

Na rysunku 3.26 pokazano jak zmienia się średni błąd bezwzględny modelu w zależności od przyjętej liczby najbliższych sąsiadów k. Wykres 3.26a sporządzono dla ważonej metody kNN, a wykres 3.26b dla modelu bazującego na mini-modelach nieliniowych. Z wykresów można odczytać optymalną (gwarantującą najmniejszy błąd rzeczywisty modelu) wartość $k = 4$ dla obu metod. Na wykresach, dolna i górna linia to ograniczenia interwałowego, średniego błędu bezwzględnego e_{IMAE}, a linia środkowa to średni błąd bezwzględny e_{MAE}.

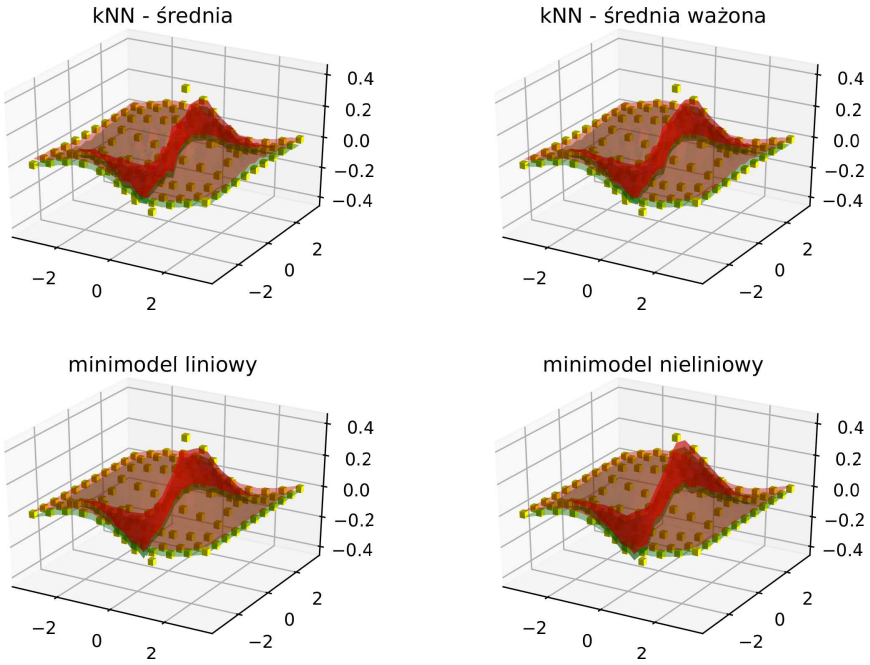


Rysunek 3.25. Mini-model liniowy (a) i nieliniowy (b) wyznaczony dla punktu wejściowego o współrzędnych $\mathbf{x}^* = (x_1^*, x_2^*) = (-1.0, 0.0)$. Odpowiedź modelu bazującego na mini-modelach: liniowych $y \approx [-0.239, -0.199]$ i nieliniowych $y \approx [-0.286, -0.246]$



Rysunek 3.26. Średni błąd bezwzględny modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla ważonej metody kNN (a) i dla modelu bazującego na mini-modelach nieliniowych (b)

Na rysunku 3.27 pokazano interwałowe charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Liczba uwzględnianych sąsiadów wynosiła $k = 4$, czyli w obliczeniach przyjęto wartość wyznaczoną jako optymalną. W tabeli 3.4 podany jest dodatkowo średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzorów (3.29) i (3.34).



Rysunek 3.27. Interwałowe charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych

Tabela 3.4. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 3

	e_{IMAE}	e_{MAE}
kNN – średnia arytmetyczna	[0.005,0.060]	0.020
kNN – średnia ważona	[0.005,0.059]	0.019
mini-model liniowy	[0.004,0.056]	0.016
mini-model nieliniowy	[0.005,0.057]	0.017

Badania nr 4 – porównanie dokładności modeli

W kolejnych eksperymentach porównana została dokładność modeli utworzonych z wykorzystaniem prostej i ważonej metody kNN oraz bazujących na mini-modelach liniowych i nieliniowych. W badaniach wykorzystano dane wygenerowane samodzielnie oraz pochodzące z popularnych repozytoriów internetowych. Atrybuty wejściowe dla wszystkich danych zostały znormalizowane, ponieważ zakresy przyjmowanych przez nie wartości często

mocno się różniły. Ponadto poddano je interwałowemu rozszerzeniu $\alpha_x = 0.1$. Atrybutów wyjściowych nie normalizowano, a przyjęte wartości rozszerzenia interwałowego α_y podane zostały w tabeli 3.5.

Tabela 3.5. Parametry danych użytych w badaniach

dane	liczba wejść	liczba próbek	α_y
$\sin(x)$ – krok próbkowania 0.5	1	13	0.1
$\sin(x)$ – krok próbkowania 0.2	1	32	0.1
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – krok próbkowania $\pi/10$	2	121	0.02
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – krok próbkowania $\pi/30$	2	961	0.02
`bodyfat`	14	252	2
`cpu`	6	209	10
`diabetes_numeric`	2	43	0.3
`elusage`	2	55	5

W tabeli 3.6 podano średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej na podstawie wzoru (3.29), a w tabeli 3.7 interwałowy, średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej oparty na wzorze (3.34). Dla każdych danych i dla każdej metody modelowania, błąd wyznaczony został dla uprzednio wyznaczonej, optymalnej liczby najbliższych sąsiadów k. Wartość błędu w tabelach zaokrąglona jest do 3 lub 4 miejsc po przecinku. Dodatkowo w tabeli 3.6, w nawiasie, podano wyznaczoną optymalną wartość k.

Tabela 3.6. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej

dane	kNN – średnia arytmetyczna	kNN – średnia ważona	mini-model liniowy	mini-model nieliniowy
$\sin(x)$ – kr. prób. 0.5	0.166 (**2**)	0.163 (**2**)	0.095 (**2**)	0.095 (**2**)
$\sin(x)$ – kr. prób. 0.2	0.0306 (**2**)	0.0299 (**2**)	0.0130 (**2**)	0.0086 (**5**)
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – kr. pr. $\pi/10$	0.0199 (**4**)	0.0195 (**4**)	0.0165 (**4**)	0.0176 (**4**)
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – kr. pr. $\pi/30$	0.0028 (**4**)	0.0028 (**4**)	0.0022 (**4**)	0.0024 (**4**)
`bodyfat`	2.486 (**3**)	2.475 (**3**)	0.449 (**39**)	0.455 (**38**)
`cpu`	28.089 (**2**)	27.463 (**6**)	25.484 (**31**)	25.875 (**31**)
`diabetes_numeric`	0.459 (**9**)	0.464 (**9**)	0.483 (**32**)	0.468 (**29**)
`elusage`	9.007 (**9**)	9.210 (**9**)	8.365 (**22**)	8.621 (**22**)

Tabela 3.7. Interwałowy, średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej

dane	kNN – średnia arytmetyczna	kNN – średnia ważona	mini-model liniowy	mini-model nieliniowy
$\sin(x)$ – kr. prób. 0.5	[0.063,0.366]	[0.059,0.363]	[0.0,0.295]	[0.0,0.295]
$\sin(x)$ – kr. prób. 0.2	[0.0057,0.2306]	[0.0050,0.2299]	[0.0,0.2130]	[0.0,0.2086]
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – kr. $\pi/10$	[0.0051,0.0599]	[0.0050,0.0595]	[0.0047,0.0565]	[0.0053,0.0576]
$\frac{\sin(x_1)*\cos(x_2)}{x_1^2+x_2^2+1}$ – kr. $\pi/30$	[0.0,0.0428]	[0.0,0.0428]	[0.0,0.0422]	[0.0,0.0424]
`bodyfat`	[0.343,6.486]	[0.337,6.475]	[0.080,4.449]	[0.081,4.455]
`cpu`	[16.098,48.088]	[15.634,47.462]	[13.542,45.484]	[13.719,45.874]
`diabetes_numeric`	[0.093,1.059]	[0.090,1.064]	[0.111,1.082]	[0.107,1.068]
`elusage`	[3.043,19.007]	[3.071,19.210]	[2.248,18.365]	[2.206,18.621]

3.7. Wnioski

Podsumowując wyniki eksperymentów możemy stwierdzić, że modele oparte na mini-modelach liniowych i nieliniowych zapewniły dokładność porównywalną bądź lepszą od modeli bazujących na metodach k najbliższych sąsiadów. Modele lokalne wyznaczane interwałową metodą najmniejszych kwadratów zapewniały dobre dopasowanie do danych, a szerokość wyznaczanych rozwiązań (zgodnie ze wzorem (3.23), s. 35) była zawsze uśrednioną szerokością wszystkich próbek sąsiednich, uwzględnianych w obliczeniach.

Czas wykonywanych obliczeń, w przypadku działań na interwałach, był tylko nieznacznie większy (około kilkanaście procent). W przypadku modeli bazujących na technice kNN wynikało to z prostoty arytmetyki interwałowej, natomiast w przypadku mini-modeli, regresja wyznaczana była dla środków interwałów, stąd większość obliczeń wykonywana była tu na liczbach rzeczywistych.

4. Liczby rozmyte

4.1. Pojęcia podstawowe

Kolejny poziom niepewności reprezentują liczby rozmyte, stanowiące szczególny przypadek zbiorów rozmytych. Podstawowe pojęcia związane z liczbami rozmytymi opisane są dobrze w wielu pozycjach literatury przedmiotu, np. [24, 46, 54, 86, 111]. Poniżej, dla przypomnienia, przedstawiono kilka najważniejszych definicji.

Rozmyty podzbiór zbioru liczb rzeczywistych $\mathbb{R}$, dla którego zdefiniowana jest funkcja przynależności $\mu : \mathbb{R} \to [0, 1]$, będziemy nazywali liczbą rozmytą A jeśli:

a) A jest normalna, czyli istnieje element $x_0 \in \mathbb{R}$, taki że $\mu(x_0) = 1$;

b) A jest wypukła, czyli $\mu(\lambda x + (1 - \lambda)y) \geqslant \mu(x) \wedge \mu(y)$, $\forall\, x, y \in \mathbb{R}$ i $0 \leqslant \lambda \leqslant 1$;

c) funkcja μ jest półciągła z góry;

d) nośnik $\mathrm{supp}(A)$ jest ograniczony.

Każdą liczbę rozmytą można przedstawić w postaci:

$$\mu(x) = \begin{cases} 0 & \text{dla} \quad x < a_1, \\ f(x) & \text{dla} \quad a_1 \leqslant x < a_2, \\ 1 & \text{dla} \quad a_2 \leqslant x < a_3, \\ g(x) & \text{dla} \quad a_3 \leqslant x < a_4, \\ 0 & \text{dla} \quad x \geqslant a_4, \end{cases} \tag{4.1}$$

gdzie: $a_1, a_2, a_3, a_4 \in \mathbb{R}$. $f(x)$ jest niemalejącą funkcją nazywaną lewym brzegiem, a $g(x)$ jest nierosnącą funkcją nazywaną prawym brzegiem liczby rozmytej.

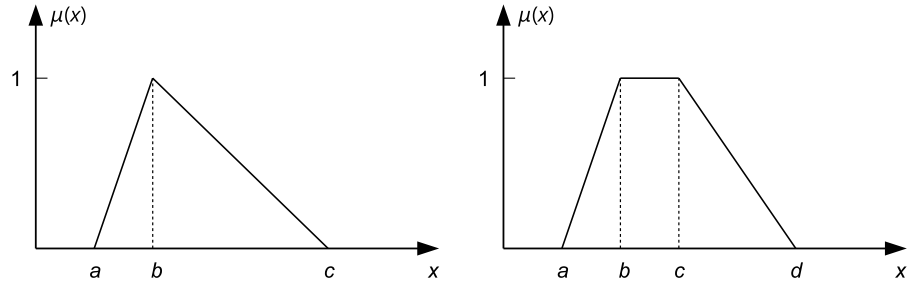
Szczególnymi przypadkami liczb rozmytych są liczby trójkątne i trapezowe (rys. 4.1), zwane interwałami rozmytymi. Funkcja przynależności liczby trapezowej $\mathrm{TrFN}(a, b, c, d)$

opisana jest zależnością:

$$\mu(x) = \begin{cases} 0 & \text{dla} \quad x < a, \\ \frac{x-a}{b-a} & \text{dla} \quad a \leqslant x < b, \\ 1 & \text{dla} \quad b \leqslant x < c, \\ \frac{d-x}{d-c} & \text{dla} \quad c \leqslant x < d, \\ 0 & \text{dla} \quad x \geqslant d, \end{cases} \tag{4.2}$$

a trójkątnej TFN(a, b, c):

$$\mu(x) = \begin{cases} 0 & \text{dla} \quad x < a, \\ \frac{x-a}{b-a} & \text{dla} \quad a \leqslant x < b, \\ 1 & \text{dla} \quad x = b, \\ \frac{c-x}{c-b} & \text{dla} \quad b < x < c, \\ 0 & \text{dla} \quad x \geqslant c. \end{cases} \tag{4.3}$$



Rysunek 4.1. Funkcje przynależności przykładowej trójkątnej i trapezowej liczby rozmytej

Kolejnym istotnym terminem jest pojęcie α-przekroju zbioru rozmytego. α-przekrój A_α zbioru rozmytego A jest nierozmytym zbiorem zdefiniowanym jako:

$$A_\alpha = \{x \in \mathbb{R} : \ \mu(x) \geqslant \alpha\}. \tag{4.4}$$

Zbiór α-przekrojów $\{A_\alpha : \ \alpha \in (0, 1]\}$ może stanowić reprezentację liczby rozmytej A:

$$A = \bigcup_{\alpha \in [0,1]} A_\alpha. \tag{4.5}$$

Z definicji liczby rozmytej wynika, że każdy α-przekrój jest spójny dla każdego $\alpha \geqslant 0$. W konsekwencji, każdy α-przekrój może być reprezentowany przez interwał:

$$A_\alpha = [f^{-1}(\alpha), g^{-1}(\alpha)], \tag{4.6}$$

gdzie: $f^{-1} = \inf\{x : \ \mu(x) \geqslant \alpha\}$ i $g^{-1} = \sup\{x : \ \mu(x) \geqslant \alpha\}$.

Ponieważ każdy α-przekrój jest interwałem, do sformułowania podstawowych zasad arytmetyki liczb rozmytych można wykorzystać arytmetykę interwałową [74] opisaną w podrozdziale 3.1 (wzory (3.2) – (3.5)).

Niech:

$$A = \bigcup_{\alpha\in[0,1]} [a_1^\alpha, a_2^\alpha] \quad \text{i} \quad B = \bigcup_{\alpha\in[0,1]} [b_1^\alpha, b_2^\alpha],$$

będą liczbami rozmytymi. Wówczas:

$$A \circ B = \bigcup_{\alpha\in[0,1]} ([a_1^\alpha, a_2^\alpha] \circ [b_1^\alpha, b_2^\alpha]), \tag{4.7}$$

gdzie: $\circ = \{+, -, \times, \div\}$ [24, 26, 46, 54].

Inaczej mówiąc, dla liczb rozmytych A i B działania zdefiniowane wzorami (3.2) – (3.5) wykonywane są dla każdej pary α-przekrojów, $\alpha \in [0, 1]$, które są interwałami i inaczej można je przedstawić w postaci $[\underline{a}(\alpha), \overline{a}(\alpha)]$ i $[\underline{b}(\alpha), \overline{b}(\alpha)]$.

4.2. Wielowymiarowa arytmetyka liczb rozmytych oparta na notacji RDM

Przedstawiona powyżej arytmetyka liczb rozmytych jest najbardziej popularna, ale nie jest jedynym możliwym podejściem [24, 26, 54, 126, 127], innym, często wykorzystywanym sposobem realizacji różnych działań arytmetycznych jest wykorzystanie zasady rozszerzenia, sformułowanej przez L. Zadeha [140, 141, 143].

Zarówno arytmetykę bazującą na działaniach na α-przekrojach, jak i opartą na zasadzie rozszerzenia, zalicza się do tzw. standardowych arytmetyk rozmytych [29, 36, 46]. W literaturze spotyka się ponadto inne przykłady działań na liczbach rozmytych, np.: uogólniona metoda wierzchołków (ang. *generalized vertex method*) [23], arytmetyka rozmyta z ograniczeniami (ang. *constrained fuzzy arithmetic*) [55, 56, 64], algorytmiczna arytmetyka rozmyta (ang. *algorithmic fuzzy arithmetic*) [84], metoda oparta na transformacji (ang. *transformation method*) [46], odwrotna arytmetyka rozmyta (ang. *inverse fuzzy arithmetic*) [46], metoda oparta na reprezentacji parametrycznej [128]. Dobry przegląd różnych arytmetyk rozmytych można znaleźć w pracach [29, 129].

Standardowa arytmetyka oparta na α-przekrojach posiada cały szereg mankamentów, odziedziczonych głównie po arytmetyce interwałowej Moore'a. Przykładowo, nie istnieje w niej element przeciwny i odwrotny do zadanej liczby rozmytej, nie działa prawo rozdzielności mnożenia względem dodawania, nie ma możliwości uwzględnienia zależności występujących pomiędzy operandami. Wady te powodują kłopoty w przekształcaniu wzorów, a konsekwencją tego jest między innymi kłopotliwe rozwiązywanie nawet naj-

prostszych równań rozmytych. Z tego powodu możliwości standardowej arytmetyki są ograniczone.

Rozwiązaniem wielu wymienionych problemów może być zastosowanie wielowymiarowej arytmetyki rozmytej RDM (ang. *multi-dimensional, relative distance measure fuzzy arithmetic*). Arytmetyka ta bazuje na zastosowaniu tzw. poziomej funkcji przynależności, a jej autorem jest prof. A. Piegat [90]. Jej koncepcja oraz liczne przykłady zastosowań zaprezentowane zostały w pracach [91, 92, 93, 95].

Na rys. 4.1 pokazano przykładowe funkcje przynależności trójkątnej i trapezowej liczby rozmytej. Liczba trójkątna może być traktowana jako szczególny przypadek liczby trapezowej, w którym $b = c$. W teorii systemów rozmytych liczby modelowane są za pomocą funkcji przynależności, które opisują zależność: $\mu = f(x)$. Dla przypomnienia, dla trapezowej liczby rozmytej ma ona postać:

$$\mu(x) = \begin{cases} (x-a)/(b-a) & \text{dla } \ x \in [a,b), \\ 1 & \text{dla } \ x \in [b,c], \\ (d-x)/(d-c) & \text{dla } \ x \in (c,d], \\ 0 & \text{poza tym.} \end{cases} \tag{4.8}$$

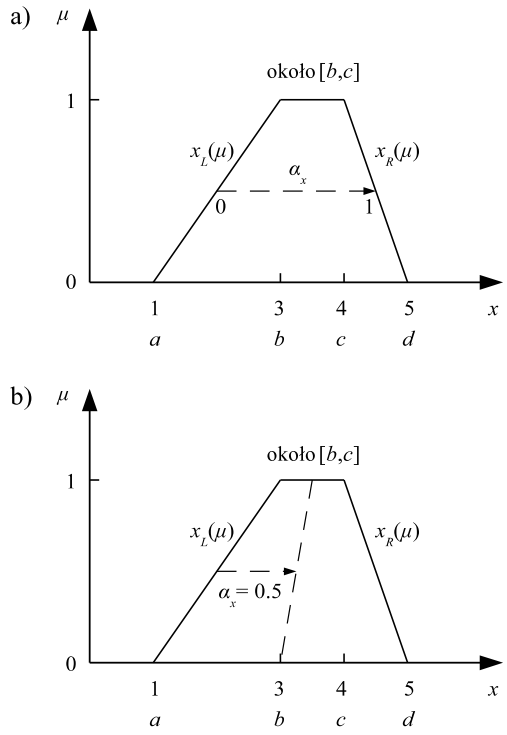
Zależność (4.8) opisuje związek pomiędzy „pionową" zmienną μ, a „poziomą" zmienną x. Tak naprawdę, modeluje jedynie brzegi funkcji przynależności, pomijając jej część wewnętrzną. Jeśli użyjemy takiej funkcji w obliczeniach na liczbach rozmytych, również wykorzystamy tylko jej brzegi. Ma to wpływ na obniżenie dokładności uzyskiwanych wyników obliczeń, a w wielu przypadkach zwiększa także ich komplikację. Funkcje przynależności typu (4.8) będziemy dalej nazywali funkcjami „pionowymi".

Można zadać pytanie: czy jest możliwe utworzenie odwrotnej (poziomej) funkcji opisującej zależność $x = f^{-1}(\mu)$. Pozornie wydaje się to niemożliwe, ponieważ zależność taka będzie niejednoznaczna i przez to nie będzie funkcją. Okazuje się jednak, że taki model może być utworzony przy założeniu, że wartością funkcji będzie interwał.

Rozważmy teraz poziomy przekrój funkcji przynależności na poziomie μ. Przekrój taki będziemy dalej nazywali μ-przekrojem (w odróżnieniu od tradycyjnie przyjętego α-przekroju), rys. 4.2a. Zmienna α_x, $\alpha_x \in [0,1]$, będzie nazywana zmienną RDM (ang. *RDM – relative distance measure*). Określa ona względną odległość punktu $x^* \in [x_L(\mu), x_R(\mu)]$ od początku lokalnego układu współrzędnych (rys. 4.2).

Zmienna RDM α_x wprowadza lokalny kartezjański układ współrzędnych w interwale, pozwalając tym samym modelować jego wnętrze. Lewy brzeg funkcji przynależności $x_L(\mu)$ oraz prawy brzeg $x_R(\mu)$ opisują zależności:

$$x_L = a + (b-a)\mu, \qquad x_R = d - (d-c)\mu. \tag{4.9}$$



Rysunek 4.2. Ilustracja opisu funkcji przynależności z wykorzystaniem zmiennej RDM

Zmienna RDM α_x umożliwia transformację lewego brzegu $x_L(\mu)$ funkcji przynależności, w brzeg prawy $x_R(\mu)$.

Zdefiniujmy teraz następującą funkcję:

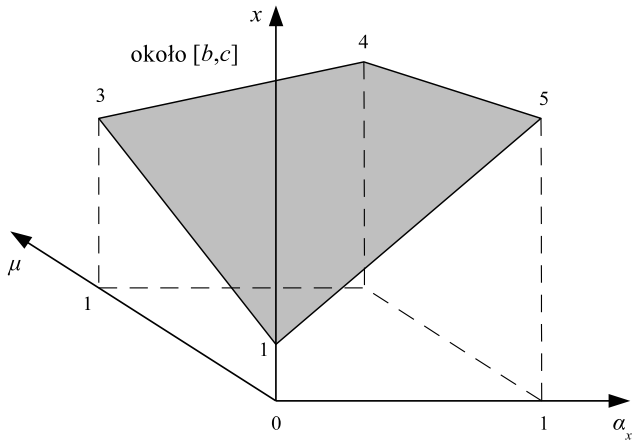
$$x(\mu, \alpha_x) = x_L + (x_R - x_L)\alpha_x, \quad \alpha_x \in [0, 1], \tag{4.10}$$

która dla trapezowej liczby rozmytej przyjmie postać:

$$x = [a + (b - a)\mu] + [(d - a) - (d - c + b - a)\mu]\alpha_x, \quad \alpha_x \in [0, 1]. \tag{4.11}$$

Funkcję taką będziemy dalej nazywać poziomą funkcją przynależności. Funkcja $x = f(\mu, \alpha_x)$ jest funkcją dwóch zmiennych, stąd jej wykres jest trójwymiarowy (rys. 4.3). Jak każda funkcja opisuje ona jednoznaczną zależność x od μ i α_x.

Pozioma funkcja przynależności $x = f(\mu, \alpha_x)$ nie definiuje pojedynczej wartości zmiennej x, ale zbiór możliwych wartości x dla zadanego μ-przekroju, opisuje zatem granulę informacyjną i będzie dalej oznaczana jako x^{gr}. Równanie (4.11) przedstawia funkcję dla liczby trapezowej. Przyjmując $b = c$, łatwo otrzymamy wzór dla liczby trójkątnej, a przyjmując $a = b$ i $c = d$ wzór dla liczby prostokątnej.



Rysunek 4.3. Pozioma funkcja przynależności $x = (1 + 2\mu) + (4 - 3\mu)\alpha_x$, $\alpha_x \in [0, 1]$, odpowiadająca pionowej funkcji pokazanej na rys. 4.2

4.2.1. Arytmetyka liczb rozmytych w postaci RDM

Załóżmy, że dana jest pozioma funkcja przynależności $x^{gr} = f(\mu, \alpha_x)$ opisująca rozmyty interwał X (4.12) i pozioma funkcja przynależności $y^{gr} = f(\mu, \alpha_y)$ opisująca rozmyty interwał Y (4.13).

$$X : x^{gr} = [a_x + (b_x - a_x)\mu] + [(d_x - a_x) - (d_x - c_x + b_x - a_x)\mu]\alpha_x, \quad \mu, \alpha_x \in [0, 1] \tag{4.12}$$

$$Y : y^{gr} = [a_y + (b_y - a_y)\mu] + [(d_y - a_y) - (d_y - c_y + b_y - a_y)\mu]\alpha_y, \quad \mu, \alpha_y \in [0, 1] \tag{4.13}$$

Dodawanie dwóch niezależnych interwałów rozmytych

$$X + Y = Z : x^{gr}(\mu, \alpha_x) + y^{gr}(\mu, \alpha_y) = z^{gr}(\mu, \alpha_x, \alpha_y), \quad \mu, \alpha_x, \alpha_y \in [0, 1] \tag{4.14}$$

Przykładowo, jeśli X jest liczbą trapezową TrFN(1,3,4,5):

$$x^{gr} = (1 + 2\mu) + (4 - 3\mu)\alpha_x, \tag{4.15}$$

a Y jest liczbą trapezową TrFN(1,2,3,4):

$$y^{gr} = (1 + \mu) + (3 - 2\mu)\alpha_y, \tag{4.16}$$

to wynik dodawania z^{gr} jest równy:

$$z^{gr} = (2 + 3\mu) + (4 - 3\mu)\alpha_x + (3 - 2\mu)\alpha_y, \quad \mu, \alpha_x, \alpha_y \in [0, 1]. \tag{4.17}$$

Wykres rozwiązania (4.17) byłby 4-wymiarowy, stąd nie można go w sposób pełny zobrazować. Można jednak pokazać jego mniej wymiarowe reprezentacje.

Taką przykładową 2-wymiarową reprezentacją może być rozpiętość $s(z^{gr})$. Można ją wyznaczyć stosując znane metody badania funkcji:

$$s(z^{gr}) = [\min_{\alpha_x,\alpha_y} z^{gr}(\mu, \alpha_x, \alpha_y), \max_{\alpha_x,\alpha_y} z^{gr}(\mu, \alpha_x, \alpha_y)].$$

Dla przedstawionego powyżej przykładu, ekstrema funkcji (4.17) znajdują się na brzegach dziedziny. Minimum mamy dla $\alpha_x = \alpha_y = 0$, a maksimum dla $\alpha_x = \alpha_y = 1$. Ostatecznie rozpiętość 4-wymiarowej granuli (4.17) opisana jest jako:

$$s(z^{gr}) = [2 + 3\mu, 9 - 2\mu], \quad \mu \in [0, 1]. \tag{4.18}$$

Rozpiętość (4.18) nie jest wynikiem dodawania, który ma postać 4-wymiarowej funkcji (4.17). Rozpiętość jest jedynie 2-wymiarową informacją o niepewności wyniku.

Odejmowanie dwóch niezależnych interwałów rozmytych

$$\begin{aligned} X - Y = Z : x^{gr}(\mu, \alpha_x) - y^{gr}(\mu, \alpha_y) = z^{gr}(\mu, \alpha_x, \alpha_y), \\ \mu, \alpha_x, \alpha_y \in [0, 1] \end{aligned} \tag{4.19}$$

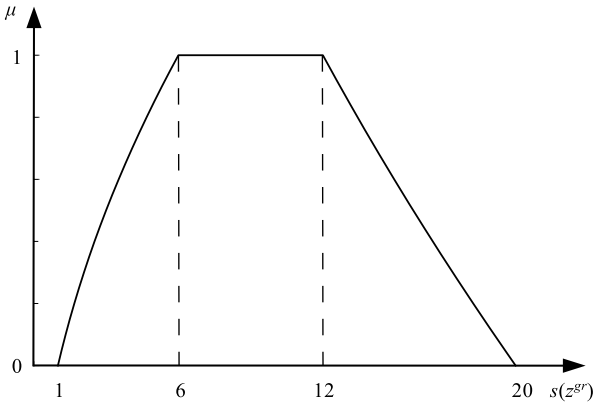
Przykładowo, jeśli X i Y są liczbami trapezowymi (4.15) i (4.16), wówczas wynik odejmowania jest równy:

$$z^{gr} = \mu + (4 - 3\mu)\alpha_x - (3 - 2\mu)\alpha_y, \quad \mu, \alpha_x, \alpha_y \in [0, 1]. \tag{4.20}$$

Jeśli chcielibyśmy otrzymać rozpiętość $s(z^{gr})$ 4-wymiarowego wyniku, możemy ją wyznaczyć z zależności:

$$\begin{aligned} s(z^{gr}) = [\min_{\alpha_x,\alpha_y} z^{gr}, \max_{\alpha_x,\alpha_y} z^{gr}] = [-3 + 3\mu, 4 - 2\mu], \\ \mu \in [0, 1]. \end{aligned} \tag{4.21}$$

Rozpiętość (4.21) wyniku z^{gr} (4.20) odpowiada $\alpha_x = 0$, $\alpha_y = 1$ dla $\min z^{gr}$ i $\alpha_x = 1$, $\alpha_y = 0$ dla $\max z^{gr}$.



Rysunek 4.4. Rozpiętość 4-wymiarowej funkcji (4.23) będącej wynikiem mnożenia

Mnożenie dwóch niezależnych interwałów rozmytych

$$X \cdot Y = Z : x^{gr}(\mu, \alpha_x) \cdot y^{gr}(\mu, \alpha_y) = z^{gr}(\mu, \alpha_x, \alpha_y), \quad \mu, \alpha_x, \alpha_y \in [0,1] \tag{4.22}$$

Przykładowo, jeśli X i Y są liczbami trapezowymi (4.15) i (4.16) wówczas wynik mnożenia z^{gr} jest równy:

$$z^{gr} = x^{gr} \cdot y^{gr} = [(1+2\mu) + (4-3\mu)\alpha_x] \cdot [(1+\mu) + (3-2\mu)\alpha_y], \quad \mu, \alpha_x, \alpha_y \in [0,1]. \tag{4.23}$$

Równanie (4.23) opisuje pełny 4-wymiarowy wynik mnożenia. Uproszczoną 2-wymiarową reprezentację w postaci rozpiętości możemy wyznaczyć z zależności (4.24). Wynik dodatkowo pokazany jest na rys. 4.4.

$$s(z^{gr}) = [\min_{\alpha_x,\alpha_y} z^{gr}, \max_{\alpha_x,\alpha_y} z^{gr}] = [(1+2\mu)(1+\mu), (5-\mu)(4-\mu)], \quad \mu \in [0,1]. \tag{4.24}$$

Dzielenie dwóch niezależnych interwałów rozmytych, $0 \notin Y$

$$X/Y = Z : x^{gr}(\mu, \alpha_x)/y^{gr}(\mu, \alpha_y) = z^{gr}(\mu, \alpha_x, \alpha_y), \quad \mu, \alpha_x, \alpha_y \in [0,1] \tag{4.25}$$

Rysunek 4.5. Rozpiętość pełnego, 4-wymiarowego wyniku dzielenia (4.26)

Przykładowo, jeśli X i Y są liczbami trapezowymi (4.15) i (4.16) wówczas wynik dzielenia z^{gr} jest równy:

$$\begin{aligned} z^{gr} = x^{gr}/y^{gr} &= \frac{(1+2\mu)+(4-3\mu)\alpha_x}{(1+\mu)+(3-2\mu)\alpha_y}, \\ &\mu, \alpha_x, \alpha_y \in [0,1]. \end{aligned} \tag{4.26}$$

Rozpiętość $s(z^{gr})$ pełnego wyniku (4.26) wyrażona jest zależnością (4.27) i pokazana jest dodatkowo na rys. 4.5.

$$\begin{aligned} s(z^{gr}) &= [\min_{\alpha_x,\alpha_y} z^{gr}, \max_{\alpha_x,\alpha_y} z^{gr}] \\ &= \left[\frac{1+2\mu}{4-\mu}, \frac{5-\mu}{1+\mu}\right], \quad \mu \in [0,1]. \end{aligned} \tag{4.27}$$

Pełne rozwiązanie (4.26) ma postać 4-wymiarowej granuli informacyjnej, stąd nie może być zobrazowane na wykresie. Można jednakże pokazać je w sposób uproszczony w przestrzeni 3-wymiarowej $X \times Y \times Z$ (ustalając zmienną μ). Rys. 4.6 pokazuje powierzchnie dla ustalonych wartości $\mu = 0$ i $\mu = 1$.

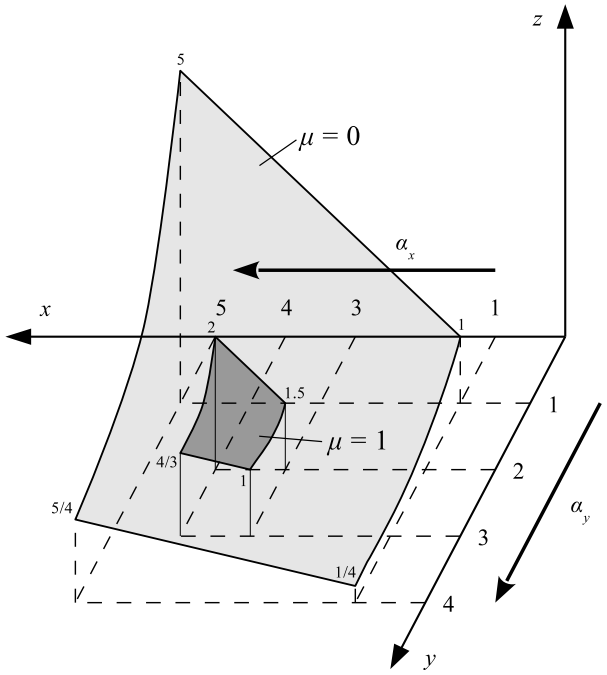
4.2.2. Własności arytmetyki liczb rozmytych w postaci RDM

Przemienność

Dla liczb rozmytych X i Y prawdziwe są równania (4.28) i (4.29).

$$X + Y = Y + X \tag{4.28}$$

$$XY = YX \tag{4.29}$$



Rysunek 4.6. Uproszczony widok 4-wymiarowej granuli informacyjnej z^{gr} opisanej równaniem (4.26), w przestrzeni 3-wymiarowej $X \times Y \times Z$

Łączność

Dla liczb rozmytych X, Y and Z, prawdziwe są równania (4.30) i (4.31).

$$X + (Y + Z) = (X + Y) + Z \tag{4.30}$$

$$X(YZ) = (XY)Z \tag{4.31}$$

Neutralny element w dodawaniu i mnożeniu

W arytmetyce liczb rozmytych w postaci RDM istnieją neutralne elementy w dodawaniu i mnożeniu. Mają one postać zdegenerowanych interwałów 0 i 1.

$$X + 0 = 0 + X = X \tag{4.32}$$

$$X \cdot 1 = 1 \cdot X = X \tag{4.33}$$

Elementy przeciwne i odwrotne

W arytmetyce liczb rozmytych w postaci RDM, interwał rozmyty:

$$-X: \; -x^{gr} = -\left[a + (b-a)\mu\right] - \left[(d-a) - \mu(d-a+b-c)\right]\alpha_x, \quad \alpha_x \in [0,1],$$

jest elementem przeciwnym interwału:

$$X: \; x^{gr} = [a + (b-a)\mu] + [(d-a) \\ - \mu(d - a + b - c)]\alpha_x, \quad \alpha_x \in [0,1].$$

Jeśli parametry dwóch interwałów rozmytych X i Y są równe: $a_x = a_y$, $b_x = b_y$, $c_x = c_y$, $d_x = d_y$, wówczas interwał $-Y$ jest interwałem przeciwnym do X, tylko wtedy, kiedy równe są także zmienne RDM: $\alpha_x = \alpha_y$. Oznacza to pełną zależność obu niepewnych zmiennych, modelowanych przez interwały rozmyte.

Zakładając, że $0 \notin X$, w arytmetyce liczb rozmytych w postaci RDM elementem odwrotnym interwału rozmytego X jest interwał:

$$\frac{1}{X}: \; \frac{1}{x^{gr}} = \frac{1}{[a + (b-a)\mu] + [(d-a) - \mu(d - a + b - c)]\alpha_x},$$
$$\alpha_x \in [0,1].$$

Jeśli parametry dwóch interwałów rozmytych X i Y są równe: $a_x = a_y$, $b_x = b_y$, $c_x = c_y$, $d_x = d_y$, wówczas interwał $1/Y$ jest interwałem odwrotnym do X tylko wtedy, kiedy równe są także zmienne RDM: $\alpha_x = \alpha_y$ (pełna zależność zmiennych).

Prawo rozdzielności mnożenia względem dodawania

W arytmetyce liczb rozmytych w postaci RDM prawdziwe jest prawo rozdzielności mnożenia względem dodawania:

$$X(Y + Z) = XY + XZ. \tag{4.34}$$

Konsekwencją tego prawa jest możliwość łatwego przekształcania wzorów.

Prawa skracania dla dodawania i mnożenia

W arytmetyce liczb rozmytych w postaci RDM prawdziwe są prawa skracania dla dodawania i mnożenia:

$$X + Z = Y + Z \Rightarrow X = Y, \tag{4.35}$$
$$XZ = YZ \Rightarrow X = Y. \tag{4.36}$$

4.3. Odległość pomiędzy liczbami rozmytymi

Podobnie jak w przypadku interwałów, aby wyznaczać najbliższych sąsiadów, a także by zdefiniować kryterium jakości dla rozmytej wersji metody najmniejszych kwadratów, musimy określić pojęcie odległości pomiędzy niepewnymi wartościami w postaci liczb rozmytych.

W literaturze opisanych jest kilka sposobów wyznaczania odległości pomiędzy liczbami rozmytymi [8, 20, 38, 61, 134]. Każdy z nich posiada swoje wady i zalety, stąd trudno określić, który z nich jest najlepszy [19]. W dalszej części niniejszej rozprawy, wykorzystana zostanie metoda zaproponowana w artykule [38]. Głównym powodem jej zastosowania jest wygoda w formułowaniu kryterium jakości dla rozmytej wersji metody najmniejszych kwadratów oraz czułość na zmiany położenia obu brzegów liczb rozmytych.

Odległość, indeksowana parametrami $p \in [1, \infty)$ i $q \in [0, 1]$, pomiędzy liczbami rozmytymi A i B może być wyznaczona za pomocą zależności:

$$d_{p,q}(A,B) = \begin{cases} \sqrt[p]{(1-q)\int_0^1 |f_B^{-1}(\alpha) - f_A^{-1}(\alpha)|^p \,\mathrm{d}\alpha + q\int_0^1 |g_B^{-1}(\alpha) - g_A^{-1}(\alpha)|^p \,\mathrm{d}\alpha} \\ \qquad\qquad \text{dla} \;\; 1 \leqslant p < \infty \\ (1-q)\sup\limits_{0<\alpha\leqslant 1}(|f_B^{-1}(\alpha) - f_A^{-1}(\alpha)|) + q\sup\limits_{0<\alpha\leqslant 1}(|g_B^{-1}(\alpha) - g_A^{-1}(\alpha)|) \\ \qquad\qquad \text{dla} \;\; p = \infty \end{cases} \tag{4.37}$$

albo inaczej, odległość indeksowaną parametrem $p \in [1, \infty)$, można obliczyć jako:

$$d_p(A,B) = \begin{cases} \max\left\{\sqrt[p]{\int_0^1 |f_B^{-1}(\alpha) - f_A^{-1}(\alpha)|^p \,\mathrm{d}\alpha}, \sqrt[p]{\int_0^1 |g_B^{-1}(\alpha) - g_A^{-1}(\alpha)|^p \,\mathrm{d}\alpha}\right\} \\ \qquad\qquad \text{dla} \;\; 1 \leqslant p < \infty \\ \max\left\{\sup\limits_{0<\alpha\leqslant 1}(|f_B^{-1}(\alpha) - f_A^{-1}(\alpha)|), \sup\limits_{0<\alpha\leqslant 1}(|g_B^{-1}(\alpha) - g_A^{-1}(\alpha)|)\right\} \\ \qquad\qquad \text{dla} \;\; p = \infty \end{cases} \tag{4.38}$$

gdzie: $A_\alpha = [f_A^{-1}(\alpha), g_A^{-1}(\alpha)]$ i $B_\alpha = [f_B^{-1}(\alpha), g_B^{-1}(\alpha)]$.

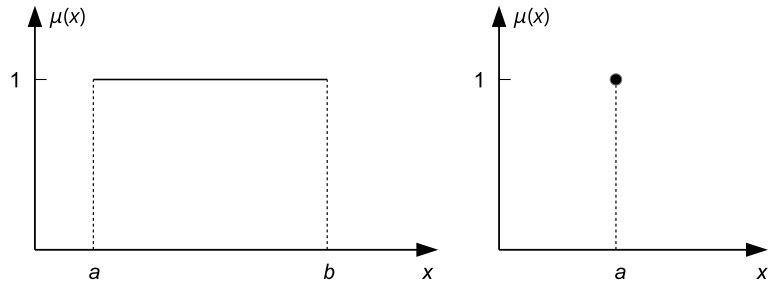
Parametr q w metryce $d_{p,q}$ charakteryzuje wagi przypisane do lewego i prawego brzegu liczby rozmytej. Jeśli nie ma powodu, aby wyróżnić któryś z nich, najlepiej przyjąć $q = 0.5$. Jeżeli założymy, że wszystkie liczby rozmyte są elementami przestrzeni $F(\mathbb{R})$, wówczas można udowodnić, że $(F(\mathbb{R}), d_{p,q})$ i $(F(\mathbb{R}), d_p)$ są przestrzeniami metrycznymi [38].

W dalszej części monografii, wykorzystana będzie metryka (4.37) z parametrem $p = 2$ i $q = 0.5$:

$$\begin{aligned} d_2 &= d_{p=2,q=0.5}(A,B) \\ &= \sqrt{(0.5\int_0^1 (f_B^{-1}(\alpha) - f_A^{-1}(\alpha))^2 \,\mathrm{d}\alpha + 0.5\int_0^1 (g_B^{-1}(\alpha) - g_A^{-1}(\alpha))^2 \,\mathrm{d}\alpha} \end{aligned} \tag{4.39}$$

Interwały można opisywać za pomocą prostokątnych liczb rozmytych. Funkcja

przynależności dla przykładowej prostokątnej liczby RFN(a, b) pokazana jest na rys. 4.7. Można zauważyć, że metryka d_2 (3.15) wykorzystywana do wyznaczania odległości między interwałami jest szczególną wersją metryki (4.39), przy założeniu, że odległość liczona jest między prostokątnymi liczbami rozmytymi.



Rysunek 4.7. Funkcja przynależności przykładowej prostokątnej liczby rozmytej RFN(a, b) i jednoelementowej liczby rozmytej (singleton) SFN(a)

4.4. Rozmyta wersja metody najmniejszych kwadratów

Koncepcja rozmytej regresji liniowej po raz pierwszy została zaprezentowana przez H. Tanakę i współautorów w pracy [132]. Opracowali oni model regresji z rozmytymi parametrami, wyznaczanymi za pomocą programowania liniowego. Jego zadaniem była minimalizacja całkowitej niepewności rozmytych parametrów, przy założeniu, że nośnik rozmytych wartości estymowanych pokrywa nośnik rozmytych wartości obserwowanych, na podstawie których tworzony jest model. W kolejnych publikacjach autorzy przedstawili udoskonaloną wersję swojej metody [130, 131, 133], jednak nie była ona pozbawiona wad [114]. Przede wszystkim okazała się bardzo wrażliwa na występowanie w danych punktów odstających (ang. *outliers*). Ponadto szerokość estymowanych wartości szybko rosła razem z ilością danych uwzględnianych w modelu, osiągając w skrajnych przypadkach wartość nieskończenie dużą. W pracy [118] M. Sakawa i H. Yano zaprezentowali rozmyty model liniowy, w którym zarówno wejście, wyjście, jak i parametry były liczbami rozmytymi. Parametry w modelu również wyznaczane były z wykorzystaniem programowania liniowego, przy czym zastosowano tu bardziej złożoną funkcję celu. D. Redden i W. Woodall w artykule [115] wykazali, że także i ten model jest bardzo wrażliwy na występowanie w danych punktów odstających, a dodatkowo w pewnych sytuacjach nie wszystkie obserwacje mogą być wykorzystywane w trakcie tworzenia modelu.

Generalnie, przy poszukiwaniu rozmytej regresji daje się zauważyć dwa zasadnicze podejścia. Pierwsze bazuje na programowaniu liniowym [85, 118, 130, 131, 132, 133], natomiast drugie na rozmytej wersji metody najmniejszych kwadratów [18, 70, 116]. W podejściu pierwszym optymalizowane kryterium opisuje rozmycie rozwiązania, poszukuje się więc rozwiązania o minimalnej niepewności, pokrywającego całkowicie lub w zadanym

stopniu dane obserwowane. Jego największą zaletą jest względna prostota obliczeń [135]. W podejściu drugim, metoda najmniejszych kwadratów błędów pozwala na utworzenie modelu o najlepszym dopasowaniu do danych. Tu z kolei największą zaletą jest mniejsza niepewność otrzymywanego rozwiązania. Opisane są także rozwiązania, w których oba podejścia stosowane są jednocześnie [71, 78, 79, 81].

W większości publikacji opisane są metody poszukiwania regresji o rozmytych parametrach [6, 8, 113, 139]. Skutkuje to zwykle dużą złożonością metod. Konsekwencją stosowania rozmytych parametrów mogą być także estymaty nadmiernie (lub zbyt słabo) rozmyte. Zjawisko takie można zaobserwować szczególnie w przypadku zastosowania regresji do zadania ekstrapolacji. W skrajnych sytuacjach może nawet dojść do zamiany lewego i prawego brzegu estymaty.

W rozwiązaniu zaprezentowanym poniżej, zastosowano uproszczony model z parametrami rzeczywistymi, wyznaczany dla środków liczb rozmytych. Tego typu podejście jest proste obliczeniowo i daje dobre, wiarygodne rozwiązania.

4.4.1. Model z jednym wejściem

Rozmyta wersja metody najmniejszych kwadratów zostanie zaprezentowana w pierwszej kolejności dla modelu z jednym wejściem. Załóżmy, że posiadamy dane w postaci par (X_i, Y_i), $i = 1 \ldots K$, przy czym zarówno wejście X_i, jak i wyjście Y_i ma postać liczby rozmytej. W przypadku, w którym jakaś wartość określona jest w postaci interwału, można zastąpić ją prostokątną liczbą rozmytą. W przypadku, w którym jakaś wartość określona jest w postaci liczby rzeczywistej, można zastąpić ją liczbą rozmytą typu singleton. Jednym z ważnych założeń zaproponowanej poniżej wersji metody jest jej skuteczna praca zarówno z danymi w postaci liczb rozmytych, jak i interwałów oraz liczb rzeczywistych.

Będziemy poszukiwali liniowego modelu aproksymującego w formie:

$$\hat{Y} = w_1 \cdot x_s + w_0 + R, \tag{4.40}$$

gdzie: x_s oznacza środek liczby rozmytej X, będącej wejściem modelu, R – trójkątna liczba rozmyta w postaci $\mathrm{TFN}(-r, 0, r)$.

Środek liczby rozmytej możemy wyznaczyć z zależności:

$$x_s = \frac{1}{2} \int_0^1 [f_X^{-1}(\alpha) + g_X^{-1}(\alpha)] \mathrm{d}\alpha. \tag{4.41}$$

gdzie: f to funkcja opisująca lewy brzeg, a g to funkcja opisująca prawy brzeg liczby rozmytej (w powyższym wzorze liczby X).

Dla liczby trójkątnej $R = \mathrm{TFN}(-r, 0, r)$ mamy:

$$f(x) = \frac{1}{r}(x + r) = \alpha \qquad g(x) = -\frac{1}{r}(x - r) = \alpha, \tag{4.42}$$

i odpowiednio:

$$f^{-1}(\alpha) = r \cdot (\alpha - 1) \qquad g^{-1}(\alpha) = r \cdot (1 - \alpha). \tag{4.43}$$

Będziemy chcieli, aby model dopasowany był do danych w taki sposób, aby minimalizować kryterium jakości oparte na metryce d_2 opisanej zależnością (4.39):

$$Q(w_1, w_0, r) = \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) \longrightarrow \min, \tag{4.44}$$

gdzie: $\hat{Y}_i = w_1 \cdot x_{si} + w_0 + R$.

Twierdzenie 4.1. Dla modelu aproksymującego w postaci (4.40) współczynniki minimalizujące kryterium jakości (4.44) można obliczyć z zależności:

$$w_1 = \frac{K \cdot \sum_{i=1}^{K} x_{si} \cdot \int_0^1 \frac{1}{2}\left[f_{Yi}^{-1}(\alpha) + g_{Yi}^{-1}(\alpha)\right] \mathrm{d}\alpha - \sum_{i=1}^{K} x_{si} \cdot \sum_{i=1}^{K} \int_0^1 \frac{1}{2}\left[f_{Yi}^{-1}(\alpha) + g_{Yi}^{-1}(\alpha)\right] \mathrm{d}\alpha}{K \cdot \sum_{i=1}^{K} x_{si}^2 - \left(\sum_{i=1}^{K} x_{si}\right)^2},$$

$$w_0 = \frac{1}{K} \cdot \left[\sum_{i=1}^{K} \int_0^1 \frac{1}{2}\left[f_{Yi}^{-1}(\alpha) + g_{Yi}^{-1}(\alpha)\right] \mathrm{d}\alpha - w_1 \cdot \sum_{i=1}^{K} x_{si}\right],$$

$$r = \frac{3}{2K} \cdot \sum_{i=1}^{K} \int_0^1 (1 - \alpha)\left[g_{Yi}^{-1}(\alpha) - f_{Yi}^{-1}(\alpha)\right] \mathrm{d}\alpha. \tag{4.45}$$

Dowód

Kryterium jakości (4.44) można inaczej zapisać w postaci:

$$\begin{aligned} Q(w_1, w_0, r) &= \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) = \sum_{i=1}^{K} d_2^2(w_1 \cdot x_{si} + w_0 + R, Y_i) \\ &= \frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot x_{si} + w_0 + r(\alpha - 1) - f_{Yi}^{-1}(\alpha)\right]^2 \mathrm{d}\alpha \\ &\quad + \frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot x_{si} + w_0 + r(1 - \alpha) - g_{Yi}^{-1}(\alpha)\right]^2 \mathrm{d}\alpha \end{aligned}$$

Aby wyznaczyć minimum, liczymy pochodne cząstkowe po parametrach modelu i przyrównujemy je do zera.

$$\begin{aligned}\frac{\partial Q(w_1,w_0,r)}{\partial w_1} =& \sum_{i=1}^{K}\int_0^1 \left[w_1\cdot x_{si}+w_0+r(\alpha-1)-f_{Yi}^{-1}(\alpha)\right]\cdot x_{si}\mathrm{d}\alpha \\ &+\sum_{i=1}^{K}\int_0^1 \left[w_1\cdot x_{si}+w_0+r(1-\alpha)-g_{Yi}^{-1}(\alpha)\right]\cdot x_{si}\mathrm{d}\alpha \\ =&\sum_{i=1}^{K}\int_0^1 w_1 xsi^2+w_0x_{si}+r(\alpha-1)x_{si}-f_{Yi}^{-1}(\alpha)x_{si} \\ &+w_1xsi^2+w_0x_{si}+r(1-\alpha)x_{si}-g_{Yi}^{-1}(\alpha)x_{si}\mathrm{d}\alpha \\ =&\sum_{i=1}^{K}\int_0^1 2w_1xsi^2+2w_0x_{si}-\left[f_{Yi}^{-1}(\alpha)+g_{Yi}^{-1}(\alpha)\right]x_{si}\mathrm{d}\alpha=0\end{aligned}$$

$$\begin{aligned}\frac{\partial Q(w_1,w_0,r)}{\partial w_0} =& \sum_{i=1}^{K}\int_0^1 \left[w_1\cdot x_{si}+w_0+r(\alpha-1)-f_{Yi}^{-1}(\alpha)\right]\mathrm{d}\alpha \\ &+\sum_{i=1}^{K}\int_0^1 \left[w_1\cdot x_{si}+w_0+r(1-\alpha)-g_{Yi}^{-1}(\alpha)\right]\mathrm{d}\alpha \\ =&\sum_{i=1}^{K}\int_0^1 2w_1xsi+2w_0-\left[f_{Yi}^{-1}(\alpha)+g_{Yi}^{-1}(\alpha)\right]\mathrm{d}\alpha=0\end{aligned}$$

$$\begin{aligned}\frac{\partial Q(w_1,w_0,r)}{\partial r} =& \sum_{i=1}^{K}\int_0^1 \left[w_1\cdot x_{si}+w_0+r(\alpha-1)-f_{Yi}^{-1}(\alpha)\right]\cdot(\alpha-1)\mathrm{d}\alpha \\ &+\sum_{i=1}^{K}\int_0^1 \left[w_1\cdot x_{si}+w_0+r(1-\alpha)-g_{Yi}^{-1}(\alpha)\right]\cdot(1-\alpha)\mathrm{d}\alpha \\ =&\sum_{i=1}^{K}\int_0^1 \left[r(\alpha-1)^2-f_{Yi}^{-1}(\alpha)(\alpha-1)+r(1-\alpha)^2-g_{Yi}^{-1}(\alpha)(1-\alpha)\right]\mathrm{d}\alpha \\ =&\sum_{i=1}^{K}\int_0^1 2r(\alpha-1)^2\mathrm{d}\alpha-\sum_{i=1}^{K}\int_0^1 \left[f_{Yi}^{-1}(\alpha)(\alpha-1)+g_{Yi}^{-1}(\alpha)(1-\alpha)\right]\mathrm{d}\alpha \\ =&2Kr\int_0^1(\alpha-1)^2\mathrm{d}\alpha-\sum_{i=1}^{K}\int_0^1 \left[f_{Yi}^{-1}(\alpha)(\alpha-1)+g_{Yi}^{-1}(\alpha)(1-\alpha)\right]\mathrm{d}\alpha \\ =&\frac{2}{3}Kr-\sum_{i=1}^{K}\int_0^1 \left[f_{Yi}^{-1}(\alpha)(\alpha-1)+g_{Yi}^{-1}(\alpha)(1-\alpha)\right]\mathrm{d}\alpha=0\end{aligned}$$

Po przekształceniach otrzymujemy ostatecznie układ 3 równań z trzema niewiadomymi parametrami:

$$\begin{aligned} w_1 \sum_{i=1}^{K} x_{si}^2 + w_0 \sum_{i=1}^{K} x_{si} &= \sum_{i=1}^{K} x_{si} \cdot \frac{1}{2} \int_0^1 f_{Yi}^{-1}(\alpha) + g_{Yi}^{-1}(\alpha) \mathrm{d}\alpha, \\ w_1 \sum_{i=1}^{K} x_{si} + w_0 K &= \sum_{i=1}^{K} \frac{1}{2} \int_0^1 f_{Yi}^{-1}(\alpha) + g_{Yi}^{-1}(\alpha) \mathrm{d}\alpha, \\ \frac{2}{3} K r &= \sum_{i=1}^{K} \int_0^1 (1-\alpha) \left[g_{Yi}^{-1}(\alpha) - f_{Yi}^{-1}(\alpha) \right] \mathrm{d}\alpha, \end{aligned}$$

po rozwiązaniu którego otrzymujemy współczynniki opisane zależnościami (4.45).

Podobnie jak w przypadku interwałów, badamy dalej warunek dostateczny. Aby kryterium (4.44), oznaczone dalej w skrócie jako Q, przyjmowało minimum w punkcie opisanym równaniami (4.45), jego Hessian w tym punkcie musi być dodatnio określony. Wyznaczamy:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 Q}{\partial w_1^2} & \frac{\partial^2 Q}{\partial w_1 \partial w_0} & \frac{\partial^2 Q}{\partial w_1 \partial r} \\ \frac{\partial^2 Q}{\partial w_0 \partial w_1} & \frac{\partial^2 Q}{\partial w_0^2} & \frac{\partial^2 Q}{\partial w_0 \partial r} \\ \frac{\partial^2 Q}{\partial r \partial w_1} & \frac{\partial^2 Q}{\partial r \partial w_0} & \frac{\partial^2 Q}{\partial r^2} \end{bmatrix} = \begin{bmatrix} 2 \sum_{i=1}^{K} \int_0^1 x_{si}^2 \mathrm{d}\alpha & 2 \sum_{i=1}^{K} \int_0^1 x_{si} \mathrm{d}\alpha & 0 \\ 2 \sum_{i=1}^{K} \int_0^1 x_{si} \mathrm{d}\alpha & \sum_{i=1}^{K} \int_0^1 2 \mathrm{d}\alpha & 0 \\ 0 & 0 & \frac{2}{3} K \end{bmatrix}$$

$$= \begin{bmatrix} 2 \sum_{i=1}^{K} x_{si}^2 & 2 \sum_{i=1}^{K} x_{si} & 0 \\ 2 \sum_{i=1}^{K} x_{si} & 2 \cdot K & 0 \\ 0 & 0 & \frac{2}{3} K \end{bmatrix}. \tag{4.46}$$

Otrzymany Hessian jest podobny do tego, który wyznaczony został wcześniej dla interwałów – wzór (3.19) na stronie 33. Wiemy już zatem, że jest on dodatnio określony i tym samym kryterium (4.44) przyjmuje minimum dla współczynników opisanych równaniami (4.45).

□

Przykład 4.1.

Dla danych z tabeli należy wyznaczyć model aproksymujący w postaci (4.40), minimalizujący kryterium jakości (4.44).

X	Y
TFN(1,1.5,2)	TFN(1,1.5,2)
TrFN(2,2.5,3.5,4)	RFN(3,4)
5	TFN(2,2.5,3)
4	5
RFN(6,7)	3
TFN(7,8,8)	TrFN(4,4.5,5.5,6)

Dla ułatwienia dalszych obliczeń, wyznaczamy wartości x_s, y_s oraz funkcje odwrotne lewego i prawego brzegu funkcji przynależności wyjściowych liczb rozmytych Y.

x_s	y_s	$f^{-1}(\alpha)$	$g^{-1}(\alpha)$
1.5	1.5	$\frac{\alpha}{2}+1$	$2-\frac{\alpha}{2}$
3	3.5	3	4
5	2.5	$\frac{\alpha}{2}+2$	$3-\frac{\alpha}{2}$
4	5	5	5
6.5	3	3	3
7.75	5	$\frac{\alpha}{2}+4$	$6-\frac{\alpha}{2}$

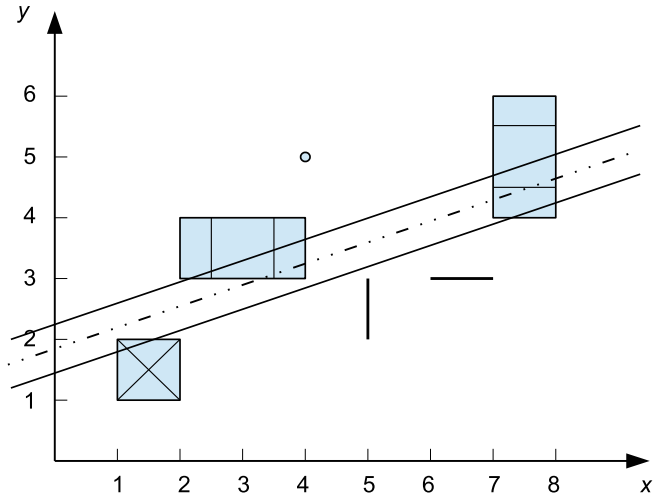
Parametry wyznaczone za pomocą zależności (4.45) mają wartość:

$$w_1 \approx 0.331, \qquad w_0 \approx 1.884.$$

Z tej samej zależności można wyznaczyć parametr r.

$$r = \frac{3}{2 \cdot 6}\left[\int_0^1 (1-\alpha)(2-\frac{\alpha}{2}-\frac{\alpha}{2}-1)\mathrm{d}\alpha + \int_0^1 (1-\alpha)(4-3)\mathrm{d}\alpha \right.$$
$$+ \int_0^1 (1-\alpha)(3-\frac{\alpha}{2}-\frac{\alpha}{2}-2)\mathrm{d}\alpha + \int_0^1 (1-\alpha)\cdot 0\mathrm{d}\alpha + \int_0^1 (1-\alpha)\cdot 0\mathrm{d}\alpha$$
$$\left. + \int_0^1 (1-\alpha)(6-\frac{\alpha}{2}-\frac{\alpha}{2}-4)\mathrm{d}\alpha\right] = 0.5$$

Na rys. 4.8 pokazane są w sposób poglądowy dane i charakterystyka modelu stanowiącego rozwiązanie. Pionowy przekrój charakterystyki jest symetryczną trójkątną liczbą rozmytą o nośniku o szerokości $2r = 1$.



Rysunek 4.8. Poglądowy widok danych wykorzystanych w przykładzie 4.1 i wyznaczone rozwiązanie

4.4.2. Model z wieloma wejściami

Załóżmy, że zmienna zależna y zależy tym razem od N zmiennych objaśniających x_i, $i = 1 \ldots N$. Będziemy poszukiwali liniowego modelu aproksymującego w postaci:

$$\hat{Y} = \mathbf{w}^T \cdot \mathbf{x_s} + R, \tag{4.47}$$

gdzie: $\mathbf{x_s} = [1, x_{1s}, \ldots x_{Ns}]^T$ oznacza wektor środków liczb rozmytych X_i wyznaczonych z zależności (4.41) i będących wejściami modelu, R – trójkątna liczba rozmyta w postaci TFN$(-r, 0, r)$, $\mathbf{w} = [w_0, w_1, \ldots w_N]^T$ – wektor rzeczywistych współczynników modelu.

Model utworzony będzie na podstawie K danych pomiarowych, przy czym zarówno dane wejściowe (zmienne objaśniające) X_{ij}, jak i wyjściowe (zadane wartości zmiennej

zależnej) Y_i mają postać liczb rozmytych, bądź mogą być do takiej postaci sprowadzone ($i = 1 \dots N$, $j = 1 \dots K$). Dane opisane będą macierzą $\mathbf{X}$ i wektorem $\mathbf{Y}$:

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{N1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1K} & X_{2K} & \dots & X_{NK} \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_K \end{bmatrix}. \tag{4.48}$$

Będziemy chcieli, aby model dopasowany był do danych w taki sposób, aby minimalizować kryterium jakości oparte na metryce d_2 opisanej zależnością (4.39):

$$Q(\mathbf{w}, r) = \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) \longrightarrow \min, \tag{4.49}$$

gdzie: $\hat{Y}_i = \mathbf{w}^T \cdot \mathbf{x_{si}} + R$. Oznaczmy przez $\mathbf{F}_Y$ wektor lewych brzegów, a przez $\mathbf{G}_Y$ wektor prawych brzegów liczb rozmytych stanowiących zadane wartości zmiennej zależnej:

$$\mathbf{F}_Y = [f_{Y1}, f_{Y2}, \dots f_{YK}]^T, \qquad \mathbf{G}_Y = [g_{Y1}, g_{Y2}, \dots g_{YK}]^T,$$

oraz przez $\mathbf{X_s}$ macierz środków liczb rozmytych będących elementami macierzy $\mathbf{X}$.

Twierdzenie 4.2. Dla modelu aproksymującego w postaci (4.47) współczynniki minimalizujące kryterium jakości (4.49) można obliczyć z zależności:

$$\mathbf{w} = (\mathbf{X_s}^T \mathbf{X_s})^{-1} \mathbf{X_s}^T \cdot \int_0^1 \frac{\mathbf{F}_Y^{-1}(\alpha) + \mathbf{G}_Y^{-1}(\alpha)}{2} \, \mathrm{d}\alpha, \tag{4.50}$$

$$r = \frac{3}{2K} \cdot \sum_{i=1}^{K} \int_0^1 (1 - \alpha) \left[g_{Yi}^{-1}(\alpha) - f_{Yi}^{-1}(\alpha) \right] \mathrm{d}\alpha.$$

Dowód

Kryterium jakości (4.49) można inaczej zapisać w postaci:

$$\begin{aligned} Q(\mathbf{w}, r) = & \sum_{i=1}^{K} d_2^2(\hat{Y}_i, Y_i) = \sum_{i=1}^{K} \frac{1}{2} \int_0^1 \left[\mathbf{w}^T \cdot \mathbf{x_{si}} + r(\alpha - 1) - f_{Yi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \\ & + \frac{1}{2} \int_0^1 \left[\mathbf{w}^T \cdot \mathbf{x_{si}} + r(1 - \alpha) - g_{Yi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha. \end{aligned}$$

W zapisie macierzowym będzie ono miało postać:

$$Q(\mathbf{w}, r) = \frac{1}{2} \int_0^1 \left[\mathbf{X_s w} + \mathbf{r}(\alpha - 1) - \mathbf{F}_Y^{-1}(\alpha)\right]^T \left[\mathbf{X_s w} + \mathbf{r}(\alpha - 1) - \mathbf{F}_Y^{-1}(\alpha)\right] + \left[\mathbf{X_s w} + \mathbf{r}(1 - \alpha) - \mathbf{G}_Y^{-1}(\alpha)\right]^T \left[\mathbf{X_s w} + \mathbf{r}(1 - \alpha) - \mathbf{G}_Y^{-1}(\alpha)\right] \mathrm{d}\alpha,$$

gdzie: $\mathbf{r}$ – kolumnowy wektor K takich samych wartości r. Po wymnożeniu i uproszczeniu wyrażenia otrzymujemy:

$$\begin{aligned} Q(\mathbf{w}, r) = \frac{1}{2} \int_0^1 & \mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w} - (\alpha - 1)\mathbf{r}^T \mathbf{X_s w} - \mathbf{F}_Y^{-1}(\alpha)^T \mathbf{X_s w} + \mathbf{w}^T \mathbf{X_s}^T \mathbf{r}(\alpha - 1) \\ & + \mathbf{r}^T \mathbf{r}(\alpha - 1)^2 - \mathbf{F}_Y^{-1}(\alpha)^T \mathbf{r}(\alpha - 1) - \mathbf{w}^T \mathbf{X_s}^T \mathbf{F}_Y^{-1}(\alpha) \\ & - \mathbf{r}^T \mathbf{F}_Y^{-1}(\alpha)(\alpha - 1) + \mathbf{F}_Y^{-1}(\alpha)^T \mathbf{F}_Y^{-1}(\alpha) + \mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w} \\ & - (1 - \alpha)\mathbf{r}^T \mathbf{X_s w} - \mathbf{G}_Y^{-1}(\alpha)^T \mathbf{X_s w} + \mathbf{w}^T \mathbf{X_s}^T \mathbf{r}(1 - \alpha) + \mathbf{r}^T \mathbf{r}(1 - \alpha)^2 \\ & - \mathbf{G}_Y^{-1}(\alpha)^T \mathbf{r}(1 - \alpha) - \mathbf{w}^T \mathbf{X_s}^T \mathbf{G}_Y^{-1}(\alpha) - \mathbf{r}^T \mathbf{G}_Y^{-1}(\alpha)(1 - \alpha) \\ & + \mathbf{G}_Y^{-1}(\alpha)^T \mathbf{G}_Y^{-1}(\alpha)\, \mathrm{d}\alpha \\ = \frac{1}{2} \int_0^1 & 2\mathbf{w}^T \mathbf{X_s}^T \mathbf{X_s w} - 2[\mathbf{F}_Y^{-1}(\alpha)^T + \mathbf{G}_Y^{-1}(\alpha)^T] \mathbf{X_s w} + 2\mathbf{r}^T \mathbf{r}(\alpha - 1)^2 \\ & - 2[\mathbf{G}_Y^{-1}(\alpha)^T - \mathbf{F}_Y^{-1}(\alpha)^T]\mathbf{r}(1 - \alpha) + \mathbf{F}_Y^{-1}(\alpha)^T \mathbf{F}_Y^{-1}(\alpha) + \mathbf{G}_Y^{-1}(\alpha)^T \mathbf{G}_Y^{-1}(\alpha)\, \mathrm{d}\alpha. \end{aligned}$$

Aby wyznaczyć minimum, liczymy pochodne cząstkowe po parametrach modelu i przyrównujemy je do zera.

$$\frac{\partial Q(\mathbf{w}, r)}{\partial \mathbf{w}} = \int_0^1 2\mathbf{X_s}^T \mathbf{X_s w} - \mathbf{X_s}^T [\mathbf{F}_Y^{-1}(\alpha) + \mathbf{G}_Y^{-1}(\alpha)]\, \mathrm{d}\alpha = 0$$

$$2\mathbf{X_s}^T \mathbf{X_s w} = \mathbf{X_s}^T \int_0^1 [\mathbf{F}_Y^{-1}(\alpha) + \mathbf{G}_Y^{-1}(\alpha)]\, \mathrm{d}\alpha$$

Możemy teraz wyznaczyć wektor parametrów:

$$\mathbf{w} = (\mathbf{X_s}^T \mathbf{X_s})^{-1} \mathbf{X_s}^T \cdot \int_0^1 \frac{\mathbf{F}_Y^{-1}(\alpha) + \mathbf{G}_Y^{-1}(\alpha)}{2}\, \mathrm{d}\alpha.$$

Kryterium jakości można inaczej zapisać w postaci:

$$Q(\mathbf{w},r) = \frac{1}{2}\int_0^1 2\mathbf{w}^T{\mathbf{X_s}}^T\mathbf{X_s}\mathbf{w} - 2[\mathbf{F}_Y^{-1}(\alpha)^T + \mathbf{G}_Y^{-1}(\alpha)^T]\mathbf{X_s}\mathbf{w} + 2Kr^2(\alpha-1)^2 - 2\sum_{i=1}^{K}[g_{Yi}^{-1}(\alpha) - f_{Yi}^{-1}(\alpha)]r(1-\alpha) + \mathbf{F}_Y^{-1}(\alpha)^T\mathbf{F}_Y^{-1}(\alpha) + \mathbf{G}_Y^{-1}(\alpha)^T\mathbf{G}_Y^{-1}(\alpha)\,\mathrm{d}\alpha.$$

Liczymy kolejną pochodną cząstkową i przyrównujemy ją do zera.

$$\frac{\partial Q(\mathbf{w},r)}{\partial r} = 2Kr\int_0^1(\alpha-1)^2\mathrm{d}\alpha - \int_0^1\sum_{i=1}^{K}[g_{Yi}^{-1}(\alpha) - f_{Yi}^{-1}(\alpha)](1-\alpha)\,\mathrm{d}\alpha = 0$$

Otrzymujemy ostatecznie zależność na parametr r:

$$r = \frac{3}{2K}\cdot\sum_{i=1}^{K}\int_0^1(1-\alpha)\left[g_{Yi}^{-1}(\alpha) - f_{Yi}^{-1}(\alpha)\right]\mathrm{d}\alpha.$$

Badamy dalej warunek dostateczny. Aby kryterium (4.49), w skrócie Q, osiągało minimum w punkcie opisanym równaniami (4.50), jego Hessian w tym punkcie musi być dodatnio określony. Ma on postać:

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 Q}{\partial\mathbf{w}\partial\mathbf{w}^T} & \dfrac{\partial^2 Q}{\partial\mathbf{w}\partial r} \\ \dfrac{\partial^2 Q}{\partial r\partial\mathbf{w}} & \dfrac{\partial^2 Q}{\partial r^2} \end{bmatrix} = \begin{bmatrix} 2{\mathbf{X_s}}^T\mathbf{X_s} & 0 \\ 0 & \frac{2}{3}K \end{bmatrix}. \tag{4.51}$$

Obliczamy wyznaczniki macierzy.

$$\Delta_1 = 2{\mathbf{X_s}}^T\mathbf{X_s}$$

$$\Delta_2 = \frac{4}{3}K\cdot{\mathbf{X_s}}^T\mathbf{X_s}$$

Ponieważ macierz ${\mathbf{X_s}}^T\mathbf{X_s}$ jest dodatnio określona jeśli tylko jest nieosobliwa, warunek dodatniej określoności Hessianu jest spełniony.

□

Przykład 4.2.

Dla danych z tabeli należy wyznaczyć model aproksymujący w postaci (4.47), minimalizujący kryterium jakości (4.49).

X_1	X_2	Y
TFN(1,1.5,2)	TFN(5,5.5,6)	TFN(1,1.5,2)
TrFN(2,2.5,3.5,4)	RFN(4,6)	RFN(3,4)
5	3	TFN(2,2.5,3)
4	TrFN(1,2,3,4)	5
RFN(6,7)	TFN(1,1.5,2)	3
TFN(7,8,8)	TFN(0,1,2)	TrFN(4,4.5,5.5,6)

Parametry wyznaczone za pomocą zależności (4.50) mają wartość:

$$\mathbf{w}^T \approx [7.162, -0.275, -0.802], \qquad r = 0.5.$$

Dane zadane na wyjściu modelu są tu takie same jak w przykładzie 4.1. Ponieważ niepewność modelu r zależy tylko od niepewności wartości wyjścia zadanego Y, podobnie jak w przykładzie poprzednim, pionowy przekrój charakterystyki jest symetryczną trójkątną liczbą rozmytą o nośniku o szerokości $2r = 1$.

4.5. Modelowanie lokalne dla danych rozmytych

Podobnie jak w przypadku interwałów, sposób modelowania lokalnego opisany w podrozdziale 2.1 będzie wymagał modyfikacji, ponieważ zarówno punkt wejściowy $\mathbf{x}^*$, jak i dane uczące będą liczbami rozmytymi. W przypadku gdy któraś z analizowanych wartości będzie liczbą rzeczywistą, zastąpimy ją jednoelementową liczbą rozmytą (typu singleton). Jeśli wartość będzie interwałem, możemy ją zastąpić prostokątną liczbą rozmytą.

Aby wyznaczyć najbliższych sąsiadów punktu wejściowego $\mathbf{x}^*$, będziemy badać odległość między wielowymiarowymi danymi rozmytymi. Wyznaczymy ją jako:

$$D(\mathbf{x}, \mathbf{x}^*) = \sqrt{\sum_{i=1}^{N} d_2^2(x_i, x_i^*)}\,, \tag{4.52}$$

gdzie: $\mathbf{x}, \mathbf{x}^*$ – wektory liczb rozmytych, N – rozmiar wektora danych, $d_2(x_i, x_i^*)$ – odległość między liczbami rozmytymi wyznaczona z zależności (4.39).

Modelowanie lokalne dla liczb rozmytych bazujące na metodach kNN, może być zrealizowane identycznie jak dla danych rzeczywistych z wykorzystaniem zasad arytmetyki rozmytej, wzór (4.7). Można w ten sposób wyznaczyć średnią arytmetyczną bądź średnią ważoną z wyjść zadanych dla próbek będących sąsiadami punktu wejściowego $\mathbf{x}^*$.

W przypadku mini-modelu liniowego, będziemy poszukiwali regresji:

$$f(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{x_s}^* + R, \tag{4.53}$$

w sposób opisany w podrozdziale 4.4. Dla mini-modelu nieliniowego regresja będzie miała postać:

$$f(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{x_s}^* + R + f_N(\mathbf{x_s}^*), \tag{4.54}$$

przy czym składnik nieliniowy wyznaczany będzie dla środków liczb rozmytych, stanowiących dane uczące i środków liczb rozmytych stanowiących punkt wejściowy $\mathbf{x}^*$.

4.5.1. Dokładność modelu

Załóżmy, podobnie jak w rozdziale poprzednim, że mamy do dyspozycji 2 znormalizowane, rozłączne zbiory danych. Pierwszy z nich będzie zawierał dane uczące (generujące modele lokalne), a każda próbka składać się będzie z wektora wejściowych liczb rozmytych $\mathbf{x}_j$ i zadanej, rozmytej wartości wyjściowej y_j, $j = 1 \dots L$. Drugi zbiór danych zawierał będzie dane walidujące i podobnie, każda próbka składać się będzie z wektora wejściowych liczb rozmytych $\mathbf{x}_i$ i zadanej, rozmytej wartości wyjściowej y_i, $i = 1 \dots M$.

Miarą dokładności modelu może być jego średni błąd bezwzględny, wyznaczony dla danych walidujących:

$$e_{MAE} = \frac{1}{M} \cdot \sum_{i=1}^{M} d_2(y_i, y_i^*), \tag{4.55}$$

gdzie: y_i – to wyjście zadane dla próbki i, a y_i^* – to wyjście wyznaczone przez model. Wyznaczony w ten sposób błąd będzie liczbą rzeczywistą.

Jeżeli nie jesteśmy w stanie wyodrębnić z posiadanych danych rozłącznej części próbek do walidacji jakości modelu, możemy zastosować kroswalidację (walidację krzyżową). Wyznaczanie błędu kroswalidacji realizowane jest podobnie jak dla danych walidujących na podstawie wzoru (4.55).

4.6. Eksperymenty

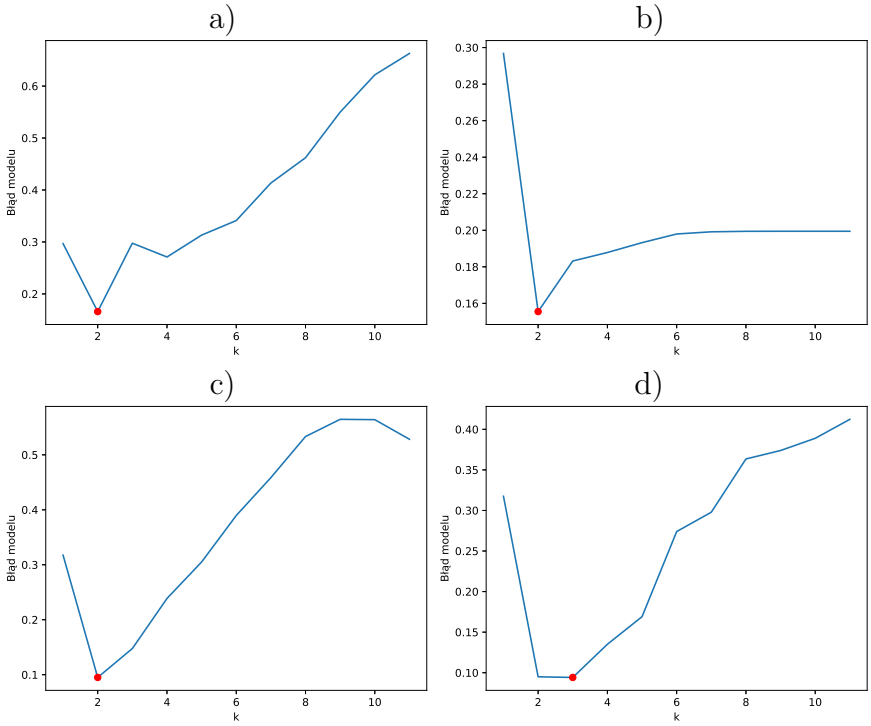
Badania nr 1 – jednorodne dane jednowejściowe

Podobnie jak w przypadku interwałów, w pierwszym eksperymencie wykorzystane zostaną proste dane jednowejściowe. Jego celem będzie przede wszystkim zbadanie poprawności metod opisanych we wcześniejszych podrozdziałach.

Dane opisują sinusoidę próbkowaną co 0.5 dla wejścia $x \in [0, 2\pi]$ i składać się będą z 13 próbek pokazanych na rys. 3.11a. Dane poddane zostały rozmyciu, przy czym

zabiegowi temu podlegał zarówno atrybut wejściowy, jak i wyjściowy danych. W trakcie rozmycia, każda rzeczywista wartość x była zastępowana trójkątną liczbą rozmytą $\mathrm{TFN}(x-\alpha, x, x+\alpha)$. W badaniach przyjęto $\alpha = 0.1$.

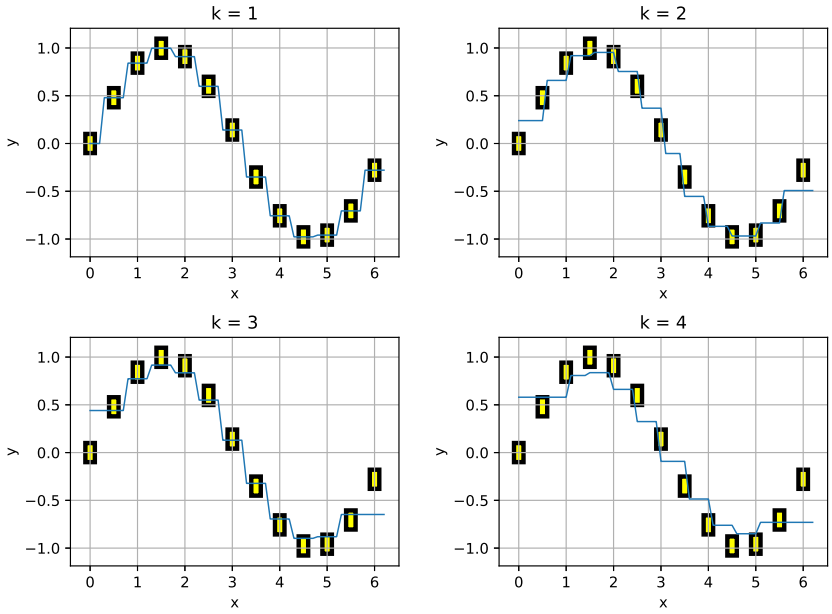
Na rys. 4.9 pokazano jak zmienia się bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla różnych metod modelowania. Wynika z niego, że optymalna liczba dla metod kNN i modelu bazującego na mini-modelach liniowych wynosi $k = 2$. Dla modelu bazującego na mini-modelach nieliniowych wartość ta wynosi $k = 3$.



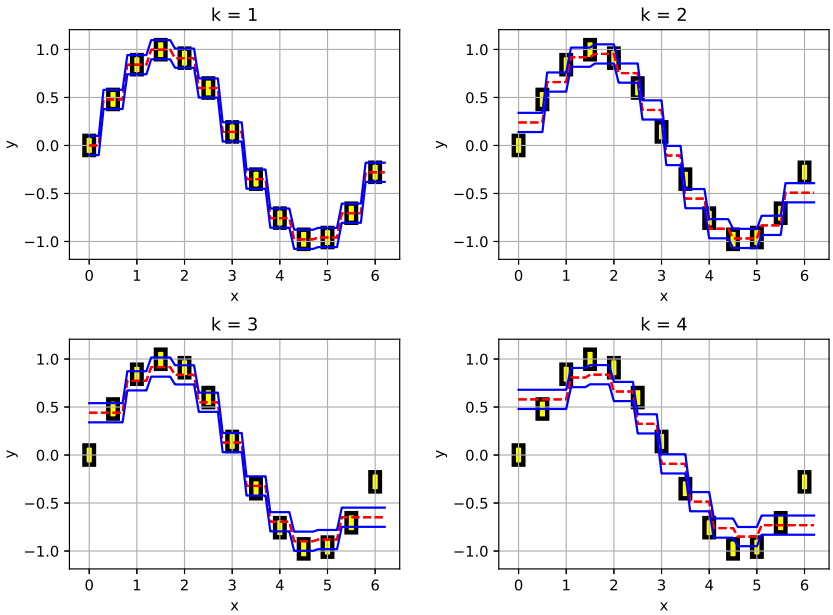
Rysunek 4.9. Bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla: prostej metody kNN (a) ważonej metody kNN (b), dla modelu bazującego na mini-modelach liniowych (c) i dla modelu bazującego na mini-modelach nieliniowych (d)

Na rysunkach od 4.10 do 4.17 pokazane są charakterystyki modeli bazujących na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Charakterystyki sporządzone zostały dla różnych ilości najbliższych sąsiadów k, przy czym dla metod kNN parametr k przyjmował wartości od 1 do 4, a dla metod bazujących na mini-modelach wartości 3 i 4 (ze względu na minimalną liczbę próbek wymaganą przy wyznaczaniu regresji).

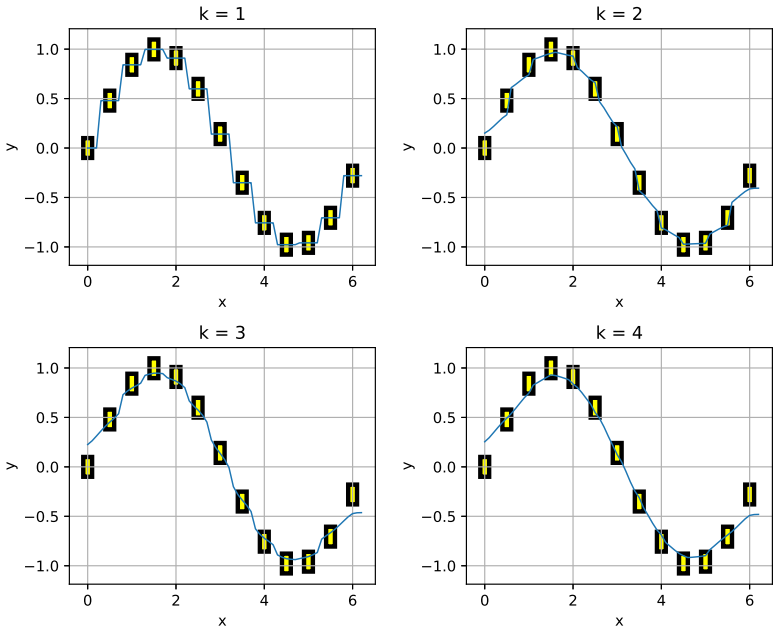
W tabeli 4.1 podany jest średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzoru (4.55), przy założeniu $k = 4$.



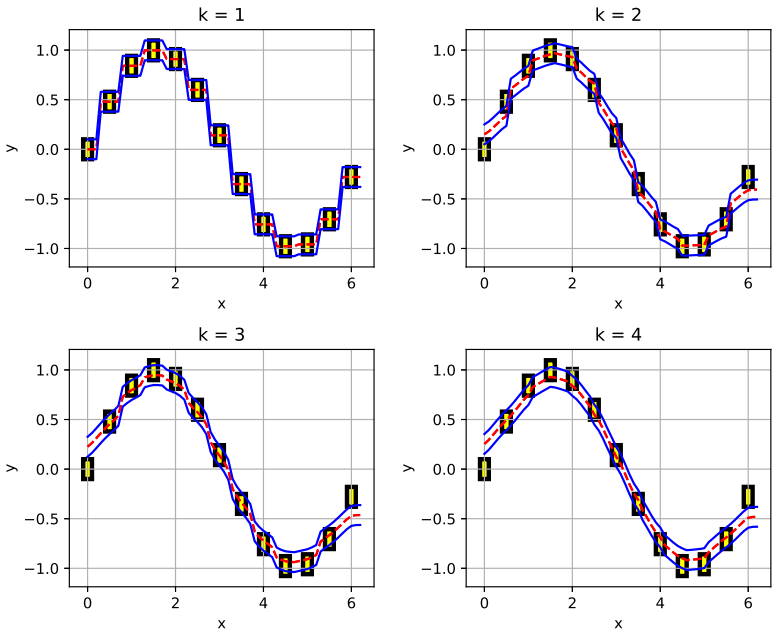
Rysunek 4.10. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



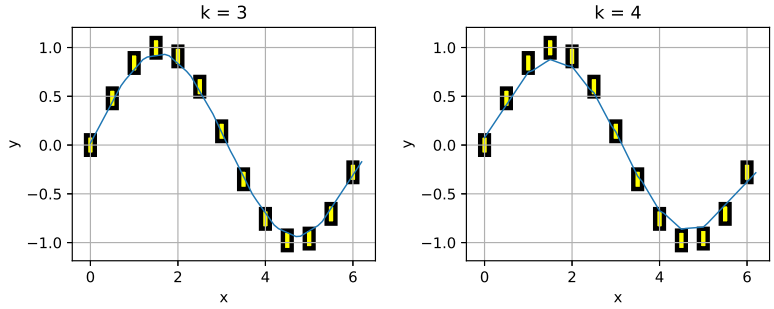
Rysunek 4.11. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



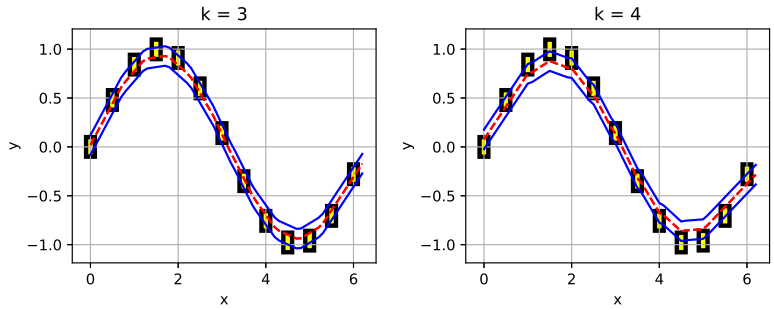
Rysunek 4.12. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



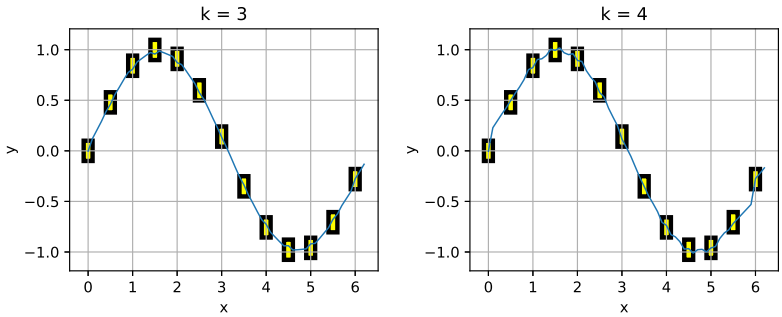
Rysunek 4.13. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



Rysunek 4.14. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



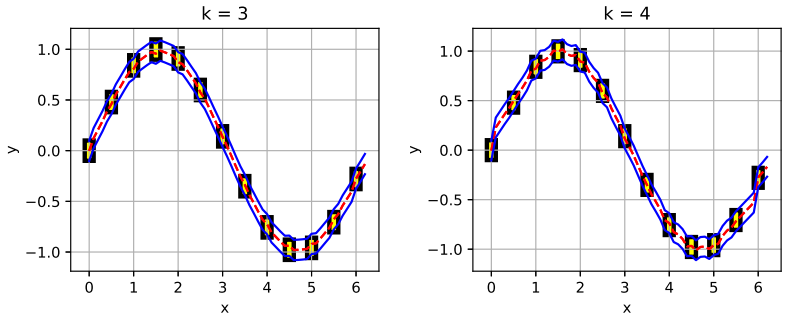
Rysunek 4.15. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



Rysunek 4.16. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej

Tabela 4.1. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 1 przy założeniu $k = 4$

model	e_{MAE}
kNN – średnia arytmetyczna	0.271
kNN – średnia ważona	0.188
mini-model liniowy	0.238
mini-model nieliniowy	0.135



Rysunek 4.17. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia

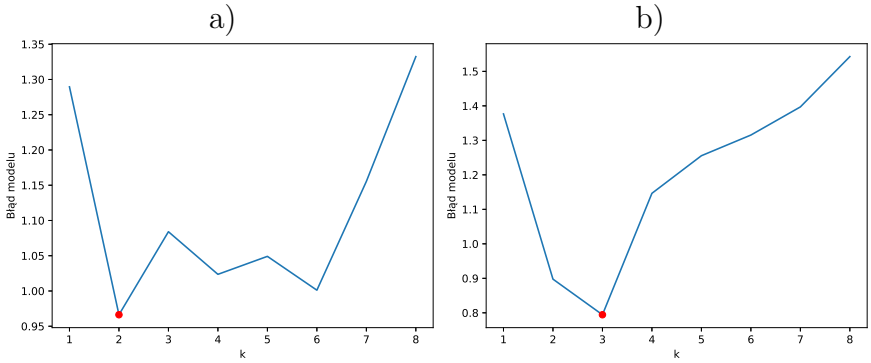
Badania nr 2 – mieszane dane jednowejściowe

W kolejnym eksperymencie zbadano jak działają modele utworzone na podstawie mieszanych danych jednowejściowych. Dane uczące zostały przygotowane tak, aby zapewnić jak największą różnorodność, stąd zarówno atrybuty wejściowe jak i wyjściowy przyjmowały postać liczb rozmytych, ale także interwałów i liczb rzeczywistych. Dane zostały przedstawione w tabeli 4.2. W poglądowy sposób pokazane są także na rys. 4.19 i kolejnych.

Tabela 4.2. Dane uczące wykorzystane w badaniach nr 2

wejście x	wyjście y
0	0
1	1
$[2, 2.5]$	$[2.5, 3]$
3	$\mathrm{TFN}(1, 1.5, 2)$
4	0
$\mathrm{TFN}(4.5, 5, 5.5)$	$[-0.5, 0]$
6	-1
$[7, 7.5]$	-2
8	$\mathrm{TrFN}(-4, -4, -3.5, -3)$
$\mathrm{TrFN}(8, 8.5, 9, 9.5)$	$[-1.5, -1]$
10	0

Na rys. 4.18 pokazano jak zmienia się bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla prostej metody kNN i dla modelu bazującego na mini-modelach liniowych. Wynika z niego, że optymalna liczba dla metody kNN wynosi $k = 2$, a dla modelu bazującego na mini-modelach liniowych wartość ta wynosi $k = 3$.



Rysunek 4.18. Bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla: prostej metody kNN (a) i dla modelu bazującego na mini-modelach liniowych (b)

Na rysunkach od 4.19 do 4.26 pokazano charakterystyki modeli bazujących na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Charakterystyki sporządzone zostały dla różnych ilości najbliższych sąsiadów k, przy czym dla metod kNN parametr k przyjmował wartości od 1 do 4, a dla metod bazujących na mini-modelach wartości 3 i 4 (ze względu na minimalną liczbę próbek wymaganą przy wyznaczaniu regresji).

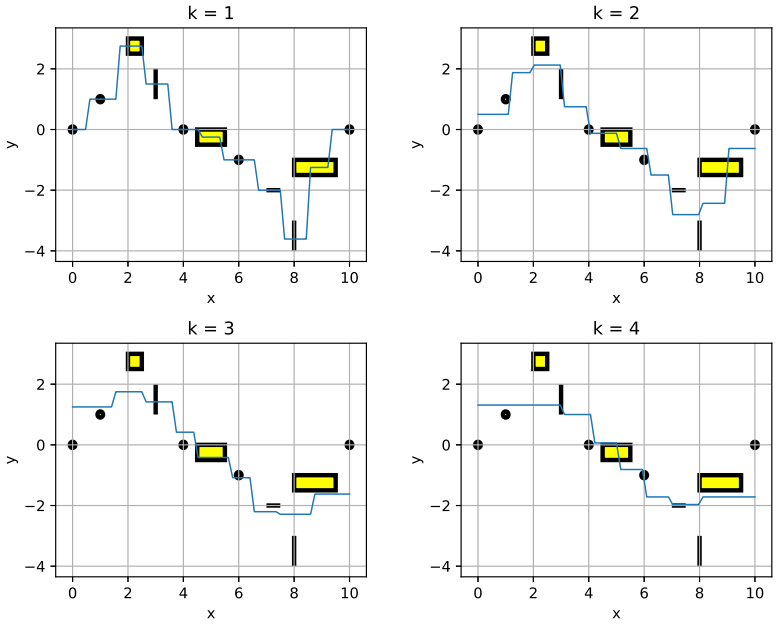
W tabeli 4.3 podany jest średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzoru (4.55), przy założeniu $k = 3$. W tabelach 4.4 i 4.5 pokazano odpowiedzi modeli dla przykładowego wejścia rzeczywistego i rozmytego.

Tabela 4.3. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 2 przy założeniu $k = 3$

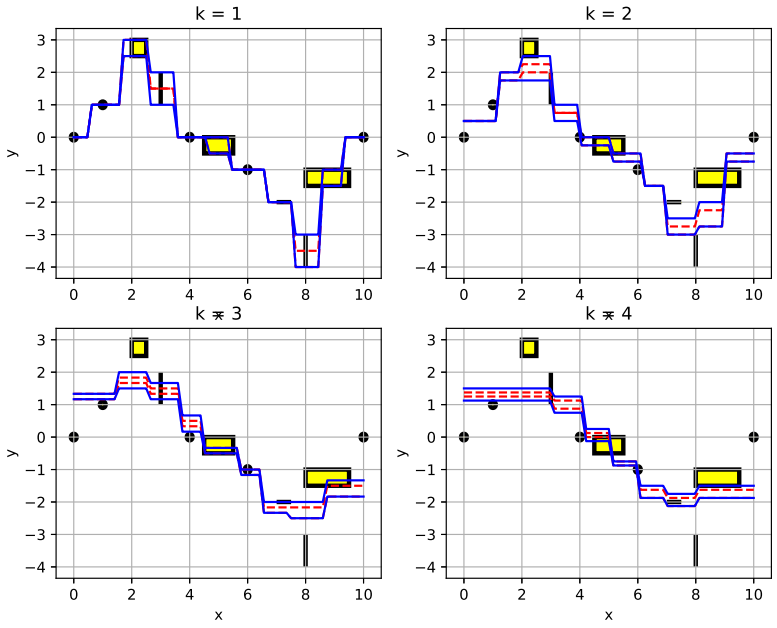
model	e_{MAE}
kNN – średnia arytmetyczna	1.084
kNN – średnia ważona	0.945
mini-model liniowy	0.794
mini-model nieliniowy	0.850

Tabela 4.4. Wartości odpowiedzi modeli dla przykładowego wejścia $x^* = 3.0$ przy założeniu $k = 3$

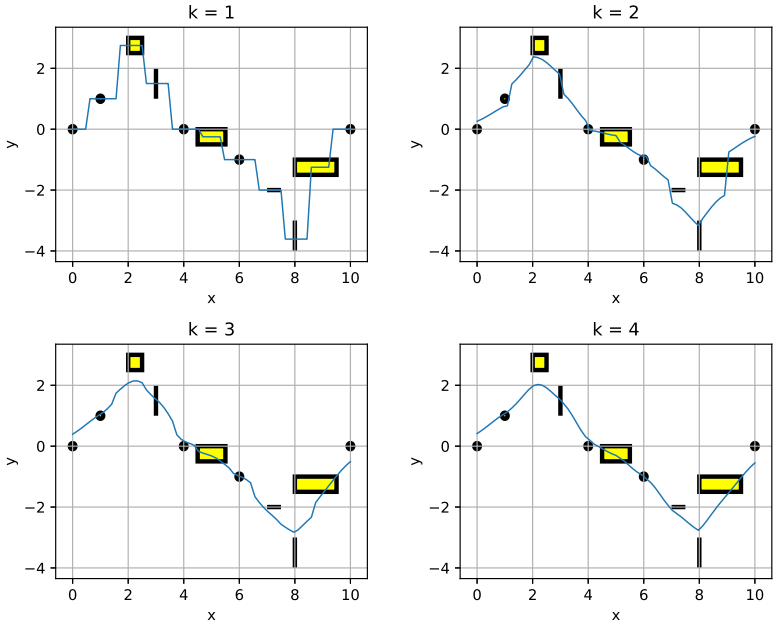
model	wyjście y
kNN – średnia arytmetyczna	TrFN(1.17,1.33,1.5,1.67)
kNN – średnia ważona	TrFN(1.16,1.49,1.59,1.92)
mini-model liniowy	TFN(1.26,1.55,1.84)
mini-model nieliniowy	TFN(1.22,1.51,1.80)



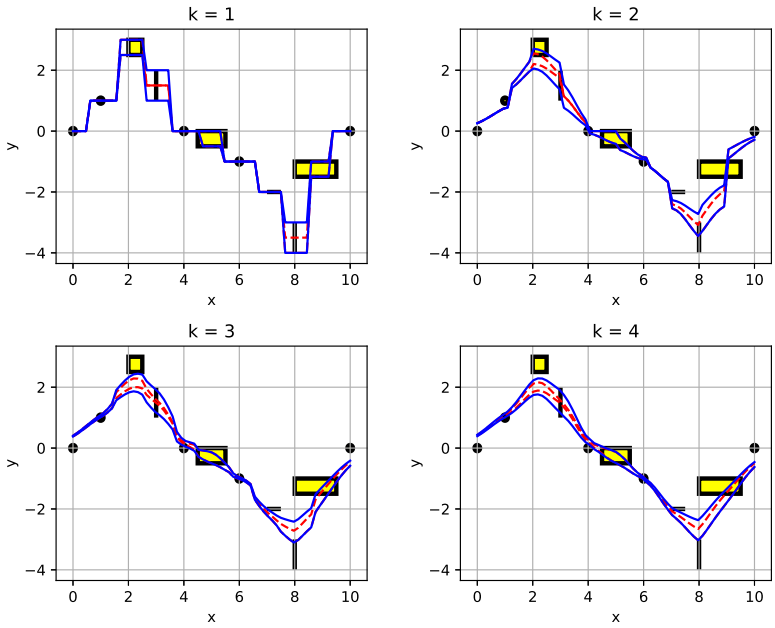
Rysunek 4.19. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



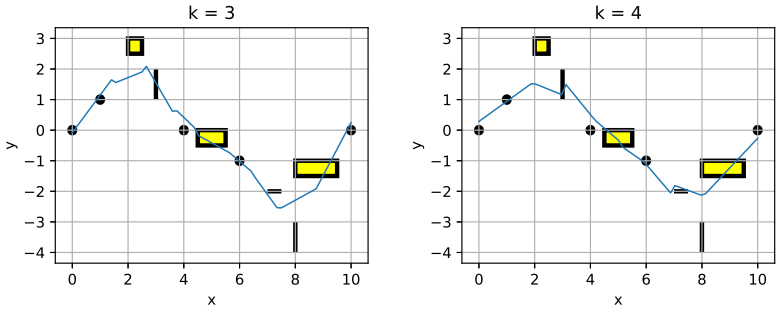
Rysunek 4.20. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linie przerywane szerokość jej rdzenia



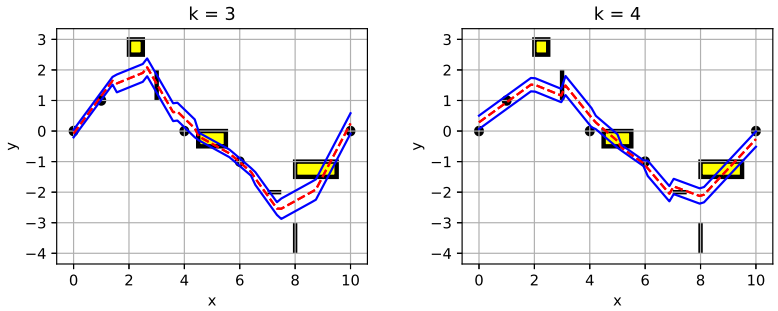
Rysunek 4.21. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



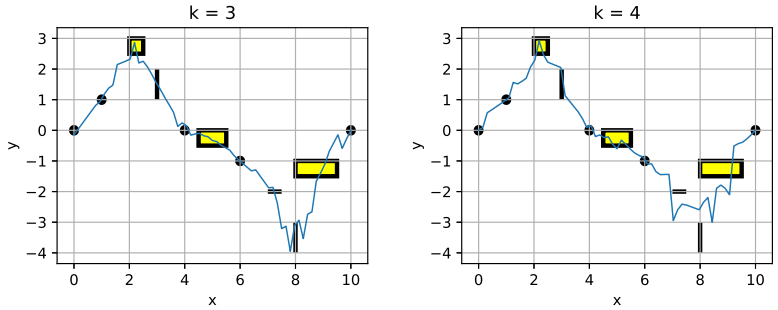
Rysunek 4.22. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linie przerywane szerokość jej rdzenia



Rysunek 4.23. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



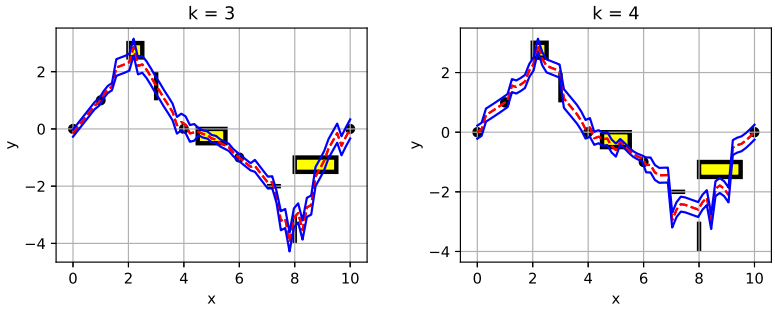
Rysunek 4.24. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



Rysunek 4.25. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej

Tabela 4.5. Wartości odpowiedzi modeli dla przykładowego rozmytego wejścia $x^* =$ TrFN(2.5,3,3.5,3.5) przy założeniu $k = 3$

model	wyjście y
kNN – średnia arytmetyczna	TrFN(1.17,1.33,1.5,1.67)
kNN – średnia ważona	TrFN(1.15, 1.42, 1.54, 1.81)
mini-model liniowy	TFN(1.06,1.35,1.64)
mini-model nieliniowy	TFN(1.03,1.32,1.61)



Rysunek 4.26. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia

Badania nr 3 – rozmyty system 2-wejściowy

Kolejny eksperyment pokazuje, że metody modelowania bazujące na regresji lokalnej mogą być z powodzeniem wykorzystane do wyznaczania odpowiedzi modeli rozmytych.

Załóżmy, że dany jest 2-wejściowy system rozmyty. Reguły modelu przedstawiono w tabeli 4.6, a funkcje przynależności wszystkich pojęć w niej występujących pokazano na rys. 4.27.

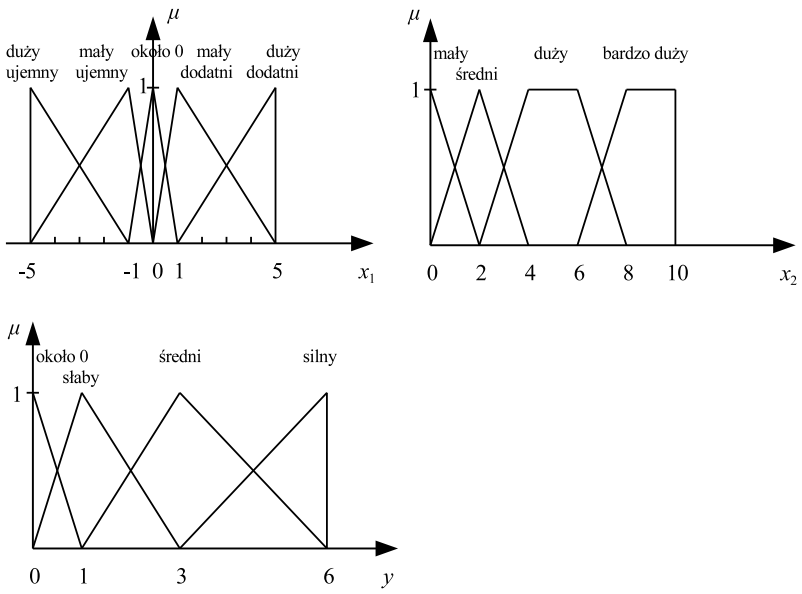
Tabela 4.6. Reguły systemu rozmytego wykorzystanego w badaniach nr 3

x_1 / x_2	duży ujemny	mały ujemny	około 0	mały dodatni	duży dodatni
mały	średni	słaby	około 0	około 0	słaby
średni	średni	słaby	słaby	słaby	średni
duży	silny	średni	średni	średni	silny
bardzo duży	silny	silny	średni	silny	silny

Tabela 4.6 może być potraktowana jako zbiór danych uczących przyjmujących postać liczb rozmytych. Można ją inaczej przedstawić w postaci tabeli 4.7 i wykorzystać do utworzenia modeli.

Na rys. 4.28 pokazano jak zmienia się bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla prostej metody kNN i dla modelu bazującego na mini-modelach liniowych. Wynika z niego, że optymalna liczba dla metody kNN wynosi $k = 4$, a dla modelu bazującego na mini-modelach liniowych wartość ta wynosi $k = 7$.

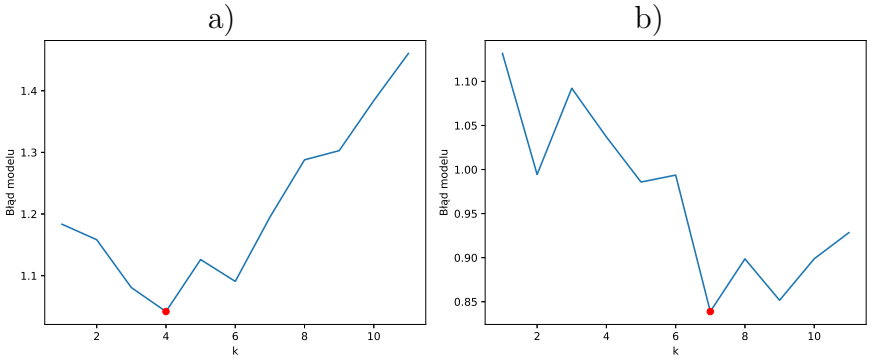
Na rysunkach od 4.29 do 4.32 pokazane są charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Charakterystyki sporządzone zostały dla różnych ilości najbliższych sąsiadów k, przy czym dla metod kNN parametr k przyjmował wartości od 2, 4, 6 i 8, a dla metod bazujących



Rysunek 4.27. Funkcje przynależności pojęć występujących w tabeli 4.6

Tabela 4.7. Dane uczące wykorzystane w badaniach nr 3

x_1	x_2	y
duży ujemny	mały	średni
duży ujemny	średni	średni
duży ujemny	duży	silny
duży ujemny	bardzo duży	silny
mały ujemny	mały	słaby
mały ujemny	średni	słaby
mały ujemny	duży	średni
mały ujemny	bardzo duży	silny
około 0	mały	około 0
około 0	średni	słaby
około 0	duży	średni
około 0	bardzo duży	średni
mały dodatni	mały	około 0
mały dodatni	średni	słaby
mały dodatni	duży	średni
mały dodatni	bardzo duży	silny
duży dodatni	mały	słaby
duży dodatni	średni	średni
duży dodatni	duży	silny
duży dodatni	bardzo duży	silny



Rysunek 4.28. Bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla: prostej metody kNN (a) i dla modelu bazującego na mini-modelach liniowych (b)

na mini-modelach wartości 4, 6, 8 i 10 (wartości większe ze względu na minimalną liczbę próbek wymaganą przy wyznaczaniu regresji).

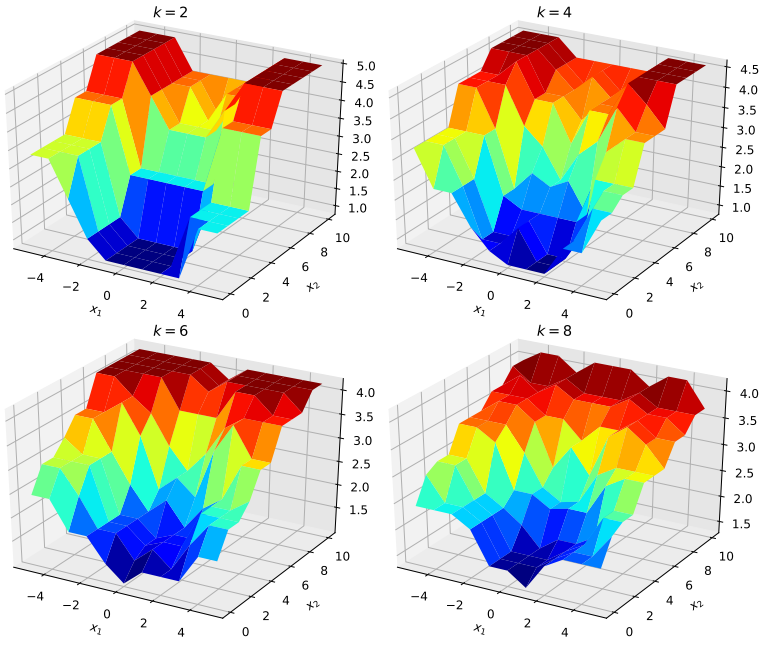
Dla porównania na rys. 4.33 pokazana jest charakterystyka systemu rozmytego otrzymana w klasyczny (znany z literatury, np. [86]) sposób, z wykorzystaniem wnioskowania bazującego na operatorach min-max i ostrzenia metodą wysokości.

Najważniejsze zalety metod modelowania bazujących na regresji lokalnej, wykorzystanych do wyznaczania odpowiedzi systemu rozmytego, wymieniono poniżej.

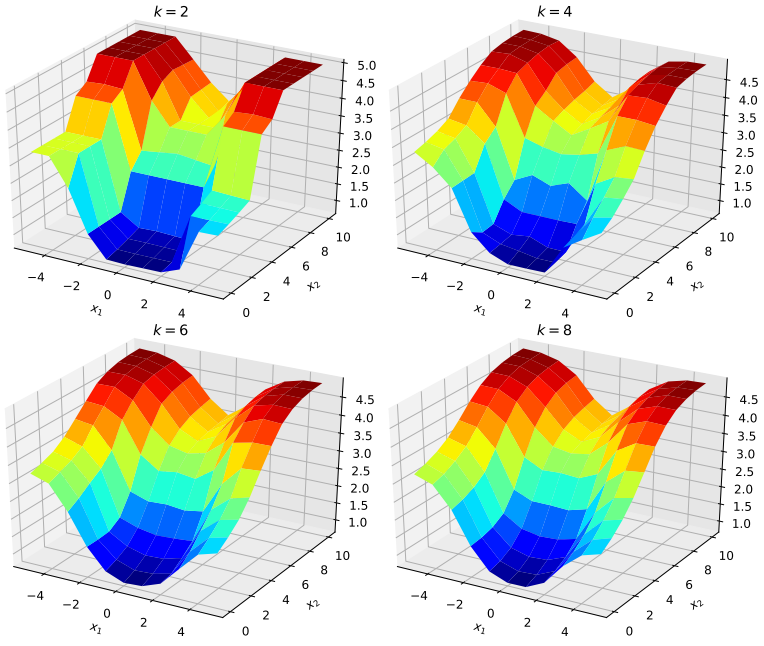
- Po pierwsze, baza reguł nie musi być kompletna. Konkluzje niektórych reguł w tabeli mogą być trudne do określenia, jednak nawet w przypadku gdy brakuje części reguł w modelu, metody bazujące na regresji lokalnej skutecznie potrafią wyznaczyć jego wyjście.
- Baza reguł nie musi być spójna. W przypadku reguł o sprzecznych konkluzjach, model uśredni je wyznaczając odpowiedź.
- Użytkownik może sam założyć ile reguł modelu wykorzystywanych jest przy wyznaczaniu wyjścia, przy czym łatwo może wyznaczyć ilość optymalną dla danej metody modelowania. W klasycznym modelowaniu rozmytym, liczba wykorzystywanych reguł jest stała i zależna od przyjętej metody wyznaczania wyjścia.

W tabeli 4.8 podany jest średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzoru (4.55), przy założeniu $k = 4$ i $k = 7$.

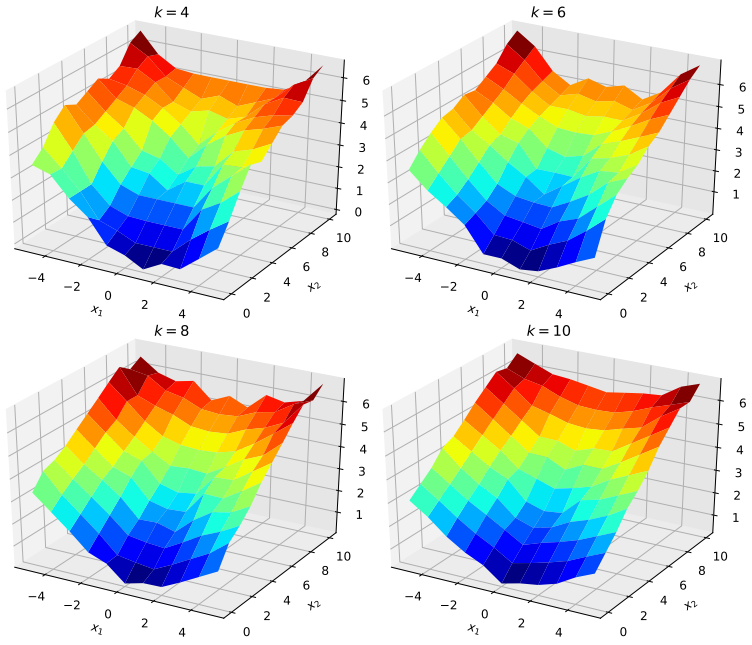
W tabelach 4.9 i 4.10 pokazane zostały odpowiedzi modeli dla przykładowego wejścia rzeczywistego i rozmytego.



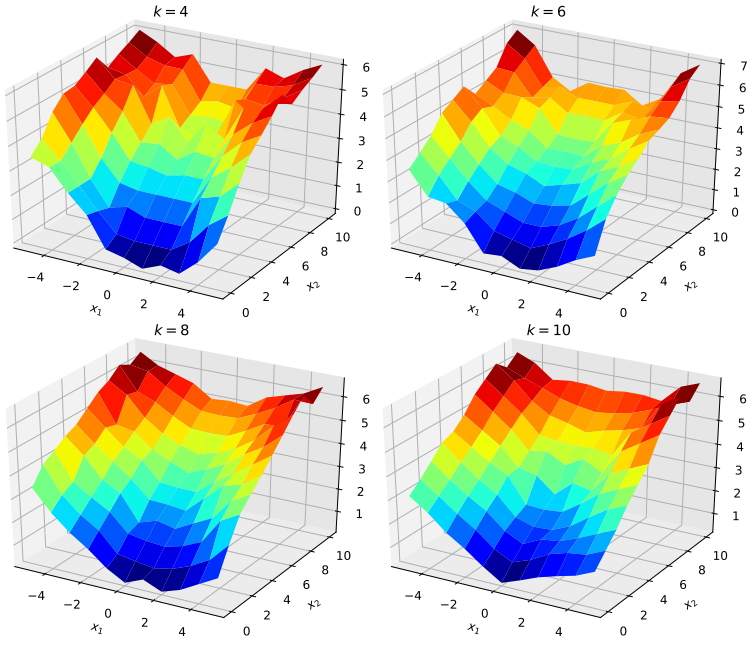
Rysunek 4.29. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Powierzchnia reprezentuje położenie środka ciężkości wynikowej liczby rozmytej



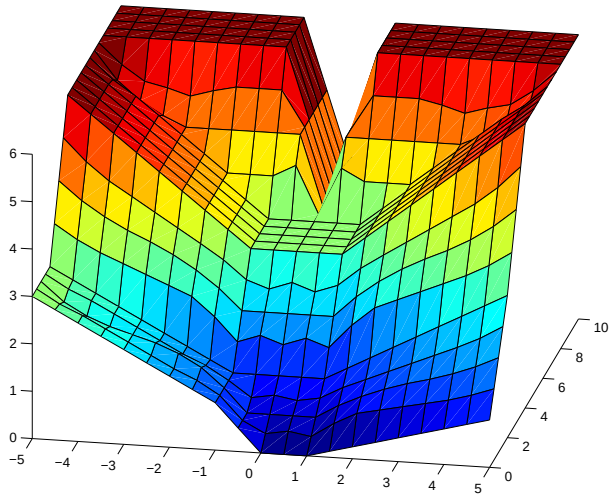
Rysunek 4.30. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Powierzchnia reprezentuje położenie środka ciężkości wynikowej liczby rozmytej



Rysunek 4.31. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Powierzchnia reprezentuje położenie środka ciężkości wynikowej liczby rozmytej



Rysunek 4.32. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Powierzchnia reprezentuje położenie środka ciężkości wynikowej liczby rozmytej



Rysunek 4.33. Charakterystyka systemu rozmytego wyznaczona z wykorzystaniem wnioskowania bazującego na operatorach min-max i ostrzenia metodą wysokości

Tabela 4.8. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 3

model	e_{MAE} dla $k = 4$	e_{MAE} dla $k = 7$
kNN – średnia arytmetyczna	1.042	1.195
kNN – średnia ważona	0.925	0.932
mini-model liniowy	1.036	0.838
mini-model nieliniowy	1.056	0.806

Tabela 4.9. Wartości odpowiedzi modeli dla przykładowego wejścia $\mathbf{x}^* = [-2.0, 2.5]^T$ przy założeniu $k = 4$

model	wyjście y
kNN – średnia arytmetyczna	TFN(0.25,1.5,3.75)
kNN – średnia ważona	TFN(0.13,1.27,3.40)
mini-model liniowy	TFN(0.22,1.97,3.72)
mini-model nieliniowy	TFN(0.08,1.83,3.58)

Tabela 4.10. Wartości odpowiedzi modeli dla przykładowego rozmytego wejścia $\mathbf{x}^* = [\text{RFN}(2,5), \text{TFN}(2.5,3,3.5)]^T = [\text{'większy od 2'},\text{'około 3'}]^T$ przy założeniu $k = 4$

model	wyjście y
kNN – średnia arytmetyczna	TFN(1.0,2.75,4.5)
kNN – średnia ważona	TFN(1.08,2.94,4.89)
mini-model liniowy	TFN(1.58,3.33,5.08)
mini-model nieliniowy	TFN(1.72,3.47,5.22)

Badania nr 4 – mieszane dane 4-wejściowe

Celem ostatniego eksperymentu będzie sprawdzenie, w jaki sposób metody poradzą sobie z mieszanymi danymi wielowejściowymi. Przykładem może tu być wycena wartości nieruchomości.

Załóżmy, że dla każdego z 15 obiektów mamy następujące informacje: wiek (w latach), standard wykończenia (w skali od 0 do 10 punktów), powierzchnia (w m^2), odległość od centrum (w kilometrach) i cenę jednego metra kwadratowego (w tys. zł za m^2). Część informacji znamy dokładnie, a część w przybliżeniu: w formie interwału lub formie pojęcia lingwistycznego, reprezentowanego przez liczbę rozmytą. Na podstawie posiadanej wiedzy, będziemy próbowali wyceniać kolejne nieruchomości. Dane przedstawiono w tabeli 4.11, przy czym w kolejnych eksperymentach ich atrybuty wejściowe były normalizowane, ponieważ rozpiętości ich dziedzin mocno się różniły.

Na rys. 4.34 pokazano jak zmienia się bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla prostej i ważonej metody kNN oraz dla modelu bazującego na mini-modelach liniowych i nieliniowych. Wynika z niego, że optymalna liczba dla metod kNN wynosi $k = 5$, a dla modelu bazującego na mini-modelach wartość ta wynosi $k = 13$.

W tabeli 4.12 podano średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzoru (4.55), przy założeniu, że liczba uwzględnianych próbek zmienia się od $k = 5$ do $k = 13$.

W tabelach 4.13 i 4.14 pokazano odpowiedzi modeli dla przykładowego wejścia rzeczywistego i rozmytego.

4.7. Wnioski

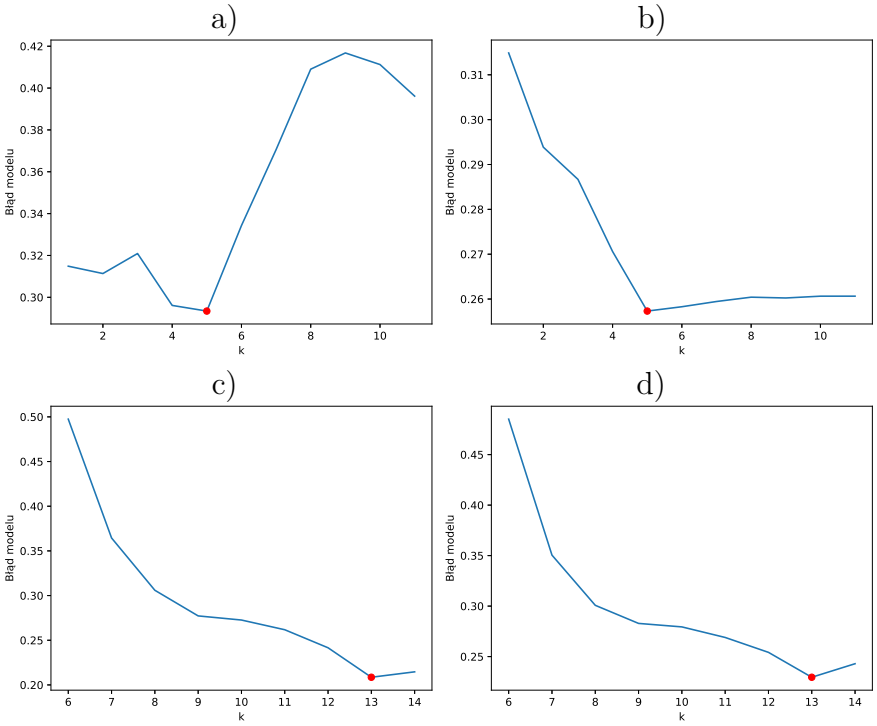
Podsumowując wyniki badań możemy stwierdzić, że podobnie jak wcześniej, modele oparte na mini-modelach liniowych i nieliniowych dały dokładność porównywalną bądź lepszą od modeli opartych na technice kNN. Przykładowe odpowiedzi modeli były zgodne z oczekiwaniami, a ich niepewność (szerokość nośnika) stanowiła uśrednioną niepewność

Tabela 4.11. Dane wykorzystane w badaniach nr 4

Wiek [lata]	Standard	Powierzchnia [m^2]	Odległość [km]	Cena [tys. zł/m^2]
10	6	90	4	4.6
'około 20' TFN(15,20,25)	7	80	[5,6]	'około 4.6' TFN(4.4,4.6,4.8)
[50,60]	4	70	8	3.9
[70,100]	2	50	3	[3.4,3.6]
5	'wysoki' TFN(8,10,10)	40	3	4.9
12	4	30	'około 7' TFN(6,7,8)	[4.5,4.7]
20	7	65	6	'około 4.8' TFN(4.7,4.8,4.9)
[0,1]	10	75	0.0	5.5
'stary' TFN(80,100,100)	0	30	'duża' TFN(12,15,15)	[3.0,3.3]
[25,35]	2	55	1	4.6
[50,70]	[8,10]	100	4	4.2
25	'średni' TFN(4,5,6)	110	'około 7' TFN(6,7,8)	3.9
10	'wysoki' TFN(8,10,10)	85	11	[4.4,4.6]
5	3	45	5	4.7
[45,55]	6	70	'około 8' TFN(7,8,9)	'około 4.1' TFN(4,4.1,4.2)

Tabela 4.12. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 4

model	e_{MAE} $k=5$	e_{MAE} $k=7$	e_{MAE} $k=9$	e_{MAE} $k=11$	e_{MAE} $k=13$
kNN – średnia arytmetyczna	0.293	0.370	0.417	0.396	0.443
kNN – średnia ważona	0.257	0.259	0.260	0.261	0.261
mini-model liniowy	1.496	0.364	0.277	0.261	0.208
mini-model nieliniowy	1.496	0.350	0.282	0.268	0.229



Rysunek 4.34. Bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla: prostej metody kNN (a) ważonej metody kNN (b), dla modelu bazującego na mini-modelach liniowych (c) i dla modelu bazującego na mini-modelach nieliniowych (d)

Tabela 4.13. Wartości odpowiedzi modeli dla przykładowego wejścia $\mathbf{x}^* = [70, 2, 60, 5]^T$

model	wyjście y dla $k = 7$	wyjście y dla $k = 13$
kNN – śr. arytmetyczna	TrFN(4.14,4.15,4.21,4.22)	TrFN(4.21,4.24,4.30,4.33)
kNN – śr. ważona	TrFN(3.60,3.61,3.73,3.74)	TrFN(3.610,3.613,3.741,3.744)
mini-model liniowy	TFN(3.54,3.59,3.64)	TFN(3.67,3.74,3.81)
mini-model nieliniowy	TFN(3.53,3.58,3.63)	TFN(3.64,3.71,3.78)

Tabela 4.14. Wartości odpowiedzi modeli dla przykładowego wejścia $\mathbf{x}^* = [\mathrm{RFN}(20, 30), \mathrm{TFN}(2.5, 5, 7.5), 45, \mathrm{RFN}(0, 2)]^T$

model	wyjście y dla $k = 7$	wyjście y dla $k = 13$
kNN – śr. arytmetyczna	TrFN(4.50,4.56,4.59,4.64)	TrFN(4.41,4.44,4.47,4.50)
kNN – śr. ważona	TrFN(4.55,4.61,4.62,4.68)	TrFN(4.54,4.60,4.61,4.66)
mini-model liniowy	TFN(5.09,5.17,5.25)	TFN(4.71,4.77,4.83)
mini-model nieliniowy	TFN(5.08,5.16,5.24)	TFN(4.67,4.72,4.77)

uwzględnianych najbliższych sąsiadów. Zgodnie z oczekiwaniami, metody bez problemów radziły sobie zarówno z danymi jednorodnymi w postaci liczb rozmytych, jak i danymi mieszanymi.

W przypadku metod pracujących na liczbach rozmytych czas wykonywanych obliczeń był już znacząco dłuższy (ok. 4-5 razy) od obliczeń wykonywanych na liczbach rzeczywistych. Wynikało to przede wszystkim z większej złożoności arytmetyki liczb rozmytych. Podobnie jak dla interwałów, szybciej wykonywane były obliczenia przez modele bazujące na mini-modelach. Wynikało to z faktu, że regresja lokalna wyznaczana była dla środków liczb rozmytych, reprezentowanych przez liczby rzeczywiste.

5. Liczby Z

5.1. Pojęcia podstawowe

Liczby Z należą do trzeciego poziomu niepewności. Liczba Z może być zdefiniowana jako uporządkowana para [1, 145]:

$$Z = (A, B), \tag{5.1}$$

gdzie:

- A – to liczba rozmyta odgrywająca rolę rozmytego ograniczenia (ang. *restriction*) na wartości jakie liczba X może przyjmować (X jest A),
- B – to liczba rozmyta odgrywająca rolę rozmytego ograniczenia na prawdopodobieństwo wystąpienia A ($P(A)$ jest B).

Za pomocą liczb Z, zdania wyrażane w języku naturalnym mogą być opisane w wygodny, ustrukturalizowany sposób. Dla przykładu zdanie:

'Istnieje wysokie prawdopodobieństwo, że stopa bezrobocia w przyszłym roku będzie niska',

może być zapisane w postaci:

$$X = \text{'stopa bezrobocia w przyszłym roku' jest } Z = (\text{'niska','wysokie'}).$$

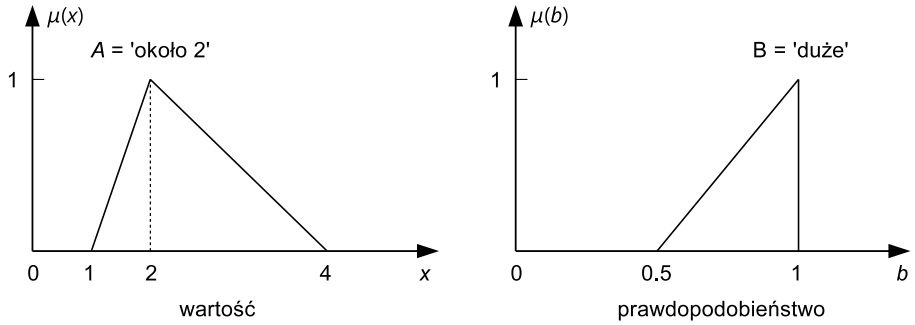
A = 'niska' i B = 'wysokie' to możliwościowe (ang. *possibilistic*) ograniczenia nałożone na zmienną X i jej prawdopodobieństwo.

Liczba Z^+ to uporządkowana para składająca się z liczby rozmytej A i liczby losowej p_X:

$$Z^+ = (A, p_X), \tag{5.2}$$

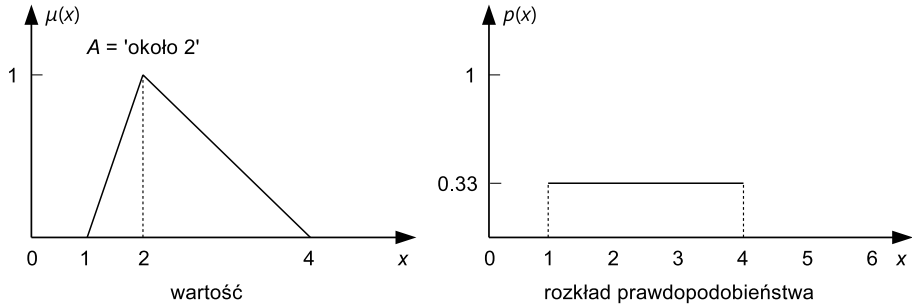
gdzie:

- A – odgrywa podobną rolę jak w liczbach Z,
- p_X – to funkcja gęstości prawdopodobieństwa losowej liczby X.

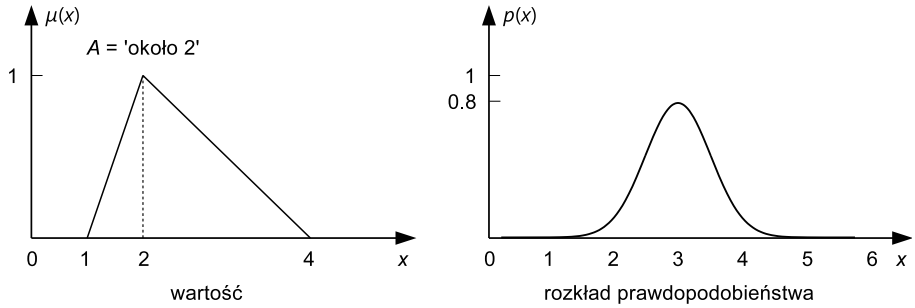


Rysunek 5.1. Przykładowa liczba $Z = (A, B) =$ ('około 2', 'duże')

Liczby Z^+ pełnią pomocniczą (choć bardzo ważną) rolę w obliczeniach na liczbach Z.



Rysunek 5.2. Przykładowa liczba Z^+ z jednostajnym rozkładem prawdopodobieństwa



Rysunek 5.3. Przykładowa liczba Z^+ z normalnym rozkładem prawdopodobieństwa $N(3, 0.5)$

Zgodnie z definicją, liczba Z^+ niesie więcej informacji niż liczba Z. Przede wszystkim, dla liczby Z^+ można obliczyć dokładną wartość prawdopodobieństwa $P(A)$ jako:

$$P(A) = \int_{\text{supp}(A)} \mu_A(u) \cdot p_X(u) \mathrm{d}u, \tag{5.3}$$

gdzie:

- $\mu_A(u)$ – to funkcja przynależności liczby rozmytej A,
- $\text{supp}(A)$ – to nośnik A.

5.2. Arytmetyka liczb Z i Z^+

Jak już wcześniej wspomniano, liczby Z^+ pełnią pomocniczą rolę przy obliczeniach na liczbach Z, stąd od nich zaczniemy omawianie sposobu wykonywania obliczeń.

Załóżmy, że $*$ to operator binarny, a jego operandami są liczby Z^+: $Z_X^+ = (A_X, p_X)$ i $Z_Y^+ = (A_Y, p_Y)$. Zgodnie z definicją:

$$Z_X^+ * Z_Y^+ = (A_X * A_Y, p_X * p_Y), \tag{5.4}$$

przy czym oczywiste jest, że działanie wykonywane jest inaczej dla liczb rozmytych: $A_X * A_Y$ i liczb losowych: $p_X * p_Y$.

W przypadku liczb rozmytych, sposób wykonywania obliczeń omówiony został w rozdziale 4. Najczęściej wykorzystywana jest tu metoda działań na α-przekrojach, omówiona na str. 55.

Bardziej szczegółowe omówienie podstaw arytmetyki liczb losowych można znaleźć w pracach [53, 125]. Niech p_X i p_Y będą funkcjami gęstości prawdopodobieństwa dwóch niezależnych, ciągłych liczb losowych. Wynikowe rozkłady, powstałe po wykonaniu działań arytmetycznych na tych liczbach, można wyznaczyć jako:

$$\begin{aligned}
p_{X+Y}(u) &= \int_{-\infty}^{\infty} p_X(v) \cdot p_Y(u-v)\,\mathrm{d}v, \\
p_{X-Y}(u) &= \int_{-\infty}^{\infty} p_X(v) \cdot p_Y(v-u)\,\mathrm{d}v, \\
p_{X\cdot Y}(u) &= \int_{-\infty}^{\infty} p_X(v) \cdot p_Y(u/v) \cdot \frac{1}{|v|}\,\mathrm{d}v, \\
p_{X/Y}(u) &= \int_{-\infty}^{\infty} p_X(u\cdot v) \cdot p_Y(v) \cdot |v|\,\mathrm{d}v.
\end{aligned} \tag{5.5}$$

Jeśli liczby losowe, na których chcemy wykonać działanie, mają postać dyskretną, obliczenia rozkładu wynikowego należy wykonać inaczej [1, 125]. Załóżmy, że mamy dane dwie dyskretne zmienne losowe X i Y przyjmujące wartości: $X = \{x_1, \ldots x_{n_x}\}$, $Y = \{y_1, \ldots y_{n_y}\}$. Dla zmiennych tych, dane są dyskretne rozkłady prawdopodobieństwa: p_x oraz p_y. Wynikowy rozkład prawdopodobieństwa dla wyrażenia $X * Y$, gdzie $*$ oznacza określone działanie arytmetyczne $\{+, -, \cdot, /\}$, możemy obliczyć jako splot:

$$p(w) = \sum_{w=x_i * y_j} p_x(x_i) \cdot p_y(y_j), \tag{5.6}$$

dla każdego: $w \in \{x_i * y_j \mid x_i \in X,\ y_j \in Y,\ i = 1, \ldots, n_x,\ j = 1, \ldots, n_y\}$.

Możemy teraz przejść do działań na liczbach Z. Załóżmy, że dane są dwie ta-

kie liczby: $Z_X = (A_X, B_X)$ i $Z_Y = (A_Y, B_Y)$, na których chcemy wykonać działanie $*$ ($* \in \{+, -, \cdot, /\}$). A_X i A_Y to liczby rozmyte, stanowiące możliwościowe ograniczenie na wartość liczb, B_X i B_Y to liczby rozmyte opisujące możliwościowe ograniczenie nałożone na prawdopodobieństwo ich wystąpienia. Wynikową wartość A_{XY} obliczyć można identycznie jak w przypadku liczb Z^+:

$$A_{XY} = A_X * A_Y, \tag{5.7}$$

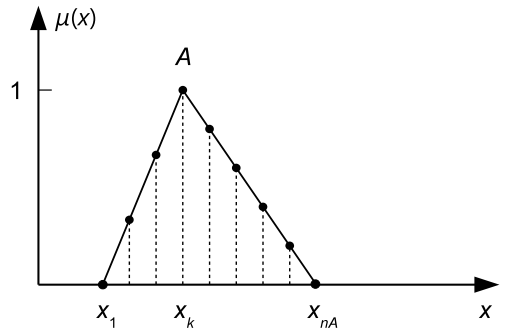
wykonując działania na liczbach rozmytych (rozdział 4, s. 55). Trudniej natomiast obliczyć wynikową wartość B_{XY}. Zacznijmy od krótkiej dygresji.

Gdyby liczba Z miała postać $Z = (A, P)$, gdzie P to prawdopodobieństwo jej wystąpienia znane dokładnie, wówczas ze wzoru:

$$P = \int_{\mathbb{R}} \mu_A(x) \cdot p(x) \mathrm{d}x, \tag{5.8}$$

moglibyśmy wyznaczyć odpowiadający jej rozkład prawdopodobieństwa $p(x)$. Zadanie to nie ma oczywiście jednoznacznego rozwiązania. Istnieje nieskończenie wiele rozkładów spełniających równanie (5.8).

Będziemy poszukiwali rozkładu w wersji dyskretnej. W tym celu podzielimy nośnik liczby A na $n_A - 1$ równych przedziałów, rys. 5.4.



Rysunek 5.4. Ilustracja sposobu dyskretyzacji liczby rozmytej A

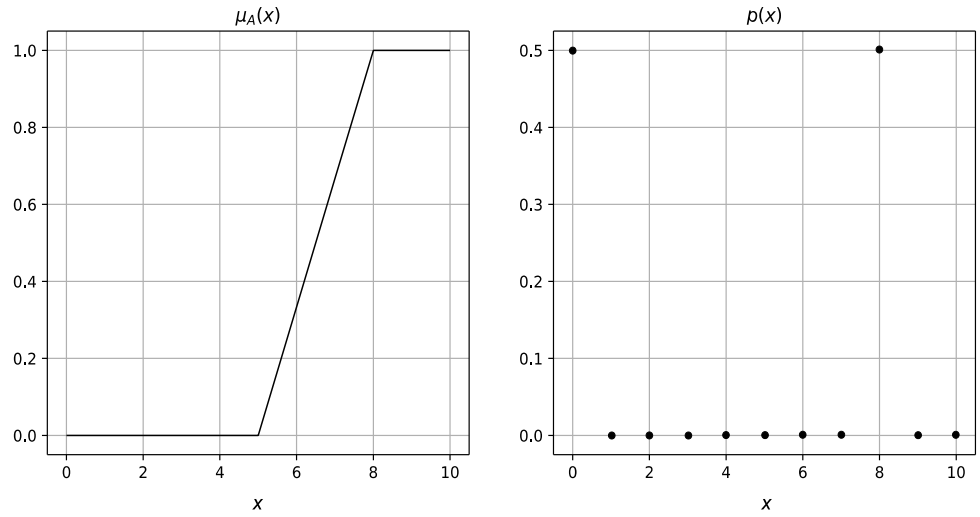
Równanie (5.8) w wersji dyskretnej będzie miało postać:

$$P = \sum_{k=1}^{n_A} \mu_A(x_k) \cdot p(x_k), \tag{5.9}$$

przy dodatkowych założeniach:

$$\sum_{k=1}^{n_A} p(x_k) = 1 \quad \text{i} \quad \forall k = 1, \ldots, n_A,\ p(x_k) \geqslant 0\,. \tag{5.10}$$

W literaturze [1] proponuje się, aby rozkład wyznaczyć rozwiązując zadanie programowania liniowego. Można to zrobić zakładając brak funkcji celu oraz przyjmując ograniczenia równościowe i nierównościowe (5.9) oraz (5.10). Wyznaczony w ten sposób wynik zależał będzie od przyjętej metody rozwiązywania. Przykładowo, metoda simplex [9, 17, 50] wyznaczać będzie zwykle rozkłady w takiej postaci, w której tylko dla 2 wartości x_k wartość rozkładu $p(x_k)$ jest różna od 0. Przykład takiego rozwiązania pokazuje rys. 5.5.

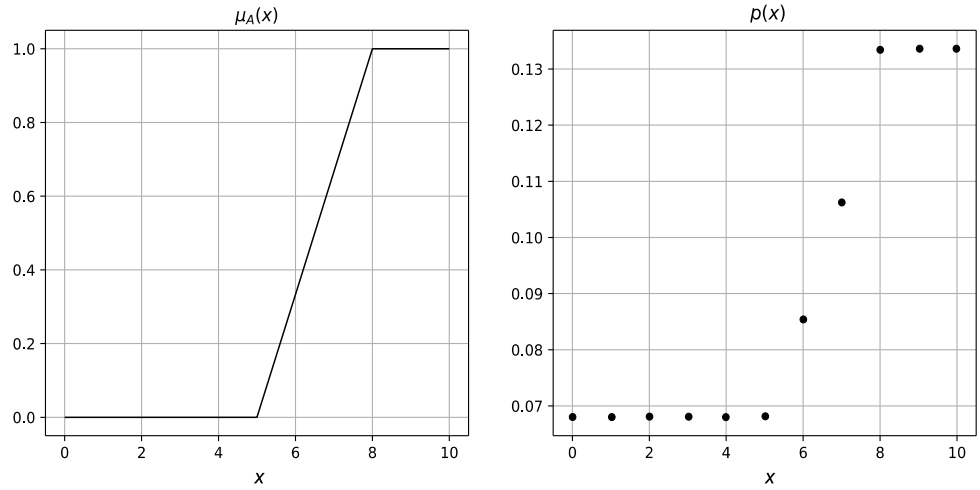


Rysunek 5.5. Przykładowy dyskretny rozkład $p(x)$ wyznaczony dla prawdopodobieństwa $P = 0.5$ metodą simplex

Dużo korzystniejsze rozwiązania daje metoda punktu wewnętrznego (ang. *interior point method*) [4, 5]. Jej zaletą jest to, że w wyznaczonym rozkładzie dla każdej wartości x_k wartość rozkładu $p(x_k)$ jest różna od 0. Przykład takiego rozwiązania pokazano na rys. 5.6.

Na rys. 5.5 i 5.6 na wykresach po lewej przedstawiono funkcję przynależności przykładowej trapezowej liczby rozmytej A=TrFN(5,8,10,10). W obu przykładach, dyskretny rozkład prawdopodobieństwa wyznaczany był dla całego obszaru rozważań zmiennej $x \in [0, 10]$ próbkowanej z krokiem 1. Na tym dygresję zakończymy.

W przypadku liczby Z, prawdopodobieństwo jej wystąpienia nie jest dokładną liczbą P, a liczbą rozmytą, którą oznaczyliśmy jako B. Oznacza to, że prawdopodobieństwo może przyjmować różne wartości (nieskończenie wiele) określone przez interwał definiujący nośnik liczby B. Tym samym, gdybyśmy chcieli wyznaczyć rozkład prawdopodobieństwa dla liczby $Z = (A, B)$ to byłby to nie jeden rozkład, a zbiór rozkładów (nieskończenie wiele) wyznaczonych dla różnych wartości prawdopodobieństwa należących do nośnika liczby B. Dokładniej, byłby to zbiór rozmyty rozkładów prawdopodobieństwa, bo z każdym rozkładem $p_b(x)$ wyznaczonym dla określonej wartości b, można powiązać odpowiadającą

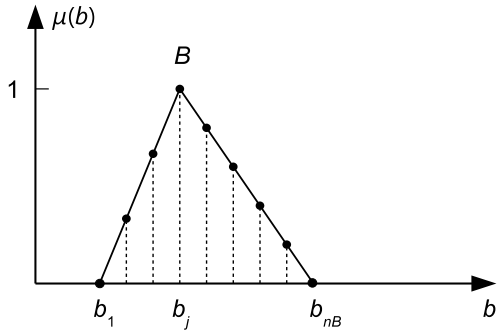


Rysunek 5.6. Przykładowy dyskretny rozkład $p(x)$ wyznaczony dla prawdopodobieństwa $P = 0.5$ metodą punktu wewnętrznego

mu wartość funkcji przynależności $\mu_B(b)$:

$$\{p_b(x), \mu_B(b)\}.$$

Aby wyznaczyć rozmyty zbiór rozkładów prawdopodobieństwa, liczbę rozmytą B również poddamy dyskretyzacji, tzn. podzielimy jej nośnik na $n_B - 1$ równych przedziałów (rys. 5.7).



Rysunek 5.7. Ilustracja sposobu dyskretyzacji liczby rozmytej B

Aby wyznaczyć wynikową wartość B_{XY} wykonać należy działania na dyskretnych rozkładach prawdopodobieństwa, wyznaczonych (w sposób opisany powyżej) dla liczb B_X i B_Y. Dla liczby B_X mamy rozkłady $p_X(x)$ wyznaczone dla wartości $b_{X1}, \ldots, b_{Xn_B}$, a dla liczby B_Y mamy rozkłady $p_Y(x)$ wyznaczone dla wartości $b_{Y1}, \ldots, b_{Yn_B}$. Dla obu liczb rozmytych mamy zatem wyznaczonych po n_B rozkładów.

W dalszej kolejności wyznaczyć należy wynikowe rozkłady $p_{XY}(x)$ dla wszystkich możliwych par prawdopodobieństw (b_{Xi}, b_{Yj}), gdzie $i, j = 1, \ldots, n_B$. Obliczenia wykonuje

się za pomocą zależności (5.6). Otrzymuje się w ten sposób n_B^2 rozkładów wynikowych, dla których należy jeszcze wyznaczyć wynikową wartość funkcji przynależności:

$$\mu(p_{XY}(x)) = \max_{p_X, p_Y}[\mu_{BX}(b) \wedge \mu_{BY}(b)], \tag{5.11}$$

przy czym $p_{XY}(x) = p_X(x) * p_Y(x)$, a operator $\wedge$ oznacza t-normę, zwykle liczoną jako minimum. W efekcie otrzymujemy zbiór rozmyty n_B^2 rozkładów prawdopodobieństw wynikowych z przypisanymi do nich wynikowymi wartościami funkcji przynależności:

$$\{p_{XY}(x), \mu(p_{XY}(x))\}.$$

Dla każdego wynikowego rozkładu prawdopodobieństwa p_{XY} możemy wyznaczyć teraz wartość prawdopodobieństwa:

$$b_{XY} = \int_{\mathbb{R}} \mu_{A_{XY}}(x) \cdot p_{XY}(x) \mathrm{d}x, \tag{5.12}$$

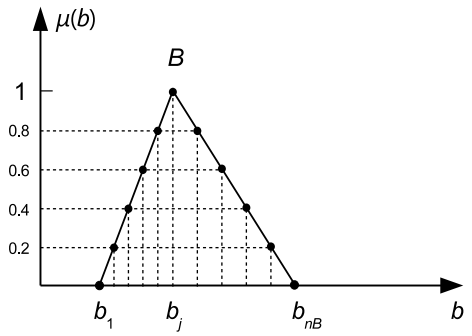
otrzymując w efekcie dyskretny, rozmyty zbiór złożony z par:

$$\{b_{XY}, \mu(b_{XY})\}.$$

Ostateczną, dyskretną postać funkcji przynależności wynikowej liczby rozmytej B_{XY} wyznaczamy jako:

$$\mu_{B_{XY}} = \sup_{b_{XY}}[\mu(b_{XY})]. \tag{5.13}$$

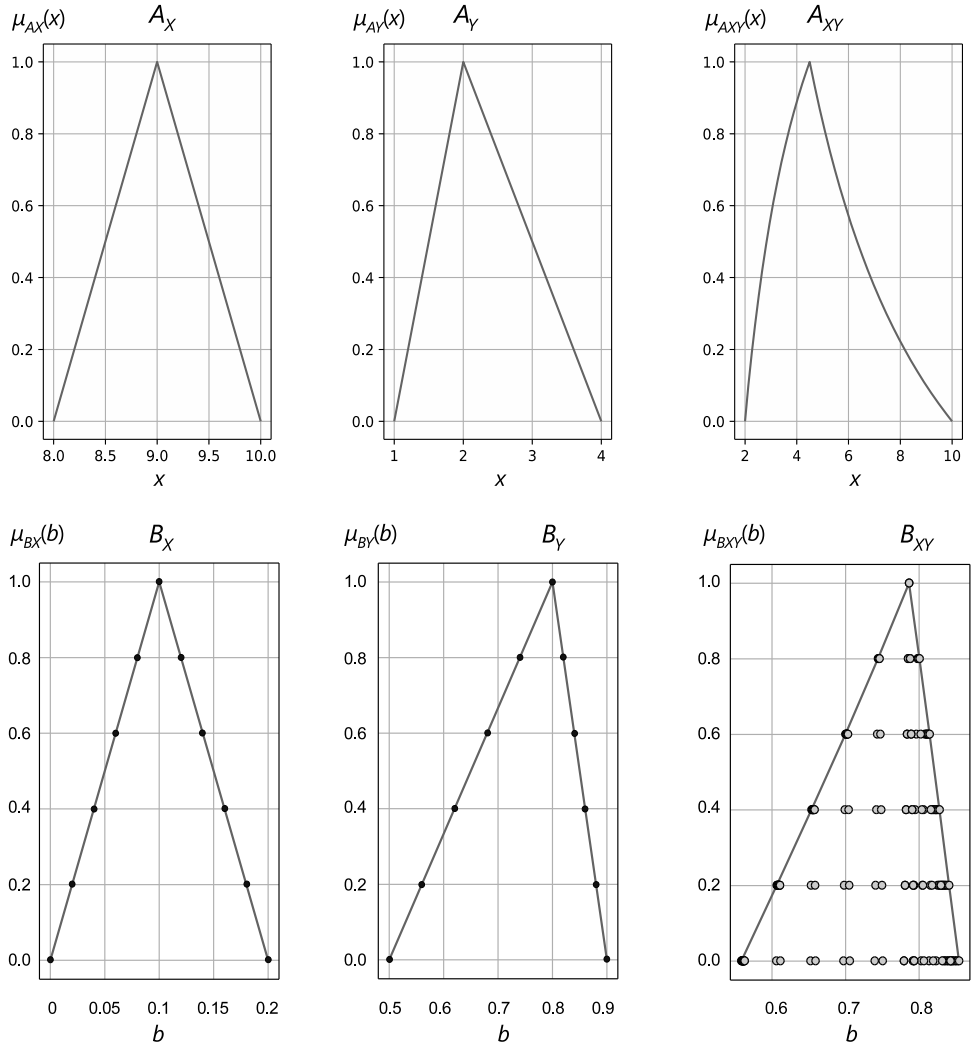
W praktyce, przy numerycznym wyznaczaniu wynikowej liczby B_{XY}, liczby B_X i B_Y korzystniej jest dyskretyzować wzdłuż osi pionowej, opisującej wartości przynależności (rys. 5.8). Dyskretyzacji podlega w ten sposób lewy i prawy brzeg liczb rozmytych na równomiernie rozłożonych α-przekrojach.



Rysunek 5.8. Ilustracja sposobu dyskretyzacji liczby rozmytej B wzdłuż osi rzędnych

Niewątpliwą zaletą takiego podejścia jest to, że przy zastosowaniu wzoru (5.11)

z operatorem minimum, wynikowa funkcja przynależności $\mu_{B_{XY}}$ przyjmuje wartości z tego samego zbioru co dyskretne funkcje μ_{B_X} i μ_{B_Y}. Zilustrowano to na rys. 5.9.



Rysunek 5.9. Ilustracja przebiegu dzielenia dwóch liczb Z

Na rys. 5.9 pokazany jest wynik $Z_{XY} = Z_X/Z_Y$ dzielenia dwóch liczb: $Z_X = (A_X, B_X)$ = (TFN(8,9,10), TFN(0,0.1,0.2)) oraz $Z_Y = (A_Y, B_Y)$ = (TFN(1,2,4), TFN (0.5,0.8,0.9)). Dodatkowo na wykresach funkcji przynależności μ_{B_X} i μ_{B_Y} zaznaczone są punkty, dla których przeprowadzane były obliczenia po dyskretyzacji funkcji (równomiernie co 0.2). Na wykresie wynikowej funkcji przynależności $\mu_{B_{XY}}$ zaznaczone są punkty, których współrzędne obliczone są za pomocą wzorów (5.11) i (5.12). Punkty te definiują kształt wynikowej funkcji przynależności, który wyznaczany jest za pomocą wzoru (5.13).

W praktyce, aby wyznaczyć wynikowy kształt dyskretnej funkcji przynależności $\mu_{B_{XY}}$, wystarczy dla każdego dyskretnego α-przekroju odnaleźć najmniejszą i najwięk-

szą wartość b, spośród wszystkich wyznaczonych punktów. W ten sposób zdefiniujemy odpowiednio lewy i prawy brzeg dyskretnej wynikowej liczby rozmytej B_{XY}[1].

Zaprezentowany powyżej schemat działań na liczbach Z może być w podobny sposób wykorzystany także do innych operacji, jak np. potęgowanie czy wyznaczanie wartości bezwzględnej.

5.3. Odległość pomiędzy liczbami Z

Pojęcie odległości między liczbami Z będzie niezbędne w modelowaniu lokalnej regresji. Wykorzystane zostanie zarówno przy poszukiwaniu najbliższych sąsiadów dla punktu wejściowego, jak i przy definiowaniu kryterium jakości dla metody najmniejszych kwadratów pracującej na liczbach Z.

Podobnie jak w przypadku liczb rozmytych, w literaturze opisanych jest kilka metod wyznaczania odległości między liczbami Z [1, 2, 112]. Mają one swoje wady i zalety, przy czym każda z nich sprowadza się do wyznaczania odległości jako:

$$D(Z_1, Z_2) = d(A_1, A_2) + d(B_1, B_2), \tag{5.14}$$

gdzie: $d(A_1, A_2)$ to odległość między liczbami rozmytymi opisującymi wartości liczb Z, a $d(B_1, B_2)$ to odległość między liczbami rozmytymi opisującymi prawdopodobieństwa ich wystąpienia [19]. Odległość d może być wyznaczana ze wzoru (4.39) znajdującego się na stronie 64.

Największym mankamentem takiego podejścia jest nieuwzględnienie, że wartości liczby Z i prawdopodobieństwa ich wystąpienia mają zupełnie inne znaczenie, a także zwykle należą do różnych dziedzin. We wzorze (5.14) niezbędne jest zatem uwzględnienie odpowiednich współczynników wagowych, umożliwiających nadanie znaczenia jego składnikom, zgodnego z potrzebą tworzonego modelu.

Liczby rozmyte B opisują prawdopodobieństwo, stąd należą zawsze do obszaru rozważań [0,1] i są znormalizowane z definicji. Wynika z tego, że we wzorze (5.14) wystarczy dodać tylko jeden współczynnik wagowy, skalujący odpowiednio składnik opisujący odległość pomiędzy wartościami liczb Z:

$$D_1(Z_1, Z_2) = \lambda_A \cdot d(A_1, A_2) + d(B_1, B_2). \tag{5.15}$$

[1] Takie podejście zastosowane zostało w bibliotece FUZCal, opracowanej przez Autora i wykorzystanej w badaniach, których wyniki zaprezentowano w niniejszej monografii. Biblioteka, napisana w języku Python, udostępnia klasy umożliwiające podstawowe operacje arytmetyczne (dodawanie, odejmowanie, mnożenie i dzielenie) na interwałach, liczbach rozmytych oraz liczbach Z. Dodatkowo, możliwe jest też wykonywanie takich działań jak potęgowanie, wyznaczanie wartości bezwzględnej, wyznaczanie odległości między liczbami i ich wizualizacja. Biblioteka dostępna jest na stronie `http://wikizmsi.zut.edu.pl/wiki/MP`.

Przez analogię do metryki stosowanej wcześniej i opisanej wzorem (4.39), odległość między liczbami Z można wyznaczyć także jako:

$$D_2(Z_1, Z_2) = \sqrt{\lambda_A \cdot d^2(A_1, A_2) + d^2(B_1, B_2)}. \tag{5.16}$$

Jeśli przez $\mathbb{Z}$ oznaczymy przestrzeń liczb Z:

$$\mathbb{Z} = \{Z(A, B) \mid A \in F(\mathbb{R}),\ B \in F(\mathbb{R})\}, \tag{5.17}$$

wówczas przestrzenie $(\mathbb{Z}, D_1)$ i $(\mathbb{Z}, D_2)$ są przestrzeniami metrycznymi [1]. Wynika to z tego, że przestrzeń $(F(\mathbb{R}), d)$ również jest przestrzenią metryczną.

5.4. Metoda najmniejszych kwadratów dla liczb Z

Teoria liczb Z jest stosunkowo nowa, stąd publikacji dotyczących regresji liniowej pracującej na takich liczbach jest niewiele. Przykładowe podejście zaprezentowane jest w pracach [1, 52]. Poszukiwana jest tam regresja liniowa z parametrami w postaci liczb Z. Parametry wyznaczane są za pomocą technik programowania liniowego. Inne przykłady można znaleźć w publikacjach [3, 28, 83, 117].

Poszukiwanie regresji z parametrami będącymi liczbami Z jest zwykle bardzo złożone obliczeniowo. W rozwiązaniu zaprezentowanym poniżej, zastosowano uproszczony model z parametrami rzeczywistymi. Tego typu podejście jest proste obliczeniowo i, podobnie jak dla liczb rozmytych, daje dobre, wiarygodne rozwiązania.

5.4.1. Model z jednym wejściem

Metoda najmniejszych kwadratów, w wersji pracującej na liczbach Z, zostanie zaprezentowana w pierwszej kolejności dla modelu z jednym wejściem. Załóżmy, że posiadamy dane w postaci par (Z_{Xi}, Z_{Yi}), $i = 1 \ldots K$, przy czym zarówno wejście Z_{Xi} jak i wyjście Z_{Yi} jest liczbą Z. Zaproponowana poniżej metoda ma w założeniu pracować na danych różnego typu. Jeśli zatem w danych znajdą się liczby rzeczywiste, interwały bądź liczby rozmyte, należy je zawsze przekonwertować do liczb Z. Będzie to wymagało dodania do każdej z nich dodatkowego elementu B, opisującego prawdopodobieństwo ich wystąpienia. Wartości tych elementów będą zależały od konkretnego przypadku, dla którego tworzymy model i od stopnia pewności posiadanych danych. Przykładowo, jeśli nasze dane są pewne, przy konwersji do liczby Z można przyjąć $B = \mathrm{SFN}(1)$, czyli jednoelementową liczbę rozmytą (typu singleton) równą 1 [1].

Ponieważ liczba $Z = (A, B)$ jest parą liczb rozmytych, będziemy poszukiwali dwóch modeli aproksymujących. Przy wykonywaniu działań na liczbach Z, wartość liczby A

zależy jedynie od wartości $A_1, A_2, \ldots$ występujących w operandach, stąd pierwszy model będzie miał postać:

$$\hat{A}_Y = w_1 \cdot A_{Xs} + w_0 + R_A. \tag{5.18}$$

Prawdopodobieństwo B wyznaczane jest w bardziej złożony sposób i zależy zarówno od wartości $A_1, A_2, \ldots$ jak i $B_1, B_2, \ldots$ występujących w operandach, stąd drugi model będzie miał postać:

$$\hat{B}_Y = u_{A1} \cdot A_{Xs} + u_{B1} \cdot B_{Xs} + u_0 + R_B. \tag{5.19}$$

W powyższych równaniach A_{Xs}, B_{Xs} oznaczają środki liczb rozmytych A_X, B_X definiujących wejściowe liczby Z modelu: $Z_X = (A_X, B_X)$. Środki liczb rozmytych możemy wyznaczyć z zależności (4.41). R_A to trójkątna liczba rozmyta w postaci $\mathrm{TFN}(-r_A, 0, r_A)$, a R_B to trójkątna liczba rozmyta w postaci $\mathrm{TFN}(-r_B, 0, r_B)$. Funkcje przynależności lewej i prawej strony trójkątnych liczb rozmytych oraz ich odwrotności opisane są zależnościami (4.42), (4.43) na stronie 67. Zmienne $w_1, w_0, r_A, u_{A1}, u_{B1}, u_0, r_B$ będą rzeczywistymi współczynnikami modelu.

Będziemy chcieli, aby modele dopasowane były do danych w taki sposób, aby minimalizować kryterium jakości oparte na metryce D_2 opisanej zależnością (5.16):

$$Q(w_1, w_0, r_A, u_{A1}, u_{B1}, u_0, r_B) = \sum_{i=1}^{K} D_2^2(\hat{Z}_{Yi}, Z_{Yi}) \longrightarrow \min, \tag{5.20}$$

gdzie: $\hat{Z}_{Yi} = (w_1 \cdot A_{Xsi} + w_0 + R_A, u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + R_B)$.

Przyjmując, że odległość między liczbami rozmytymi obliczana będzie z zależności (4.39), kryterium jakości (5.20) możemy zapisać inaczej w postaci:

$$
\begin{aligned}
Q(w_1, w_0, r_A, u_{A1}, u_{B1}, u_0, r_B) &= \sum_{i=1}^{K} D_2^2(\hat{Z}_{Yi}, Z_{Yi}) \qquad (5.21)\\
&= \sum_{i=1}^{K} \lambda_A \cdot d_2^2(\hat{A}_{Yi}, A_{Yi}) + d_2^2(\hat{B}_{Yi}, B_{Yi})\\
&= \lambda_A \cdot \left(\frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(\alpha - 1) - f_{AYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right.\\
&\quad \left. + \frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(1 - \alpha) - g_{AYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right)\\
&\quad + \left(\frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(\alpha - 1) - f_{BYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right.\\
&\quad \left. + \frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(1 - \alpha) - g_{BYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right).
\end{aligned}
$$

Aby wyznaczyć minimum kryterium (w skrócie oznaczanego dalej jako Q), liczymy pochodne cząstkowe po parametrach modelu i przyrównujemy je do zera. Pierwsze dwa składniki w kryterium (5.21) zależą jedynie od parametrów: w_1, w_0, r_A, natomiast pozostałe dwa od parametrów: u_{A1}, u_{B1}, u_0, r_B. Wynika z tego, że zadanie wyznaczania parametrów możemy rozdzielić na dwa niezależne etapy.

W pierwszym etapie, przeprowadzane obliczenia są bardzo podobne do tych, jakie wcześniej wykonywane były dla liczb rozmytych.

$$
\begin{aligned}
\frac{\partial Q}{\partial w_1} =& \lambda_A \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(\alpha - 1) - f_{AYi}^{-1}(\alpha) \right] \cdot A_{Xsi} \mathrm{d}\alpha\\
&+ \lambda_A \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(1 - \alpha) - g_{AYi}^{-1}(\alpha) \right] \cdot A_{Xsi} \mathrm{d}\alpha\\
=& \lambda_A \sum_{i=1}^{K} \int_0^1 w_1 A_{Xsi}^2 + w_0 A_{Xsi} + r_A(\alpha - 1) A_{Xsi} - f_{AYi}^{-1}(\alpha) A_{Xsi}\\
&+ w_1 A_{Xsi}^2 + w_0 A_{Xsi} + r_A(1 - \alpha) A_{Xsi} - g_{AYi}^{-1}(\alpha) A_{Xsi} \mathrm{d}\alpha\\
=& \lambda_A \sum_{i=1}^{K} \int_0^1 2 w_1 A_{Xsi}^2 + 2 w_0 A_{Xsi} - \left[f_{AYi}^{-1}(\alpha) + g_{AYi}^{-1}(\alpha) \right] A_{Xsi} \mathrm{d}\alpha = 0
\end{aligned}
$$

$$\begin{aligned}\frac{\partial Q}{\partial w_0} =&\lambda_A \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(\alpha - 1) - f_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha \\ &+ \lambda_A \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(1 - \alpha) - g_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha \\ =&\lambda_A \sum_{i=1}^{K} \int_0^1 2w_1 A_{Xsi} + 2w_0 - \left[f_{AYi}^{-1}(\alpha) + g_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha = 0\end{aligned}$$

$$\begin{aligned}\frac{\partial Q}{\partial r_A} =&\lambda_A \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(\alpha - 1) - f_{AYi}^{-1}(\alpha)\right] \cdot (\alpha - 1)\mathrm{d}\alpha \\ &+ \lambda_A \sum_{i=1}^{K} \int_0^1 \left[w_1 \cdot A_{Xsi} + w_0 + r_A(1 - \alpha) - g_{AYi}^{-1}(\alpha)\right] \cdot (1 - \alpha)\mathrm{d}\alpha = 0\end{aligned}$$

Przekształcając dalej ostatnią pochodną, mamy:

$$\begin{aligned}&\sum_{i=1}^{K} \int_0^1 \left[r_A(\alpha - 1)^2 - f_{AYi}^{-1}(\alpha)(\alpha - 1) + r_A(1 - \alpha)^2 - g_{AYi}^{-1}(\alpha)(1 - \alpha)\right] \mathrm{d}\alpha \\ &= \sum_{i=1}^{K} \int_0^1 2r_A(\alpha - 1)^2 \mathrm{d}\alpha - \sum_{i=1}^{K} \int_0^1 \left[f_{AYi}^{-1}(\alpha)(\alpha - 1) + g_{AYi}^{-1}(\alpha)(1 - \alpha)\right] \mathrm{d}\alpha \\ &= 2Kr_A \int_0^1 (\alpha - 1)^2 \mathrm{d}\alpha - \sum_{i=1}^{K} \int_0^1 \left[f_{AYi}^{-1}(\alpha)(\alpha - 1) + g_{AYi}^{-1}(\alpha)(1 - \alpha)\right] \mathrm{d}\alpha \\ &= \frac{2}{3} Kr_A - \sum_{i=1}^{K} \int_0^1 \left[f_{AYi}^{-1}(\alpha)(\alpha - 1) + g_{AYi}^{-1}(\alpha)(1 - \alpha)\right] \mathrm{d}\alpha = 0.\end{aligned}$$

Po przekształceniach dostajemy ostatecznie układ 3 równań z trzema niewiadomymi parametrami:

$$\begin{aligned}w_1 \sum_{i=1}^{K} A_{Xsi}^2 + w_0 \sum_{i=1}^{K} A_{Xsi} &= \sum_{i=1}^{K} A_{Xsi} \cdot \frac{1}{2} \int_0^1 f_{AYi}^{-1}(\alpha) + g_{AYi}^{-1}(\alpha)\mathrm{d}\alpha, \\ w_1 \sum_{i=1}^{K} A_{Xsi} + w_0 K &= \sum_{i=1}^{K} \frac{1}{2} \int_0^1 f_{AYi}^{-1}(\alpha) + g_{AYi}^{-1}(\alpha)\mathrm{d}\alpha, \\ \frac{2}{3} Kr_A &= \sum_{i=1}^{K} \int_0^1 (1 - \alpha) \left[g_{AYi}^{-1}(\alpha) - f_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha,\end{aligned}$$

po rozwiązaniu którego otrzymujemy współczynniki opisane zależnościami:

$$w_1 = \frac{K \cdot \sum_{i=1}^{K} A_{Xsi} \cdot \int_0^1 \frac{1}{2}\left[f_{AYi}^{-1}(\alpha) + g_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha - \sum_{i=1}^{K} A_{Xsi} \cdot \sum_{i=1}^{K} \int_0^1 \frac{1}{2}\left[f_{AYi}^{-1}(\alpha) + g_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha}{K \cdot \sum_{i=1}^{K} A_{Xsi}^2 - \left(\sum_{i=1}^{K} A_{Xsi}\right)^2},$$

$$w_0 = \frac{1}{K} \cdot \left[\sum_{i=1}^{K} \int_0^1 \frac{1}{2}\left[f_{AYi}^{-1}(\alpha) + g_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha - w_1 \cdot \sum_{i=1}^{K} A_{Xsi}\right],$$

$$r_A = \frac{3}{2K} \cdot \sum_{i=1}^{K} \int_0^1 (1-\alpha)\left[g_{AYi}^{-1}(\alpha) - f_{AYi}^{-1}(\alpha)\right] \mathrm{d}\alpha. \tag{5.22}$$

W drugim etapie, liczymy kolejne pochodne cząstkowe.

$$\begin{aligned}\frac{\partial Q}{\partial u_{A1}} =& \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(\alpha - 1) - f_{BYi}^{-1}(\alpha)\right] \cdot A_{Xsi} \mathrm{d}\alpha \\ &+ \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(1-\alpha) - g_{BYi}^{-1}(\alpha)\right] \cdot A_{Xsi} \mathrm{d}\alpha = 0\end{aligned}$$

$$\begin{aligned}\frac{\partial Q}{\partial u_{B1}} =& \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(\alpha - 1) - f_{BYi}^{-1}(\alpha)\right] \cdot B_{Xsi} \mathrm{d}\alpha \\ &+ \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(1-\alpha) - g_{BYi}^{-1}(\alpha)\right] \cdot A_{Bsi} \mathrm{d}\alpha = 0\end{aligned}$$

$$\begin{aligned}\frac{\partial Q}{\partial u_0} =& \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(\alpha - 1) - f_{BYi}^{-1}(\alpha)\right] \mathrm{d}\alpha \\ &+ \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(1-\alpha) - g_{BYi}^{-1}(\alpha)\right] \mathrm{d}\alpha = 0\end{aligned}$$

$$\begin{aligned}\frac{\partial Q}{\partial r_B} =& \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(\alpha - 1) - f_{BYi}^{-1}(\alpha)\right] \cdot (\alpha - 1) \mathrm{d}\alpha \\ &+ \sum_{i=1}^{K} \int_0^1 \left[u_{A1} \cdot A_{Xsi} + u_{B1} \cdot B_{Xsi} + u_0 + r_B(1-\alpha) - g_{BYi}^{-1}(\alpha)\right] \cdot (1-\alpha) \mathrm{d}\alpha = 0\end{aligned}$$

Ostatnie równanie, po przekształceniach i uproszczeniu, przyjmuje postać:

$$\sum_{i=1}^{K} \int_0^1 \left[r_B(\alpha-1)^2 - f_{BYi}^{-1}(\alpha)(\alpha-1) + r_B(1-\alpha)^2 - g_{BYi}^{-1}(\alpha)(1-\alpha) \right] \mathrm{d}\alpha = 0.$$

Możemy z niego wyznaczyć parametr r_B jako:

$$r_B = \frac{3}{2K} \cdot \sum_{i=1}^{K} \int_0^1 (1-\alpha) \left[g_{BYi}^{-1}(\alpha) - f_{BYi}^{-1}(\alpha) \right] \mathrm{d}\alpha. \tag{5.23}$$

Pozostałe 3 równania, po przekształceniach utworzą układ z trzema niewiadomymi.

$$\begin{aligned}
u_{A1} \cdot \sum_{i=1}^{K} A_{Xsi}^2 + u_{B1} \cdot \sum_{i=1}^{K} A_{Xsi}B_{Xsi} + u_0 \cdot \sum_{i=1}^{K} A_{Xsi} &= \frac{1}{2} \sum_{i=1}^{K} A_{Xsi} \cdot \int_0^1 \left[f_{BYi}^{-1}(\alpha) + g_{BYi}^{-1}(\alpha) \right] \mathrm{d}\alpha \\
u_{A1} \cdot \sum_{i=1}^{K} A_{Xsi}B_{Xsi} + u_{B1} \cdot \sum_{i=1}^{K} B_{Xsi}^2 + u_0 \cdot \sum_{i=1}^{K} B_{Xsi} &= \frac{1}{2} \sum_{i=1}^{K} B_{Xsi} \cdot \int_0^1 \left[f_{BYi}^{-1}(\alpha) + g_{BYi}^{-1}(\alpha) \right] \mathrm{d}\alpha \\
u_{A1} \cdot \sum_{i=1}^{K} A_{Xsi} + u_{B1} \cdot \sum_{i=1}^{K} B_{Xsi} + u_0 \cdot K &= \frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[f_{BYi}^{-1}(\alpha) + g_{BYi}^{-1}(\alpha) \right] \mathrm{d}\alpha
\end{aligned} \tag{5.24}$$

Całkę po prawej stronie równań oznaczmy jako:

$$B_{Ysi} = \int_0^1 \left[f_{BYi}^{-1}(\alpha) + g_{BYi}^{-1}(\alpha) \right] \mathrm{d}\alpha.$$

Opisuje ona środek liczby rozmytej B_{Yi}.

Dla układu równań obliczamy wyznaczniki.

$$\begin{aligned}
\Delta =& N \cdot \sum_{i=1}^{K} A_{Xsi}^2 \cdot \sum_{i=1}^{K} B_{Xsi}^2 + 2 \cdot \sum_{i=1}^{K} A_{Xsi}B_{Xsi} \cdot \sum_{i=1}^{K} A_{Xsi} \cdot \sum_{i=1}^{K} B_{Xsi} \\
&- \left(\sum_{i=1}^{K} A_{Xsi} \right)^2 \cdot \sum_{i=1}^{K} B_{Xsi}^2 - \sum_{i=1}^{K} A_{Xsi}^2 \cdot \left(\sum_{i=1}^{K} B_{Xsi} \right)^2 - N \cdot \left(\sum_{i=1}^{K} A_{Xsi}B_{Xsi} \right)^2
\end{aligned}$$

$$\begin{aligned}
\Delta_{u_{A1}} =& N \cdot \sum_{i=1}^{K} B_{Xsi}^2 \cdot \sum_{i=1}^{K} A_{Xsi}B_{Ysi} + \sum_{i=1}^{K} A_{Xsi} \cdot \sum_{i=1}^{K} B_{Xsi} \cdot \sum_{i=1}^{K} B_{Xsi}B_{Ysi} \\
&+ \sum_{i=1}^{K} A_{Xsi}B_{Xsi} \cdot \sum_{i=1}^{K} B_{Xsi} \cdot \sum_{i=1}^{K} B_{Ysi} - \sum_{i=1}^{K} A_{Xsi} \cdot \sum_{i=1}^{K} B_{Xsi}^2 \cdot \sum_{i=1}^{K} B_{Ysi} \\
&- \left(\sum_{i=1}^{K} B_{Xsi} \right)^2 \cdot \sum_{i=1}^{K} A_{Xsi}B_{Ysi} - N \cdot \sum_{i=1}^{K} A_{Xsi}B_{Xsi} \cdot \sum_{i=1}^{K} B_{Xsi}B_{Ysi}
\end{aligned}$$

$$\begin{aligned}\Delta_{u_{B1}} =& N \cdot \sum_{i=1}^{K} A_{Xsi}^2 \cdot \sum_{i=1}^{K} B_{Xsi}B_{Ysi} + \sum_{i=1}^{K} A_{Xsi} \cdot \sum_{i=1}^{K} A_{Xsi}B_{Xsi} \cdot \sum_{i=1}^{K} B_{Ysi} \\ &+ \sum_{i=1}^{K} A_{Xsi} \cdot \sum_{i=1}^{K} B_{Xsi} \cdot \sum_{i=1}^{K} A_{Xsi}B_{Ysi} - \left(\sum_{i=1}^{K} A_{Xsi}\right)^2 \cdot \sum_{i=1}^{K} B_{Xsi}B_{Ysi} \\ &- \sum_{i=1}^{K} B_{Ysi} \cdot \sum_{i=1}^{K} B_{Xsi} \cdot \sum_{i=1}^{K} A_{Xsi}^2 - N \cdot \sum_{i=1}^{K} A_{Xsi}B_{Xsi} \cdot \sum_{i=1}^{K} A_{Xsi}B_{Ysi}\end{aligned}$$

Ostatecznie, rozwiązanie możemy wyznaczyć w postaci:

$$\begin{aligned} u_{A1} &= \frac{\Delta_{u_{A1}}}{\Delta}, \\ u_{B1} &= \frac{\Delta_{u_{B1}}}{\Delta}, \\ u_0 &= \frac{1}{N} \cdot \left(\sum_{i=1}^{K} B_{Ysi} - u_{A1} \cdot \sum_{i=1}^{K} A_{Xsi} - u_{B1} \cdot \sum_{i=1}^{K} B_{Xsi}\right). \end{aligned} \tag{5.25}$$

Parametry modelu, wyznaczone według powyższych wzorów, nie zależą od współczynnika wagowego λ_A.

Ponieważ kryterium jakości (5.20) jest funkcją siedmiu parametrów, wyznaczony dla niego Hessian będzie macierzą o rozmiarze 7×7. W celu zbadania warunku dostatecznego, określającego czy współczynniki opisane równaniami (5.22), (5.23) i (5.25) faktycznie definiują minimum kryterium, należy zbadać, czy Hessian jest dodatnio określony. Badanie to jest realizowane podobnie jak dla interwałów (strona 33) i liczb rozmytych (strona 69), stąd ze względu na znaczny rozmiar niezbędnych wyprowadzeń nie zostanie tu zamieszczone. W podrozdziale kolejnym natomiast, przedstawione zostanie badanie warunku dostatecznego dla modelu z kilkoma wejściami (strona 119).

Przykład 5.1.

Dla danych z tabeli należy wyznaczyć modele aproksymujące w postaci (5.18) i (5.19), minimalizujące kryterium jakości (5.20).

A_X	B_X	A_Y	B_Y
TFN(1,1.5,2)	RFN(0.2,0.4)	TFN(1,1.5,2)	TFN(0.2,0.4,0.6)
TrFN(2,2.5,3.5,4)	TFN(0,0,0.5)	RFN(3,4)	TFN(0.3,0.5,0.7)
5	TFN(0,0.5,1)	TFN(2,2.5,3)	TFN(0,0,0.5)
4	TFN(0.5,1,1)	5	1
RFN(6,7)	0.5	3	RFN(0.2,0.4)
TFN(7,8,8)	RFN(0.5,1)	TrFN(4,4.5,5.5,6)	TFN(0.5,1,1)

Dla ułatwienia dalszych obliczeń, wyznaczamy wartości A_{Xs}, A_{Ys} oraz funkcje odwrotne lewego i prawego brzegu funkcji przynależności wyjściowych liczb rozmytych A_Y.

A_{Xs}	A_{Ys}	$f_{AY}^{-1}(\alpha)$	$g_{AY}^{-1}(\alpha)$
1.5	1.5	$\frac{\alpha}{2}+1$	$2-\frac{\alpha}{2}$
3	3.5	3	4
5	2.5	$\frac{\alpha}{2}+2$	$3-\frac{\alpha}{2}$
4	5	5	5
6.5	3	3	3
7.75	5	$\frac{\alpha}{2}+4$	$6-\frac{\alpha}{2}$

Podobnie, wyznaczamy wartości B_{Xs}, B_{Ys} oraz funkcje odwrotne lewego i prawego brzegu funkcji przynależności wyjściowych liczb rozmytych B_Y.

B_{Xs}	B_{Ys}	$f_{BY}^{-1}(\alpha)$	$g_{BY}^{-1}(\alpha)$
0.3	0.4	$0.2+\frac{\alpha}{5}$	$0.6-\frac{\alpha}{5}$
0.125	0.5	$0.3+\frac{\alpha}{5}$	$0.7-\frac{\alpha}{5}$
0.5	0.125	0	$0.5-\frac{\alpha}{2}$
0.875	1	1	1
0.5	0.3	0.2	0.4
0.75	0.875	$0.5+\frac{\alpha}{2}$	1

Parametry wyznaczone za pomocą zależności (5.22) mają wartość:

$$w_1 \approx 0.331, \qquad w_0 \approx 1.884.$$

Z tej samej zależności można wyznaczyć parametr $r_A = 0.5$. Wartości A_X i A_Y są identyczne jak w przykładzie 4.1 (s. 69) i tam też można znaleźć szczegóły obliczeń.

Współczynniki drugiego równania modelu: u_{A1}, u_{B1}, u_0 wyznaczyć można rozwiązując układ równań (5.24). Po podstawieniach i uproszczeniach przyjmuje on postać:

$$\begin{aligned} u_{A1} \cdot 154.562 + u_{B1} \cdot 15.8875 + u_0 \cdot 27.75 &= 15.45625 \\ u_{A1} \cdot 15.8875 + u_{B1} \cdot 1.93375 + u_0 \cdot 3.05 &= 1.92625 \\ u_{A1} \cdot 27.75 + u_{B1} \cdot 3.05 + u_0 \cdot 6 &= 3.2 \end{aligned}$$

Po rozwiązaniu otrzymujemy:

$$u_{A1} \approx -0.04101, \qquad u_{B1} \approx 0.9721, \qquad u_0 \approx 0.22887.$$

Ostatnim niewiadomym parametrem jest r_B. Można go obliczyć ze wzoru (5.23).

$$\begin{aligned} r_B =& \frac{3}{2 \cdot 6} \left[\int_0^1 (1-\alpha)(0.6 - \frac{\alpha}{5} - 0.2 - \frac{\alpha}{5}) \mathrm{d}\alpha + \int_0^1 (1-\alpha)(0.7 - \frac{\alpha}{5} - 0.3 - \frac{\alpha}{5}) \mathrm{d}\alpha \right. \\ &+ \int_0^1 (1-\alpha)(0.5 - \frac{\alpha}{2}) \mathrm{d}\alpha + \int_0^1 (1-\alpha)(1-1) \mathrm{d}\alpha + \int_0^1 (1-\alpha) \cdot 0.2 \mathrm{d}\alpha \\ &\left. + \int_0^1 (1-\alpha)(1 - 0.5 - \frac{\alpha}{2}) \mathrm{d}\alpha \right] = \frac{1}{4} \cdot \left[\frac{2}{15} + \frac{2}{15} + \frac{1}{6} + \frac{1}{10} + \frac{1}{6} \right] = \frac{7}{40} = 0.175 \end{aligned}$$

5.4.2. Model z wieloma wejściami

Załóżmy dalej, że zmienna zależna Z_Y zależy od N zmiennych objaśniających Z_{Xi}, $i = 1 \ldots N$, przy czym zarówno Z_{Xi} jak i Z_Y będą miały postać liczb Z.

Ponieważ liczba $Z = (A, B)$ jest parą liczb rozmytych, będziemy poszukiwali dwóch modeli aproksymujących. Przy wykonywaniu działań na liczbach Z, wartość liczby A zależy jedynie od wartości $A_1, A_2, \ldots$ występujących w operandach, stąd pierwszy model będzie miał postać:

$$\hat{A}_Y = \mathbf{w}^T \cdot \mathbf{A_{Xs}} + R_A, \tag{5.26}$$

gdzie: $\mathbf{A_{Xs}} = [1, A_{Xs1}, \ldots, A_{XsN}]$ oznacza wektor środków liczb rozmytych, składających się na wartości A wejściowych liczb Z, R_A – trójkątna liczba rozmyta w postaci TFN$(-r_A, 0, r_A)$, $\mathbf{w} = [w_0, w_1, \ldots w_N]^T$ – wektor rzeczywistych współczynników modelu.

Prawdopodobieństwo B wyznaczane jest w bardziej złożony sposób i zależy zarówno od wartości $A_1, A_2, \ldots$ jak i $B_1, B_2, \ldots$ występujących w operandach, stąd drugi model będzie miał postać:

$$\hat{B}_Y = \mathbf{u}^T \cdot \mathbf{C_{Xs}} + R_B, \tag{5.27}$$

gdzie: $\mathbf{C_{Xs}} = [1, A_{Xs1}, \ldots, A_{XsN}, B_{Xs1}, \ldots, B_{XsN}]$ oznacza wektor środków liczb rozmytych, składających się na wartości A i prawdopodobieństwa B wejściowych liczb Z, R_B – trójkątna liczba rozmyta w postaci TFN$(-r_B, 0, r_B)$, $\mathbf{u} = [u_0, u_1, \ldots u_N]^T$ – wektor rzeczywistych współczynników modelu.

Model utworzony będzie na podstawie K danych pomiarowych, które zapisać możemy w formie macierzy:

$$\mathbf{Z_X} = \begin{bmatrix} 1 & Z_{X_{11}} & Z_{X_{21}} & \ldots & Z_{X_{N1}} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & Z_{X_{1K}} & Z_{X_{2K}} & \ldots & Z_{X_{NK}} \end{bmatrix}, \qquad \mathbf{Z_Y} = \begin{bmatrix} Z_{Y_1} \\ \vdots \\ Z_{Y_K} \end{bmatrix}. \tag{5.28}$$

Jak już wspominano wcześniej, zarówno dane opisujące zmienne objaśniające $Z_{X_{ij}}$, $i = 1 \ldots N$, $j = 1 \ldots K$, jak i dane opisujące zmienną zależną Z_{Y_j}, mają postać liczb Z bądź mogą być do takiej postaci sprowadzone.

Będziemy chcieli, aby modele (5.26) i (5.27) dopasowane były do danych w taki sposób, aby minimalizować kryterium jakości oparte na metryce D_2 opisanej zależnością (5.16):

$$Q(\mathbf{w}, r_A, \mathbf{u}, r_B) = \sum_{i=1}^{K} D_2^2(\hat{Z}_{Yi}, Z_{Yi}) \longrightarrow \min, \tag{5.29}$$

gdzie: $\hat{Z}_{Yi} = (\mathbf{w}^T \cdot \mathbf{A_{Xs}} + R_A, \mathbf{u}^T \cdot \mathbf{C_{Xs}} + R_B)$.

Oznaczmy przez $\mathbf{F_{AY}^{-1}}$, $\mathbf{F_{BY}^{-1}}$ wektory odwrotności funkcji lewych brzegów, a przez $\mathbf{G_{AY}^{-1}}$, $\mathbf{G_{BY}^{-1}}$ wektory odwrotności funkcji prawych brzegów liczb rozmytych stanowiących zadane wartości zmiennej zależnej $Z_Y = (A_Y, B_Y)$:

$$\mathbf{F_{AY}^{-1}} = [f_{AY1}^{-1}, f_{AY2}^{-1}, \ldots f_{AYK}^{-1}]^T, \qquad \mathbf{G_{AY}^{-1}} = [g_{AY1}^{-1}, g_{AY2}^{-1}, \ldots g_{AYK}^{-1}]^T,$$

$$\mathbf{F_{BY}^{-1}} = [f_{BY1}^{-1}, f_{BY2}^{-1}, \ldots f_{BYK}^{-1}]^T, \qquad \mathbf{G_{BY}^{-1}} = [g_{BY1}^{-1}, g_{BY2}^{-1}, \ldots g_{BYK}^{-1}]^T,$$

oraz przez $\mathbf{A_{Xs}}$, $\mathbf{C_{Xs}}$ macierze środków liczb rozmytych będących elementami macierzy $\mathbf{Z_X}$:

$$\mathbf{A_{Xs}} = \begin{bmatrix} 1 & A_{Xs_{11}} & A_{Xs_{21}} & \ldots & A_{Xs_{N1}} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & A_{Xs_{1K}} & A_{Xs_{2K}} & \ldots & A_{Xs_{NK}} \end{bmatrix},$$

$$\mathbf{C_{Xs}} = \begin{bmatrix} 1 & A_{Xs_{11}} & A_{Xs_{21}} & \ldots & A_{Xs_{N1}} & B_{Xs_{11}} & B_{Xs_{21}} & \ldots & B_{Xs_{N1}} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & A_{Xs_{1K}} & A_{Xs_{2K}} & \ldots & A_{Xs_{NK}} & B_{Xs_{1K}} & B_{Xs_{2K}} & \ldots & B_{Xs_{NK}} \end{bmatrix}.$$

Przyjmując, że odległość między liczbami rozmytymi obliczana będzie z zależności (4.39), kryterium jakości (5.29) możemy zapisać inaczej w postaci:

$$\begin{aligned}
Q(\mathbf{w}, r_A, \mathbf{u}, r_B) &= \sum_{i=1}^{K} D_2^2(\hat{Z}_{Yi}, Z_{Yi}) = \sum_{i=1}^{K} \lambda_A \cdot d_2^2(\hat{A}_{Yi}, A_{Yi}) + d_2^2(\hat{B}_{Yi}, B_{Yi}) \\
&= \lambda_A \cdot \left(\frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[\mathbf{w}^T \cdot \mathbf{A_{Xsi}} + r_A(\alpha - 1) - f_{AYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right. \\
&\quad \left. + \frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[\mathbf{w}^T \cdot \mathbf{A_{Xsi}} + r_A(1 - \alpha) - g_{AYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right) \\
&\quad + \left(\frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[\mathbf{u}^T \cdot \mathbf{C_{Xsi}} + r_B(\alpha - 1) - f_{BYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right. \\
&\quad \left. + \frac{1}{2} \sum_{i=1}^{K} \int_0^1 \left[\mathbf{u}^T \cdot \mathbf{C_{Xsi}} + r_B(1 - \alpha) - g_{BYi}^{-1}(\alpha) \right]^2 \mathrm{d}\alpha \right).
\end{aligned}$$

W zapisie macierzowym, kryterium będzie miało postać:

$$\begin{aligned}
&Q(\mathbf{w}, r_A, \mathbf{u}, r_B) \\
&= \lambda_A \cdot \left(\frac{1}{2} \int_0^1 \left[\mathbf{A_{Xs}w} + \mathbf{r_A}(\alpha - 1) - \mathbf{F_{AY}}^{-1}(\alpha) \right]^T \left[\mathbf{A_{Xs}w} + \mathbf{r_A}(\alpha - 1) - \mathbf{F_{AY}}^{-1}(\alpha) \right] \right. \\
&\qquad \left. + \left[\mathbf{A_{Xs}w} + \mathbf{r_A}(1 - \alpha) - \mathbf{G_{AY}}^{-1}(\alpha) \right]^T \left[\mathbf{A_{Xs}w} + \mathbf{r_A}(1 - \alpha) - \mathbf{G_{AY}}^{-1}(\alpha) \right] \mathrm{d}\alpha \right) \\
&\quad + \left(\frac{1}{2} \int_0^1 \left[\mathbf{C_{Xs}u} + \mathbf{r_B}(\alpha - 1) - \mathbf{F_{BY}}^{-1}(\alpha) \right]^T \left[\mathbf{C_{Xs}u} + \mathbf{r_B}(\alpha - 1) - \mathbf{F_{BY}}^{-1}(\alpha) \right] \right. \\
&\qquad \left. + \left[\mathbf{C_{Xs}u} + \mathbf{r_B}(1 - \alpha) - \mathbf{G_{BY}}^{-1}(\alpha) \right]^T \left[\mathbf{C_{Xs}u} + \mathbf{r_B}(1 - \alpha) - \mathbf{G_{BY}}^{-1}(\alpha) \right] \mathrm{d}\alpha \right),
\end{aligned}$$

gdzie: $\mathbf{r_A}$ – kolumnowy wektor K takich samych wartości r_A, $\mathbf{r_B}$ – kolumnowy wektor K takich samych wartości r_B.

Po wymnożeniu i uproszczeniu otrzymujemy:

$$
\begin{aligned}
&Q(\mathbf{w}, r_A, \mathbf{u}, r_B) \\
&\quad = \lambda_A \cdot \left(\frac{1}{2} \int_0^1 \mathbf{w}^T {\mathbf{A_{Xs}}}^T \mathbf{A_{Xs}} \mathbf{w} - (\alpha - 1) {\mathbf{r_A}}^T \mathbf{A_{Xs}} \mathbf{w} - {\mathbf{F_{AY}}^{-1}(\alpha)}^T \mathbf{A_{Xs}} \mathbf{w} + \mathbf{w}^T {\mathbf{A_{Xs}}}^T \mathbf{r_A} (\alpha - 1) \right. \\
&\qquad\qquad + {\mathbf{r_A}}^T \mathbf{r_A} (\alpha - 1)^2 - {\mathbf{F_{AY}}^{-1}(\alpha)}^T \mathbf{r_A} (\alpha - 1) - \mathbf{w}^T {\mathbf{A_{Xs}}}^T {\mathbf{F_{AY}}^{-1}(\alpha)} \\
&\qquad\qquad - {\mathbf{r_A}}^T {\mathbf{F_{AY}}^{-1}(\alpha)} (\alpha - 1) + {\mathbf{F_{AY}}^{-1}(\alpha)}^T {\mathbf{A_{Xs}}^{-1}(\alpha)} + \mathbf{w}^T {\mathbf{A_{Xs}}}^T \mathbf{A_{Xs}} \mathbf{w} \\
&\qquad\qquad - (1 - \alpha) {\mathbf{r_A}}^T \mathbf{A_{Xs}} \mathbf{w} - {\mathbf{G_{AY}}^{-1}(\alpha)}^T \mathbf{A_{Xs}} \mathbf{w} + \mathbf{w}^T {\mathbf{A_{Xs}}}^T \mathbf{r} (1 - \alpha) + \mathbf{r}^T \mathbf{r_A} (1 - \alpha)^2 \\
&\qquad\qquad - {\mathbf{G_{AY}}^{-1}(\alpha)}^T \mathbf{r_A} (1 - \alpha) - \mathbf{w}^T {\mathbf{A_{Xs}}}^T {\mathbf{G_{AY}}^{-1}(\alpha)} - {\mathbf{r_A}}^T {\mathbf{G_{AY}}^{-1}(\alpha)} (1 - \alpha) \\
&\qquad\qquad \left. + {\mathbf{G_{AY}}^{-1}(\alpha)}^T {\mathbf{G_{AY}}^{-1}(\alpha)} \, \mathrm{d}\alpha \right) \\
&\qquad + \left(\frac{1}{2} \int_0^1 \mathbf{u}^T {\mathbf{C_{Xs}}}^T \mathbf{C_{Xs}} \mathbf{u} - (\alpha - 1) {\mathbf{r_B}}^T \mathbf{C_{Xs}} \mathbf{u} - {\mathbf{F_{BY}}^{-1}(\alpha)}^T \mathbf{C_{Xs}} \mathbf{u} + \mathbf{u}^T {\mathbf{C_{Xs}}}^T \mathbf{r_B} (\alpha - 1) \right. \\
&\qquad\qquad + {\mathbf{r_B}}^T \mathbf{r_B} (\alpha - 1)^2 - {\mathbf{F_{BY}}^{-1}(\alpha)}^T \mathbf{r_B} (\alpha - 1) - \mathbf{u}^T {\mathbf{C_{Xs}}}^T {\mathbf{F_{BY}}^{-1}(\alpha)} \\
&\qquad\qquad - {\mathbf{r_B}}^T {\mathbf{F_{BY}}^{-1}(\alpha)} (\alpha - 1) + {\mathbf{F_{BY}}^{-1}(\alpha)}^T {\mathbf{F_{BY}}^{-1}(\alpha)} + \mathbf{u}^T {\mathbf{C_{Xs}}}^T \mathbf{C_{Xs}} \mathbf{u} \\
&\qquad\qquad - (1 - \alpha) {\mathbf{r_B}}^T \mathbf{C_{Xs}} \mathbf{u} - {\mathbf{G_{BY}}^{-1}(\alpha)}^T \mathbf{C_{Xs}} \mathbf{u} + \mathbf{u}^T {\mathbf{C_{Xs}}}^T \mathbf{r} (1 - \alpha) + \mathbf{r}^T \mathbf{r_B} (1 - \alpha)^2 \\
&\qquad\qquad - {\mathbf{G_{BY}}^{-1}(\alpha)}^T \mathbf{r_B} (1 - \alpha) - \mathbf{u}^T {\mathbf{C_{Xs}}}^T {\mathbf{G_{BY}}^{-1}(\alpha)} - {\mathbf{r_B}}^T {\mathbf{G_{BY}}^{-1}(\alpha)} (1 - \alpha) \\
&\qquad\qquad \left. + {\mathbf{G_{BY}}^{-1}(\alpha)}^T {\mathbf{G_{BY}}^{-1}(\alpha)} \, \mathrm{d}\alpha \right) \\
&\quad = \lambda_A \cdot \left(\frac{1}{2} \int_0^1 2 \mathbf{w}^T {\mathbf{A_{Xs}}}^T \mathbf{A_{Xs}} \mathbf{w} - 2 [{\mathbf{F_{AY}}^{-1}(\alpha)}^T + {\mathbf{G_{AY}}^{-1}(\alpha)}^T] \mathbf{A_{Xs}} \mathbf{w} + 2 {\mathbf{r_A}}^T \mathbf{r_A} (\alpha - 1)^2 \right. \\
&\qquad\qquad - 2 [{\mathbf{G_{AY}}^{-1}(\alpha)}^T - {\mathbf{F_{AY}}^{-1}(\alpha)}^T] \mathbf{r_A} (1 - \alpha) + {\mathbf{F_{AY}}^{-1}(\alpha)}^T {\mathbf{F_{AY}}^{-1}(\alpha)} \\
&\qquad\qquad \left. + {\mathbf{G_{AY}}^{-1}(\alpha)}^T {\mathbf{G_{AY}}^{-1}(\alpha)} \mathrm{d}\alpha \right) \\
&\qquad + \left(\frac{1}{2} \int_0^1 2 \mathbf{u}^T {\mathbf{C_{Xs}}}^T \mathbf{C_{Xs}} \mathbf{u} - 2 [{\mathbf{F_{BY}}^{-1}(\alpha)}^T + {\mathbf{G_{BY}}^{-1}(\alpha)}^T] \mathbf{C_{Xs}} \mathbf{u} + 2 {\mathbf{r_B}}^T \mathbf{r_B} (\alpha - 1)^2 \right. \\
&\qquad\qquad - 2 [{\mathbf{G_{BY}}^{-1}(\alpha)}^T - {\mathbf{F_{BY}}^{-1}(\alpha)}^T] \mathbf{r_B} (1 - \alpha) + {\mathbf{F_{BY}}^{-1}(\alpha)}^T {\mathbf{F_{BY}}^{-1}(\alpha)} \\
&\qquad\qquad \left. + {\mathbf{G_{BY}}^{-1}(\alpha)}^T {\mathbf{G_{BY}}^{-1}(\alpha)} \mathrm{d}\alpha \right).
\end{aligned}
$$

Aby wyznaczyć minimum kryterium (w skrócie oznaczanego dalej jako Q) liczymy pochodne cząstkowe po parametrach modelu i przyrównujemy je do 0. Wartość pierwszej całki w kryterium zależy jedynie od wektora $\mathbf{w}$ i szerokości r_A, natomiast wartość całki

drugiej zależy tylko od wektora $\mathbf{u}$ i szerokości r_B. Wynika z tego, że zadanie wyznaczania parametrów możemy rozdzielić na dwa niezależne etapy.

W etapie pierwszym wyznaczamy $\mathbf{w}$ i r_A.

$$\frac{\partial Q}{\partial \mathbf{w}} = \lambda_A \cdot \int_0^1 2{\mathbf{A_{Xs}}}^T \mathbf{A_{Xs}}\mathbf{w} - {\mathbf{A_{Xs}}}^T[{\mathbf{F_{AY}}}^{-1}(\alpha) + {\mathbf{G_{AY}}}^{-1}(\alpha)]\,\mathrm{d}\alpha = 0$$

$$2{\mathbf{A_{Xs}}}^T \mathbf{A_{Xs}}\mathbf{w} = {\mathbf{A_{Xs}}}^T \int_0^1 [{\mathbf{F_{AY}}}^{-1}(\alpha) + {\mathbf{G_{AY}}}^{-1}(\alpha)]\,\mathrm{d}\alpha$$

Ostatecznie:

$$\mathbf{w} = ({\mathbf{A_{Xs}}}^T \mathbf{A_{Xs}})^{-1}{\mathbf{A_{Xs}}}^T \cdot \int_0^1 \frac{{\mathbf{F_{AY}}}^{-1}(\alpha) + {\mathbf{G_{AY}}}^{-1}(\alpha)}{2}\,\mathrm{d}\alpha. \tag{5.30}$$

$$\begin{aligned}
\frac{\partial Q}{\partial \mathbf{r_A}} &= \frac{\partial}{\partial \mathbf{r_A}}\left[\lambda_A \cdot \int_0^1 {\mathbf{r_A}}^T \mathbf{r_A}(\alpha-1)^2 - [{\mathbf{G_{AY}}}^{-1}(\alpha)^T - {\mathbf{F_{AY}}}^{-1}(\alpha)^T]\mathbf{r_A}(1-\alpha)\mathrm{d}\alpha\right] \\
&= \frac{\partial}{\partial \mathbf{r_A}}\left[\lambda_A \cdot \int_0^1 K r_A^2(\alpha-1)^2 - \sum_{i=1}^{K}[g_{AYi}^{-1} - f_{AYi}^{-1}]r_A(1-\alpha)\mathrm{d}\alpha\right] \\
&= 2Kr_A\lambda_A \int_0^1 (\alpha-1)^2\mathrm{d}\alpha - \lambda_A \int_0^1 \sum_{i=1}^{K}[g_{AYi}^{-1} - f_{AYi}^{-1}](1-\alpha)\mathrm{d}\alpha = 0
\end{aligned}$$

Rozwiązując równanie:

$$2Kr_A\frac{1}{3} = \int_0^1 \sum_{i=1}^{K}[g_{AYi}^{-1} - f_{AYi}^{-1}](1-\alpha)\mathrm{d}\alpha,$$

otrzymujemy wzór, z którego można wyznaczyć r_A:

$$r_A = \frac{3}{2K}\int_0^1 \sum_{i=1}^{K}[g_{AYi}^{-1} - f_{AYi}^{-1}](1-\alpha)\mathrm{d}\alpha. \tag{5.31}$$

W etapie drugim wyznaczamy $\mathbf{u}$ i r_B.

$$\frac{\partial Q}{\partial \mathbf{u}} = \int_0^1 2{\mathbf{C_{Xs}}}^T \mathbf{C_{Xs}}\mathbf{u} - {\mathbf{C_{Xs}}}^T[{\mathbf{F_{BY}}}^{-1}(\alpha) + {\mathbf{G_{BY}}}^{-1}(\alpha)]\,\mathrm{d}\alpha = 0$$

$$2{\mathbf{C_{Xs}}}^T \mathbf{C_{Xs}}\mathbf{w} = {\mathbf{C_{Xs}}}^T \int_0^1 [{\mathbf{F_{BY}}}^{-1}(\alpha) + {\mathbf{G_{BY}}}^{-1}(\alpha)]\,\mathrm{d}\alpha$$

Ostatecznie:

$$\mathbf{u} = (\mathbf{C_{Xs}}^T\mathbf{C_{Xs}})^{-1}\mathbf{C_{Xs}}^T \cdot \int_0^1 \frac{\mathbf{F_{BY}}^{-1}(\alpha) + \mathbf{G_{BY}}^{-1}(\alpha)}{2}\,\mathrm{d}\alpha. \tag{5.32}$$

Podobnie jak wcześniej:

$$\begin{aligned}\frac{\partial Q}{\partial \mathbf{r_B}} &= \frac{\partial}{\partial \mathbf{r_B}}\left[\int_0^1 \mathbf{r_B}^T\mathbf{r_B}(\alpha-1)^2 - [\mathbf{G_{BY}}^{-1}(\alpha)^T - \mathbf{F_{BY}}^{-1}(\alpha)^T]\mathbf{r_B}(1-\alpha)\mathrm{d}\alpha\right]\\ &= \frac{\partial}{\partial \mathbf{r_B}}\left[\int_0^1 Kr_B^2(\alpha-1)^2 - \sum_{i=1}^K [g_{BYi}^{-1} - f_{BYi}^{-1}]r_B(1-\alpha)\mathrm{d}\alpha\right]\\ &= 2Kr_B\int_0^1(\alpha-1)^2\mathrm{d}\alpha - \int_0^1\sum_{i=1}^K[g_{BYi}^{-1} - f_{BYi}^{-1}](1-\alpha)\mathrm{d}\alpha = 0.\end{aligned}$$

Po przekształceniach otrzymujemy:

$$r_B = \frac{3}{2K}\int_0^1\sum_{i=1}^K[g_{BYi}^{-1} - f_{BYi}^{-1}](1-\alpha)\mathrm{d}\alpha. \tag{5.33}$$

Badamy dalej warunek dostateczny. Aby kryterium (5.29) osiągało minimum w punkcie opisanym zależnościami (5.30) – (5.33), jego Hessian w tym punkcie musi być dodatnio określony. Ma on postać:

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 Q}{\partial \mathbf{w}\partial \mathbf{w}^T} & \frac{\partial^2 Q}{\partial \mathbf{w}\partial r_A} & \frac{\partial^2 Q}{\partial \mathbf{w}\partial \mathbf{u}^T} & \frac{\partial^2 Q}{\partial \mathbf{w}\partial r_B}\\ \frac{\partial^2 Q}{\partial r_A\partial \mathbf{w}} & \frac{\partial^2 Q}{\partial r_A^2} & \frac{\partial^2 Q}{\partial r_A\partial \mathbf{u}} & \frac{\partial^2 Q}{\partial r_A\partial r_B}\\ \frac{\partial^2 Q}{\partial \mathbf{u}\partial \mathbf{w}^T} & \frac{\partial^2 Q}{\partial \mathbf{u}\partial r_A} & \frac{\partial^2 Q}{\partial \mathbf{u}\partial \mathbf{u}^T} & \frac{\partial^2 Q}{\partial \mathbf{u}\partial r_B}\\ \frac{\partial^2 Q}{\partial r_B\partial \mathbf{w}} & \frac{\partial^2 Q}{\partial r_B\partial r_A} & \frac{\partial^2 Q}{\partial r_B\partial \mathbf{u}} & \frac{\partial^2 Q}{\partial r_B^2} \end{bmatrix} = \begin{bmatrix} \lambda_A\cdot 2\mathbf{A_{Xs}}^T\mathbf{A_{Xs}} & 0 & 0 & 0\\ 0 & \lambda_A\cdot\frac{2}{3}K & 0 & 0\\ 0 & 0 & 2\mathbf{C_{Xs}}^T\mathbf{C_{Xs}} & 0\\ 0 & 0 & 0 & \frac{2}{3}K \end{bmatrix}. \tag{5.34}$$

Współczynnik λ_A jest zawsze dodatni. Ponieważ macierze $\mathbf{A_{Xs}}^T\mathbf{A_{Xs}}$ i $\mathbf{C_{Xs}}^T\mathbf{C_{Xs}}$ są dodatnio określone jeśli tylko są nieosobliwe, stąd warunek dodatniej określoności Hessianu jest spełniony.

Przykład 5.2.

Dla danych z tabeli należy wyznaczyć modele aproksymujące w postaci (5.26) i (5.26), minimalizujące kryterium jakości (5.29).

A_{X1}	B_{X1}	A_{X2}	B_{X2}	A_Y	B_Y
TFN(1,1.5,2)	RFN(0.2,0.4)	RFN(0,1)	RFN(0,0.1)	TFN(1,1.5,2)	TFN(0.2,0.4,0.6)
TrFN(2,2.5,3.5,4)	TFN(0,0,0.5)	TFN(1,2,3)	TFN(0.2,0.3,0.4)	RFN(3,4)	TFN(0.3,0.5,0.7)
5	TFN(0,0.5,1)	2	0.6	TFN(2,2.5,3)	TFN(0,0,0.5)
4	TFN(0.5,1,1)	Tr(3,3.5,4.5,5)	TFN(0.5,1,1)	5	1
RFN(6,7)	0.5	RFN(3,5)	RFN(0.5,0.7)	3	RFN(0.2,0.4)
TFN(7,8,8)	RFN(0.5,1)	TFN(4,4.5,5)	RFN(0.8,1)	TrFN(4,4.5,5.5,6)	TFN(0.5,1,1)

Dla ułatwienia dalszych obliczeń, wyznaczamy wartości A_{Xs1}, A_{Xs2}, A_{Ys} oraz funkcje odwrotne lewego i prawego brzegu funkcji przynależności wyjściowych liczb rozmytych A_Y.

A_{Xs1}	A_{Xs2}	A_{Ys}	$f_{AY}^{-1}(\alpha)$	$g_{AY}^{-1}(\alpha)$
1.5	0.5	1.5	$\frac{\alpha}{2}+1$	$2-\frac{\alpha}{2}$
3	2	3.5	3	4
5	2	2.5	$\frac{\alpha}{2}+2$	$3-\frac{\alpha}{2}$
4	4	5	5	5
6.5	4	3	3	3
7.75	4.5	5	$\frac{\alpha}{2}+4$	$6-\frac{\alpha}{2}$

Podobnie wyznaczamy wartości B_{Xs1}, B_{Xs2}, B_{Ys} oraz funkcje odwrotne lewego i prawego brzegu funkcji przynależności wyjściowych liczb rozmytych B_Y.

B_{Xs1}	B_{Xs2}	B_{Ys}	$f_{BY}^{-1}(\alpha)$	$g_{BY}^{-1}(\alpha)$
0.3	0.05	0.4	$0.2+\frac{\alpha}{5}$	$0.6-\frac{\alpha}{5}$
0.125	0.3	0.5	$0.3+\frac{\alpha}{5}$	$0.7-\frac{\alpha}{5}$
0.5	0.6	0.125	0	$0.5-\frac{\alpha}{2}$
0.875	0.875	1	1	1
0.5	0.6	0.3	0.2	0.4
0.75	0.9	0.875	$0.5+\frac{\alpha}{2}$	1

Na podstawie danych z tabel, możemy utworzyć macierze $\mathbf{A_{Xs}}$ i $\mathbf{C_{Xs}}$.

$$\mathbf{A_{Xs}} = \begin{bmatrix} 1 & 1.5 & 0.5 \\ 1 & 3 & 2 \\ 1 & 5 & 2 \\ 1 & 4 & 4 \\ 1 & 6.5 & 4 \\ 1 & 7.75 & 4.5 \end{bmatrix}, \quad \mathbf{C_{Xs}} = \begin{bmatrix} 1 & 1.5 & 0.5 & 0.3 & 0.05 \\ 1 & 3 & 2 & 0.125 & 0.3 \\ 1 & 5 & 2 & 0.5 & 0.6 \\ 1 & 4 & 4 & 0.875 & 0.875 \\ 1 & 6.5 & 4 & 0.5 & 0.6 \\ 1 & 7.75 & 4.5 & 0.75 & 0.9 \end{bmatrix}.$$

Ze wzoru (5.30) wyznaczamy wektor współczynników $\mathbf{w}$:

$$\mathbf{w} = [w_0, w_1, w_2]^T = [1.6747, -0.3086, 1.1186]^T,$$

a ze wzoru (5.32) wektor współczynników $\mathbf{u}$:

$$\mathbf{u} = [u_0, u_{A1}, u_{A2}, u_{B1}, u_{B2}]^T = [0.2945, -0.1396, 0.2351, 0.4033, 0.0237]^T.$$

Parametry $r_A = 0.5$ i $r_B = 0.175$ przyjmują identyczne wartości jak w przykładzie 5.1 (s. 112).

5.5. Modelowanie lokalne dla danych w postaci liczb Z

Podobnie jak w przypadku interwałów i liczb rozmytych, sposób modelowania lokalnego opisany w podrozdziale 2.1 będzie wymagał modyfikacji, ponieważ zarówno punkt wejściowy $\mathbf{x}^*$ jak i dane uczące będą teraz liczbami Z [110]. W przypadku gdy któraś z analizowanych wartości będzie liczbą rzeczywistą, zastąpimy ją jednoelementową liczbą rozmytą (typu singleton). Jeśli wartość będzie interwałem, możemy ją zastąpić prostokątną liczbą rozmytą. W obu tych przypadkach jednakże, nie jest określona w sposób

jednoznaczny wartość prawdopodobieństwa ich wystąpienia. Ponieważ w obliczeniach na liczbach Z wartość ta musi być znana, będziemy zakładali, że wartości te są pewne, czyli opisuje je jednoelementowa liczba rozmyta równa 1.

Aby wyznaczyć najbliższych sąsiadów punktu wejściowego $\mathbf{x}^*$, będziemy badać odległość między wielowymiarowymi danymi w postaci liczb Z. Wyznaczymy ją jako:

$$D(\mathbf{x}, \mathbf{x}^*) = \sqrt{\sum_{i=1}^{N} D_2^2(x_i, x_i^*)}, \tag{5.35}$$

gdzie: $\mathbf{x}, \mathbf{x}^*$ – wektory liczb Z, N – rozmiar wektora danych, $D_2(x_i, x_i^*)$ – odległość między liczbami Z wyznaczona z zależności (5.16).

Modelowanie lokalne dla liczb Z, z wykorzystaniem metody kNN, może być zrealizowane identycznie jak dla danych rzeczywistych z wykorzystaniem zasad arytmetyki liczb Z opisanej w podrozdziale 5.2. Można w ten sposób wyznaczyć średnią arytmetyczną bądź średnią ważoną z wyjść zadanych dla próbek będących sąsiadami punktu wejściowego $\mathbf{x}^*$.

W przypadku mini-modelu liniowego będziemy poszukiwali dwóch regresji:

$$f_A(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{A_{xs}}^* + R_A, \tag{5.36}$$

$$f_B(\mathbf{x}^*) = \mathbf{u}^T \cdot \mathbf{C_{xs}}^* + R_B, \tag{5.37}$$

w sposób opisany w podrozdziale 5.4. Dla mini-modelu nieliniowego regresje będą miały postać:

$$f_A(\mathbf{x}^*) = \mathbf{w}^T \cdot \mathbf{A_{xs}}^* + R_A + f_N(\mathbf{A_{xs}}^*), \tag{5.38}$$

$$f_B(\mathbf{x}^*) = \mathbf{u}^T \cdot \mathbf{C_{xs}}^* + R_B + f_N(\mathbf{C_{xs}}^*), \tag{5.39}$$

przy czym składnik nieliniowy wyznaczany będzie dla środków liczb rozmytych stanowiących dane uczące i środków liczb rozmytych stanowiących punkt wejściowy $\mathbf{x}^*$.

5.5.1. Dokładność modelu

Załóżmy, że podobnie jak wcześniej mamy do dyspozycji 2 znormalizowane, rozłączne zbiory danych. Pierwszy z nich będzie zawierał dane uczące (generujące modele lokalne), a każda próbka składać się będzie z wektora wejściowych liczb Z $\mathbf{x}_j$ i zadanej wartości wyjściowej y_j, również w postaci liczby Z, $j = 1 \ldots L$. Drugi zbiór danych zawierał będzie dane walidujące i, podobnie, każda próbka składać się będzie z wektora wejściowych liczb Z $\mathbf{x}_i$ i zadanej Z-wartości wyjściowej y_i, $i = 1 \ldots M$.

Miarą dokładności modelu może być jego średni błąd bezwzględny wyznaczony dla danych walidujących:

$$e_{MAE} = \frac{1}{M} \cdot \sum_{i=1}^{M} D_2(y_i, y_i^*), \tag{5.40}$$

gdzie: y_i – to wyjście zadane dla próbki i, a y_i^* – to wyjście wyznaczone przez model. Wyznaczony w ten sposób błąd będzie liczbą rzeczywistą.

Jeżeli nie jesteśmy w stanie wyodrębnić z posiadanych danych rozłącznej części próbek do walidacji jakości modelu, możemy zastosować kroswalidację (walidację krzyżową). Wyznaczanie błędu kroswalidacji realizowane jest podobnie jak dla danych walidujących za pomocą wzoru (5.40).

5.6. Eksperymenty

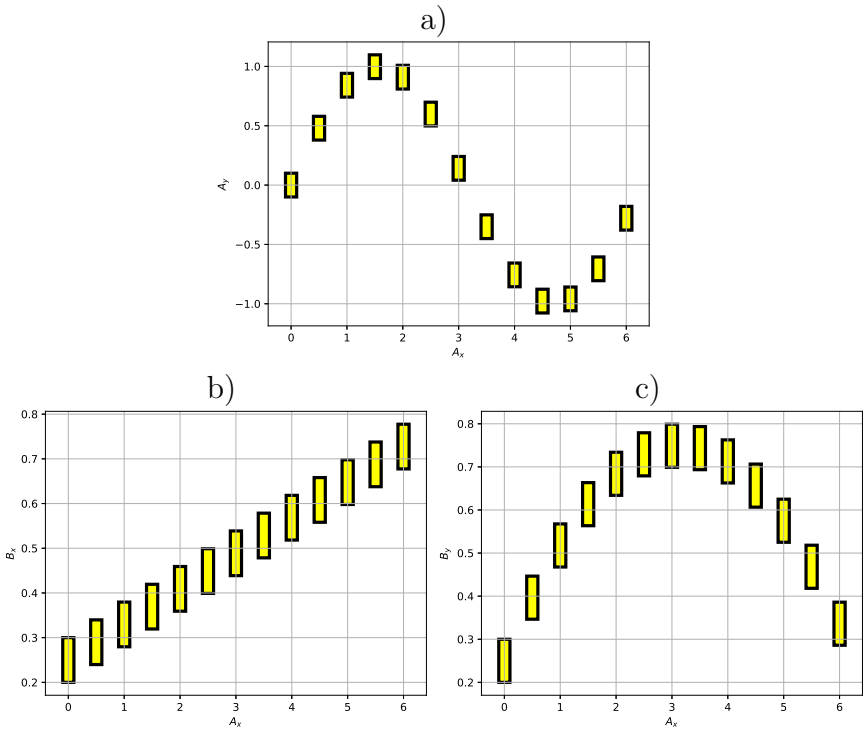
Badania nr 1 – jednorodne dane jednowejściowe

Podobnie jak w przypadku interwałów i liczb rozmytych, w pierwszym eksperymencie wykorzystane zostaną proste dane jednowejściowe. Jego celem będzie, przede wszystkim, zbadanie poprawności metod opisanych we wcześniejszych podrozdziałach.

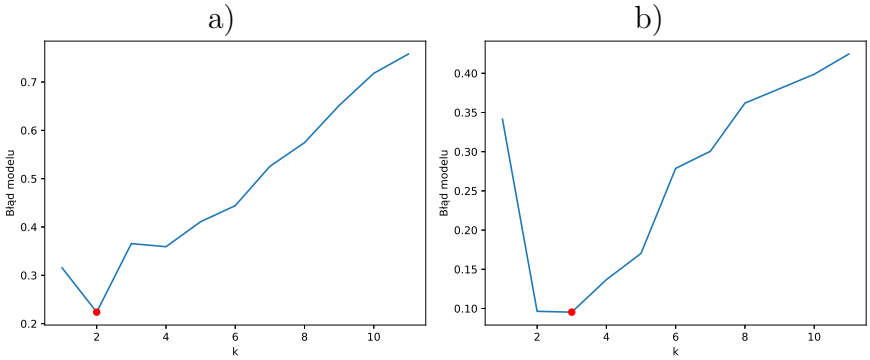
Dane opisują sinusoidę próbkowaną co 0.5 dla wejścia $x \in [0, 2\pi]$ i składać się będą z 13 próbek pokazanych na rys. 5.10a. Dane poddane zostały rozmyciu, przy czym zabiegowi temu podlegał zarówno atrybut wejściowy, jak i wyjściowy danych. W trakcie rozmycia, każda rzeczywista wartość x była zastępowana trójkątną liczbą rozmytą $\mathrm{TFN}(x - \alpha_A, x, x + \alpha_A)$. W badaniach przyjęto $\alpha_A = 0.1$. Dodatkowo, każdej wartości wejściowej i wyjściowej przypisane zostało prawdopodobieństwo wystąpienia. Miało ono formę symetrycznej, trójkątnej liczby rozmytej z nośnikiem o szerokości równej 0.1. Wartości prawdopodobieństw dla poszczególnych próbek zaprezentowano na rys. 5.10b i 5.10c.

Na rys. 5.11 pokazano w jaki sposób zmienia się bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla różnych metod modelowania. Wynika z niego, że optymalna liczba dla prostej metody kNN wynosi $k = 2$ (podobny przebieg wykresu zaobserwowano dla ważonej metody kNN i modelu bazującego na mini-modelach liniowych). Dla modelu bazującego na mini-modelach nieliniowych wartość ta wynosi $k = 3$.

Na rysunkach 5.12 – 5.19 pokazane są charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Charakterystyki sporządzone zostały dla różnych ilości najbliższych sąsiadów k, przy czym dla metod kNN parametr k przyjmował wartości od 1 do 4, a dla metod bazujących na mini-modelach wartości od 2 do 4 (ze względu na minimalną liczbę próbek wymaganą przy wyznacza-

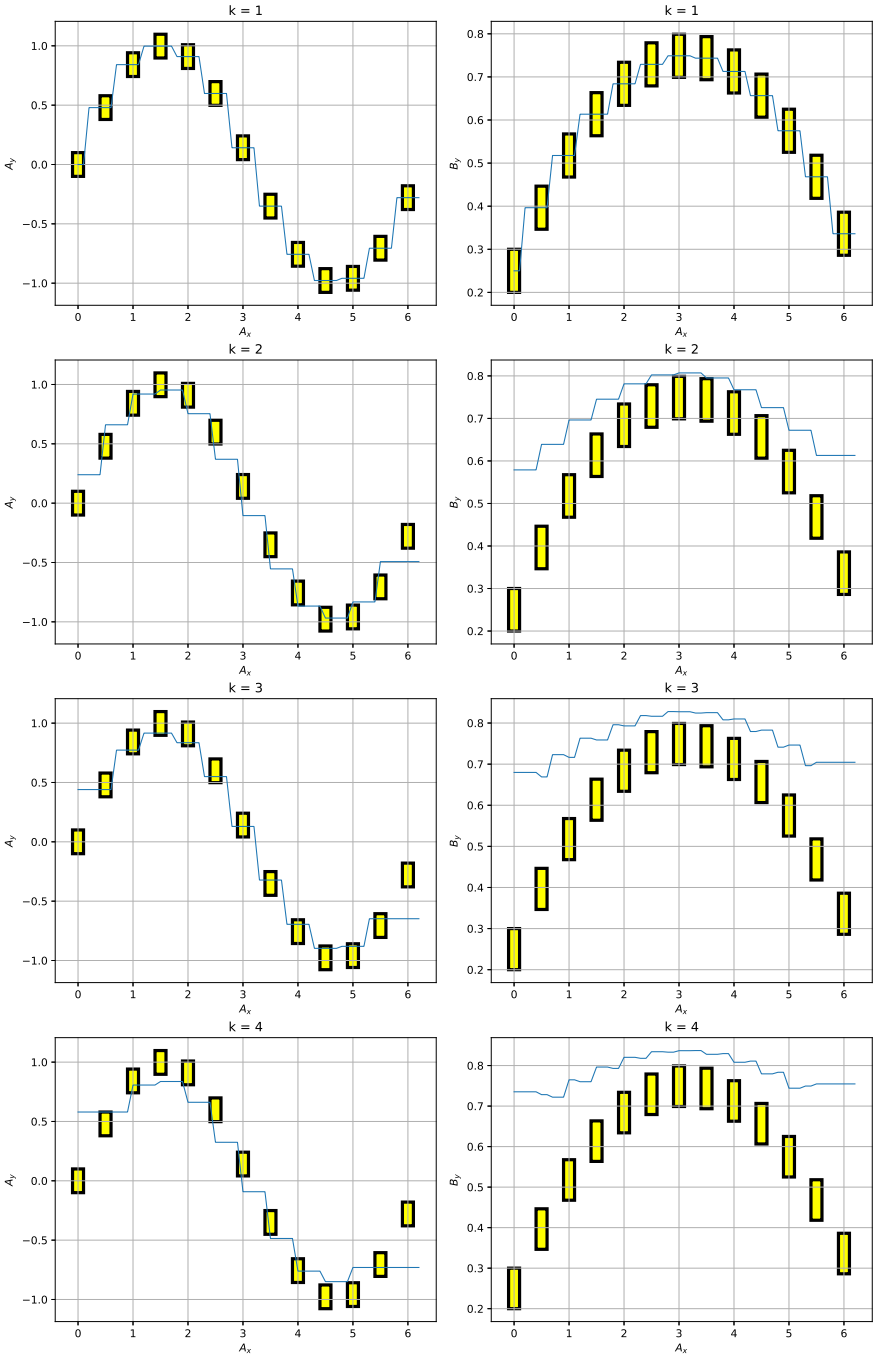


Rysunek 5.10. Dane wykorzystane w badaniach nr 1: (a) wartości danych, (b) prawdopodobieństwa wystąpienia wartości wejściowych, (c) prawdopodobieństwa wystąpienia wartości wyjściowych

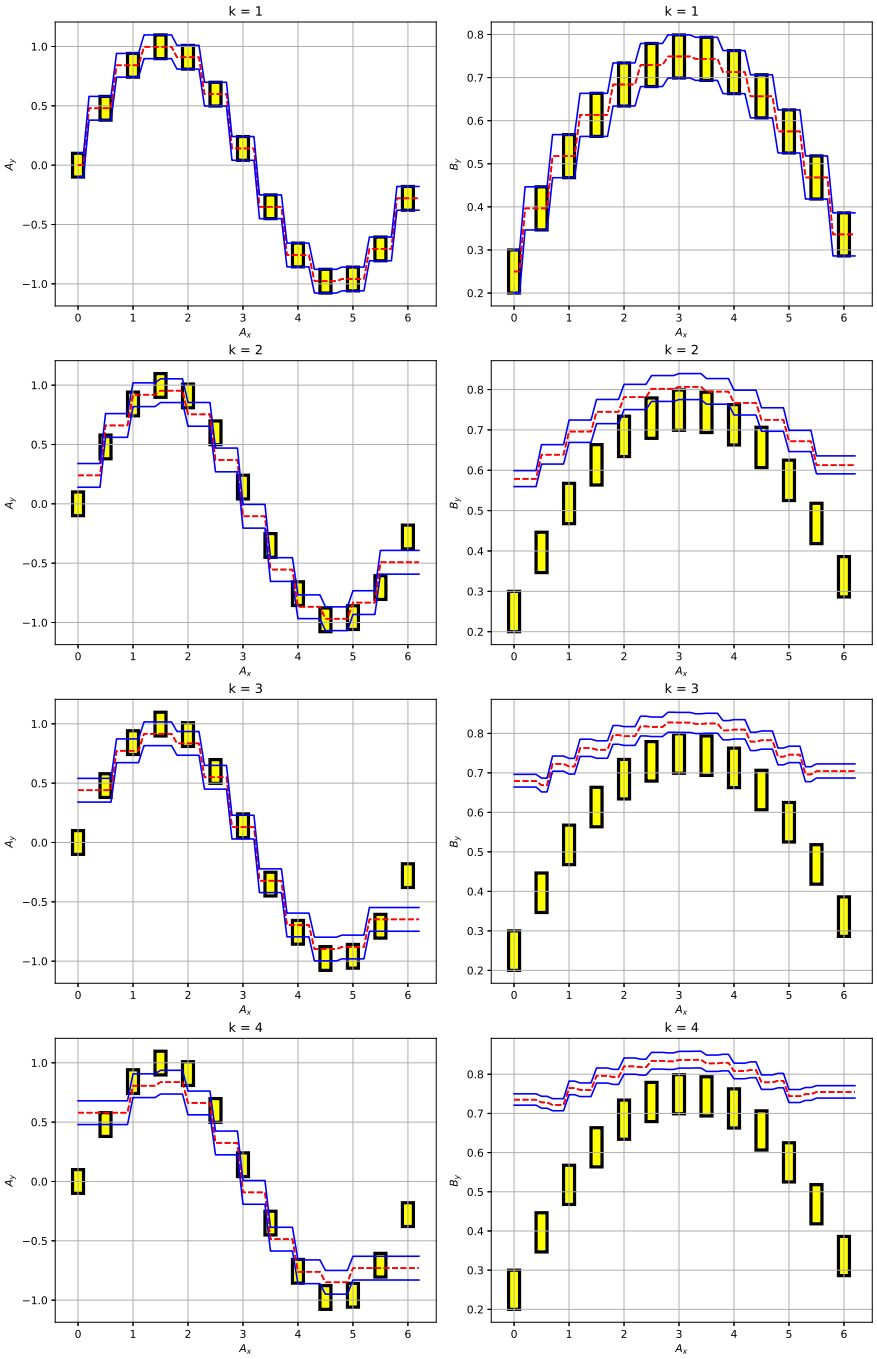


Rysunek 5.11. Bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla: prostej metody kNN (a) i dla modelu bazującego na mini-modelach nieliniowych (b)

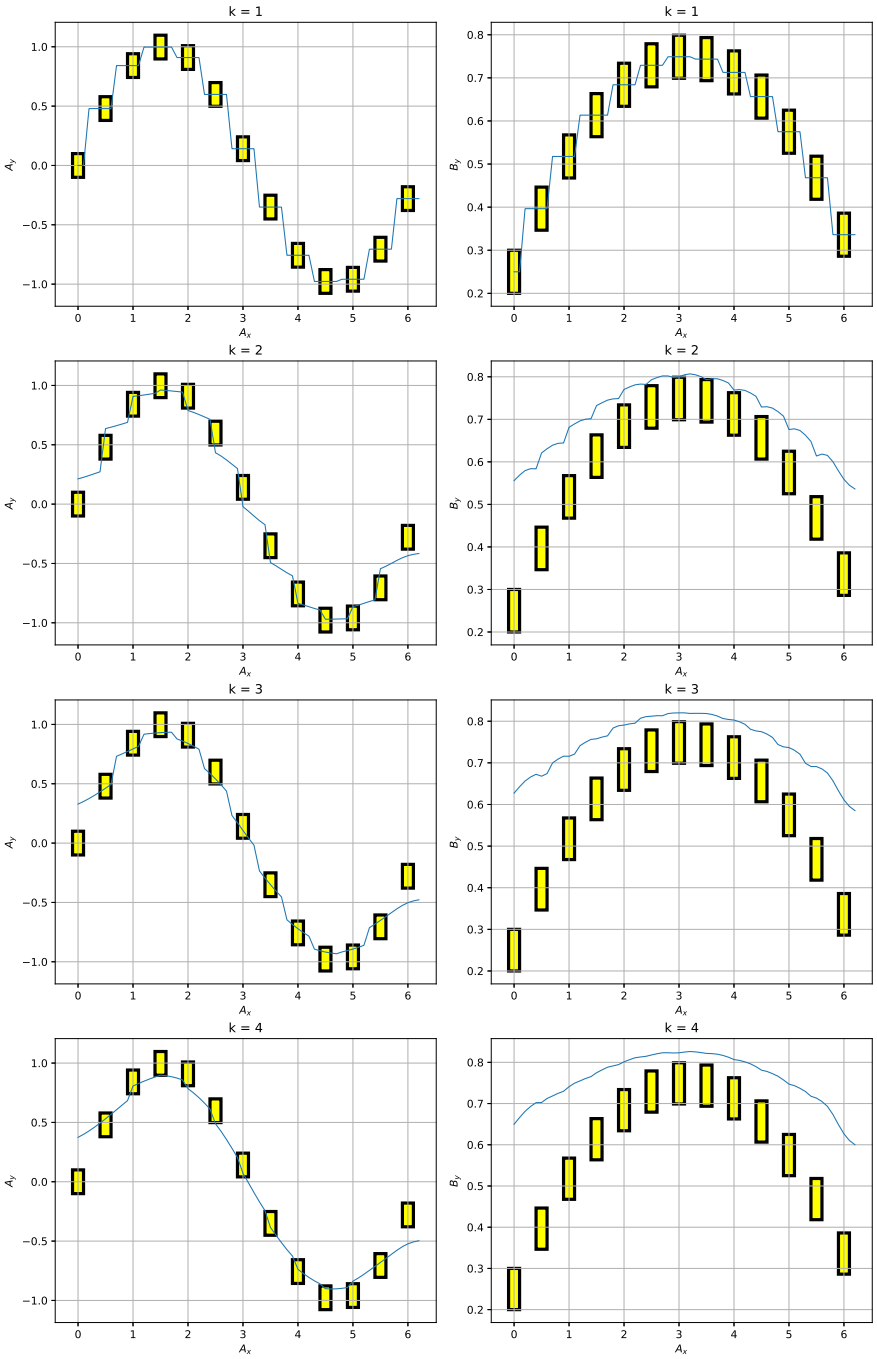
niu regresji). Na wykresach po lewej stronie widoczna jest charakterystyka dla wartości, a po prawej stronie dla prawdopodobieństwa wynikowej liczby Z. Na rysunkach 5.12 – 5.15, dla $k > 1$, charakterystyki wyznaczane dla prawdopodobieństw położone są powyżej prawdopodobieństw danych wejściowych. Wynika to z własności arytmetyki liczb Z [1], zgodnie z którą uśrednione prawdopodobieństwo wynikowe operacji jest zawsze wyższe od średniego prawdopodobieństwa operandów.



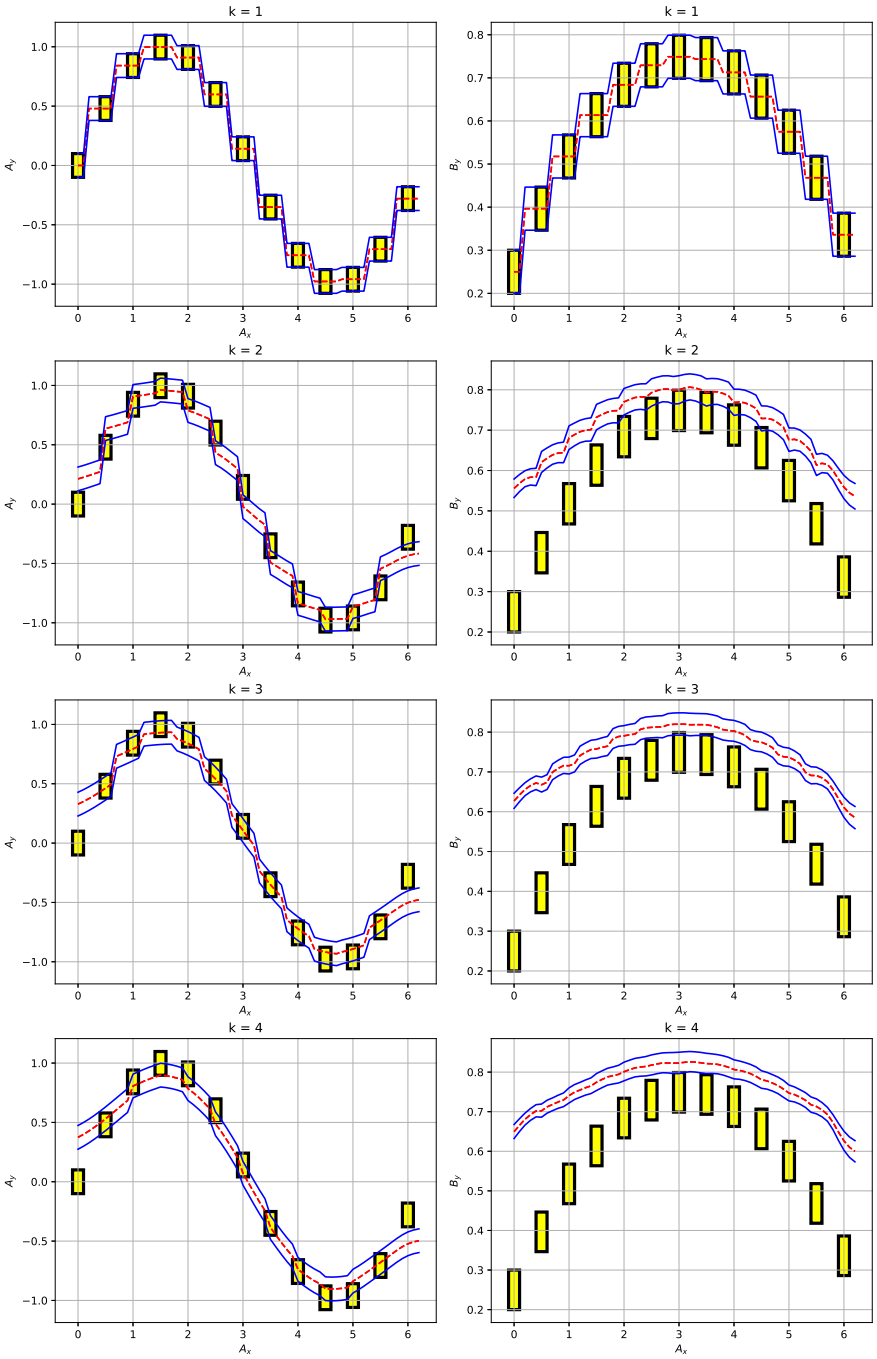
Rysunek 5.12. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



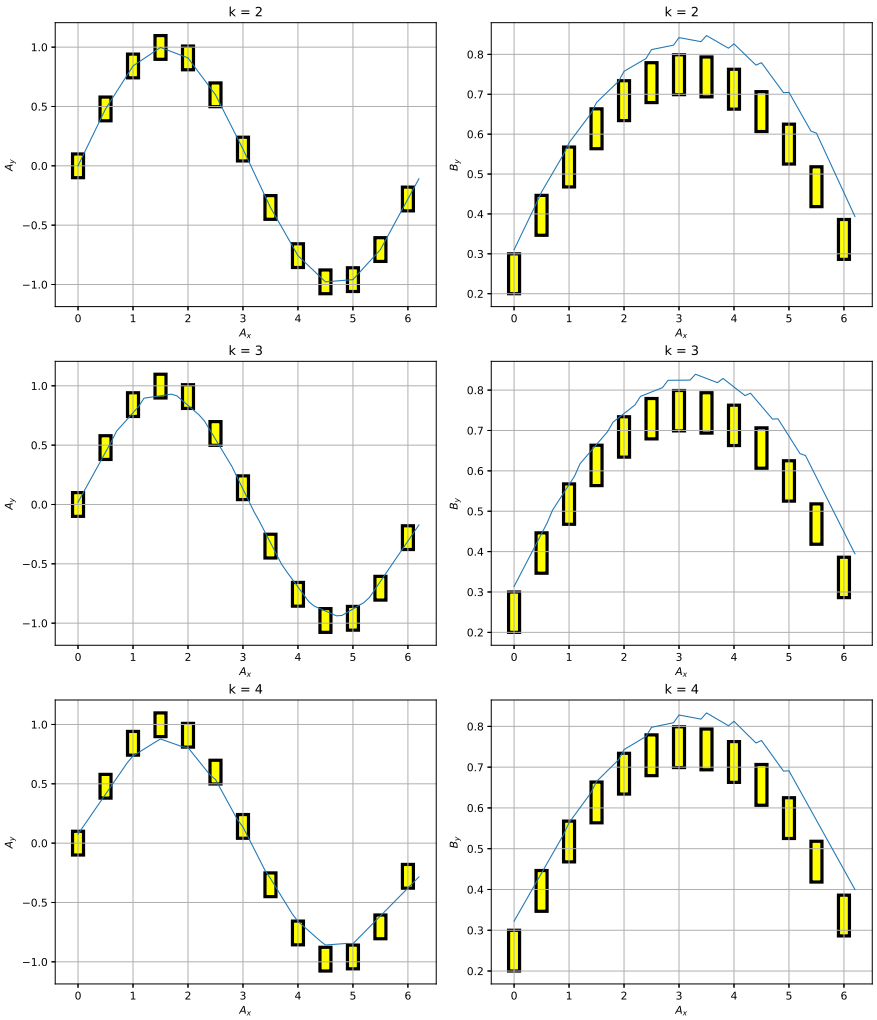
Rysunek 5.13. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



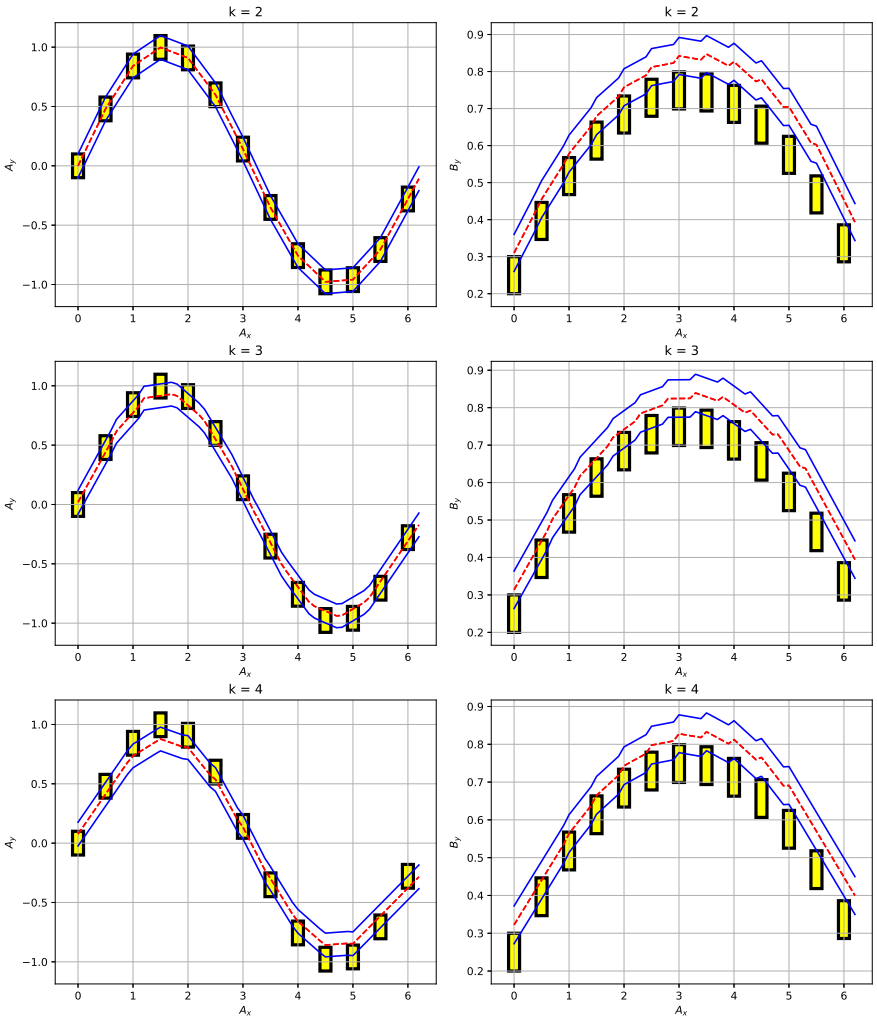
Rysunek 5.14. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



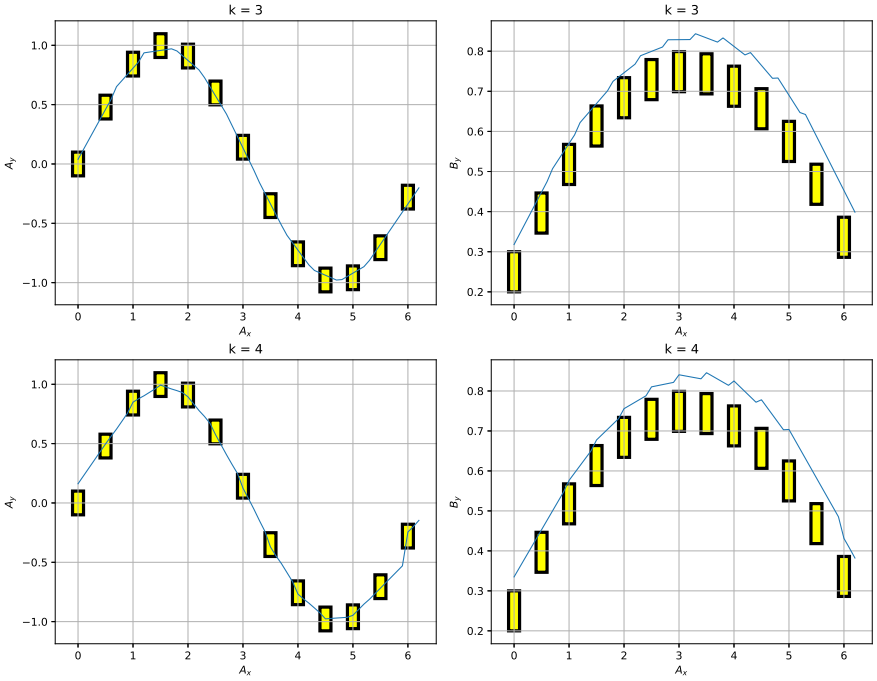
Rysunek 5.15. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



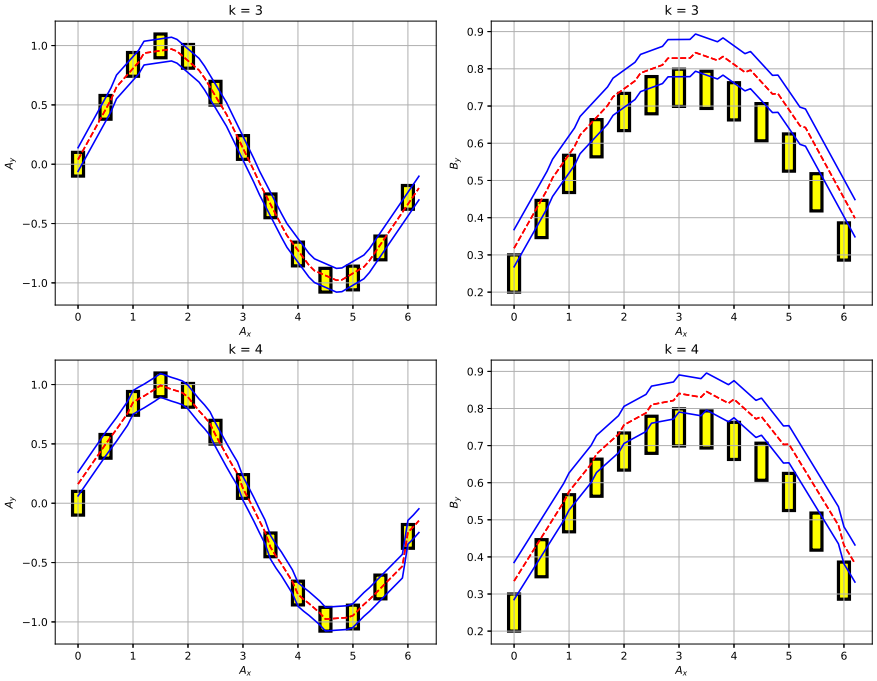
Rysunek 5.16. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



Rysunek 5.17. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



Rysunek 5.18. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



Rysunek 5.19. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia

W tabeli 5.1 podany jest średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzoru (5.40) przy założeniu $k = 2$ i $k = 3$.

Tabela 5.1. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 1

model	e_{MAE} dla $k = 2$	e_{MAE} dla $k = 3$
kNN – średnia arytmetyczna	0.224	0.366
kNN – średnia ważona	0.209	0.252
mini-model liniowy	0.096	0.149
mini-model nieliniowy	0.096	0.095

Badania nr 2 – mieszane dane jednowejściowe

W kolejnym eksperymencie zbadano w jaki sposób działają modele utworzone na podstawie mieszanych danych jednowejściowych. Dane uczące zostały przygotowane tak, aby zapewnić jak największą różnorodność, stąd zarówno wartości atrybutów wejściowych i wyjściowych, jak i ich prawdopodobieństwa przyjmowały postać liczb rozmytych, ale także interwałów i liczb rzeczywistych. Dane przedstawiono w tabelach 5.2 i 5.3. W poglądowy sposób zaprezentowano je także na rys. 5.20 i 5.21.

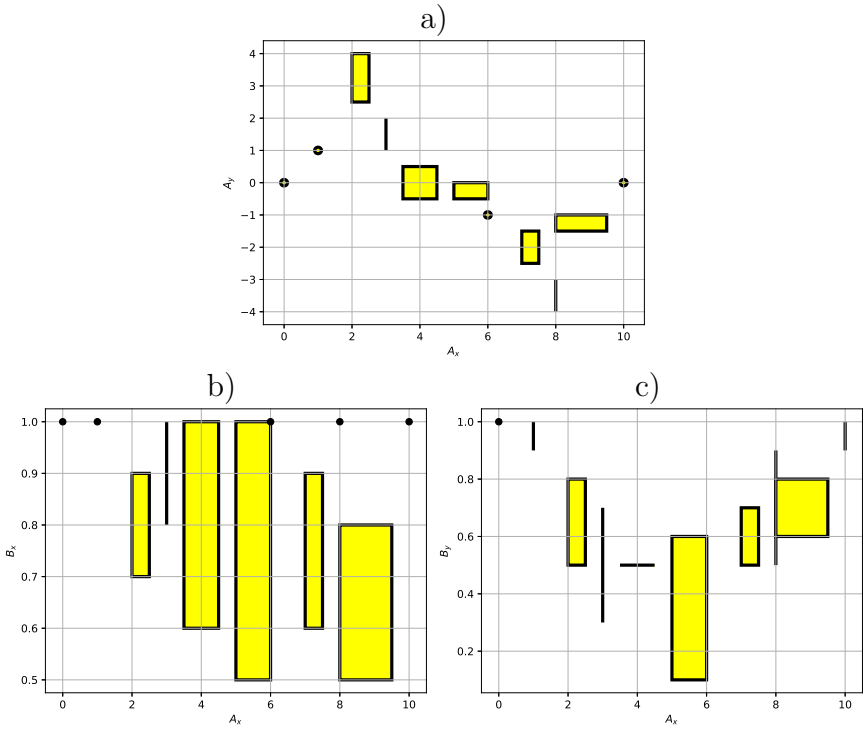
Tabela 5.2. Dane uczące wykorzystane w badaniach nr 2 – wartości

wejście A_x	wyjście A_y
0	0
1	1
$[2, 2.5]$	TFN(2.5,4,4)
3	TFN$(1, 1.5, 2)$
TFN(3.5,4,4.5)	TFN$(-0.5, 0, 0.5)$
TFN$(5, 5.5, 6)$	TFN$(-0.5, -0.5, 0)$
6	-1
$[7, 7.5]$	TFN$(-2.5, -2, -1.5)$
8	TrFN$(-4, -4, -3.5, -3)$
TrFN$(8, 8.5, 9, 9.5)$	TFN$(-1.5, -1, -1)$
10	0

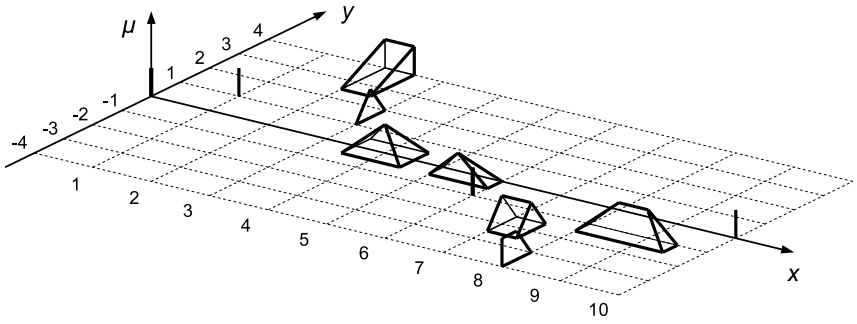
Na rys. 5.22 pokazano jak zmienia się bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla prostej metody kNN i dla modelu bazującego na mini-modelach nieliniowych. Wynika z niego, że optymalna liczba dla metody kNN wynosi $k = 2$ (podobnie jak dla ważonej metody kNN i modelu bazującego

Tabela 5.3. Dane uczące wykorzystane w badaniach nr 2 – prawdopodobieństwa

wejście B_x	wyjście B_y
1	1
1	$[0.9, 1]$
$[0.7, 0.9]$	$[0.5, 0.8]$
TFN(0.8,1,1)	TFN$(0.3, 0.5, 0.7)$
TFN(0.6,0.8,1)	0.5
$[0.5, 1]$	TFN$(0.1, 0.3, 0.6)$
1	$[0.4, 0.6]$
TrFN(0.6,0.7,0.8,0.9)	TFN$(0.5, 0.6, 0.7)$
1	TrFN$(0.5, 0.7, 0.8, 0.9)$
$[0.5, 0.8]$	$[0.6, 0.8]$
1	$[0.9, 1]$

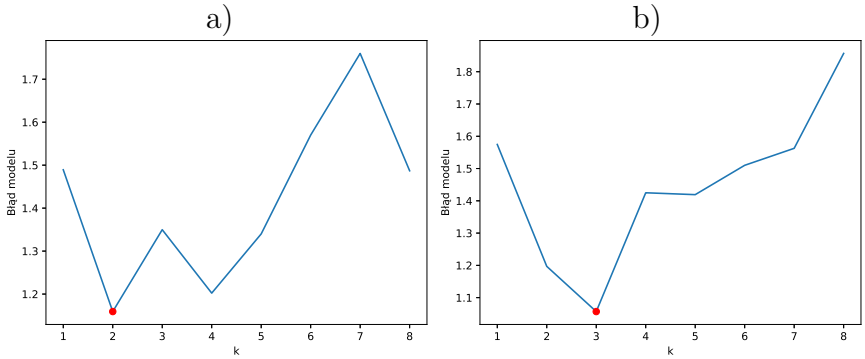


Rysunek 5.20. Dane wykorzystane w badaniach nr 2: (a) wartości danych, (b) prawdopodobieństwa wystąpienia wartości wejściowych, (c) prawdopodobieństwa wystąpienia wartości wyjściowych



Rysunek 5.21. Poglądowy widok wartości danych wykorzystanych w badaniach nr 2

na mini-modelach liniowych), a dla modelu bazującego na mini-modelach nieliniowych wartość ta wynosi $k = 3$.

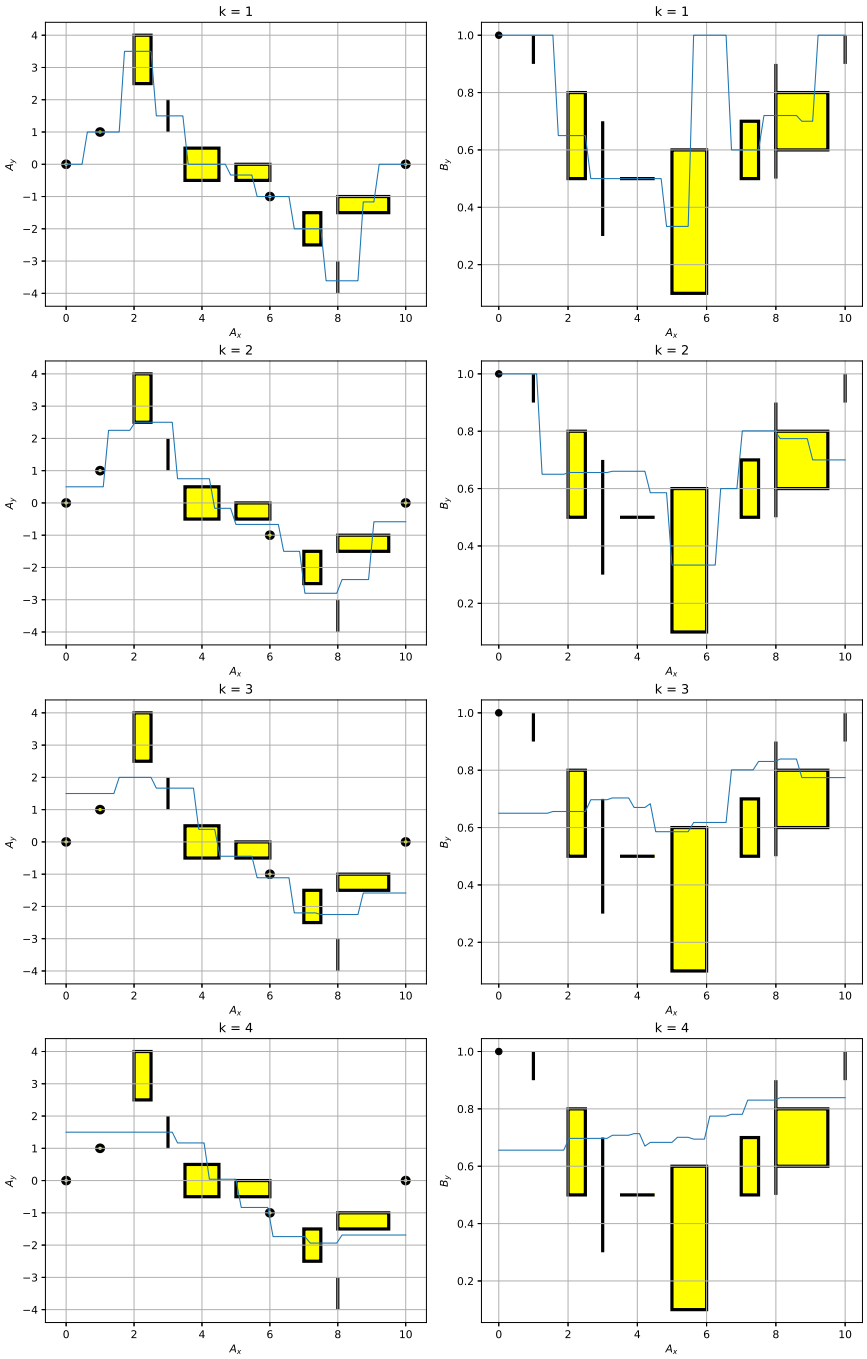


Rysunek 5.22. Bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla: prostej metody kNN (a) i dla modelu bazującego na mini-modelach nieliniowych (b)

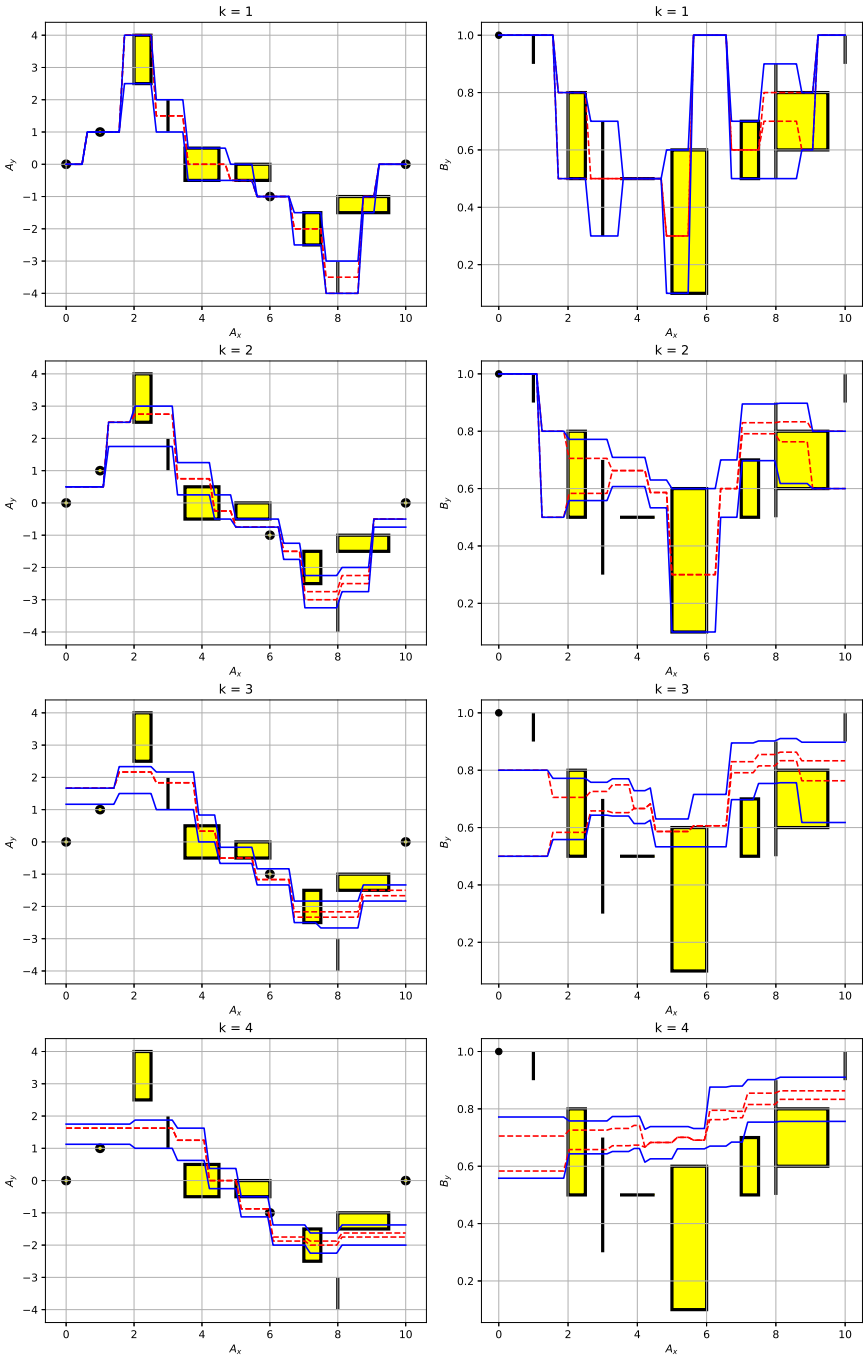
Na rysunkach 5.23 – 5.30 pokazano charakterystyki modeli bazujące na prostej i ważonej metodzie kNN oraz na mini-modelach liniowych i nieliniowych. Charakterystyki sporządzone zostały dla różnych ilości najbliższych sąsiadów k, przy czym dla metod kNN parametr k przyjmował wartości od 1 do 4, a dla metod bazujących na mini-modelach wartości od 2 do 4 (ze względu na minimalną liczbę próbek wymaganą przy wyznaczaniu regresji). Na wykresach po lewej stronie widoczna jest charakterystyka dla wartości, a po prawej stronie dla prawdopodobieństwa wynikowej liczby Z.

W tabeli 5.4 podany jest średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzoru (5.40) przy założeniu $k = 2$ i $k = 3$.

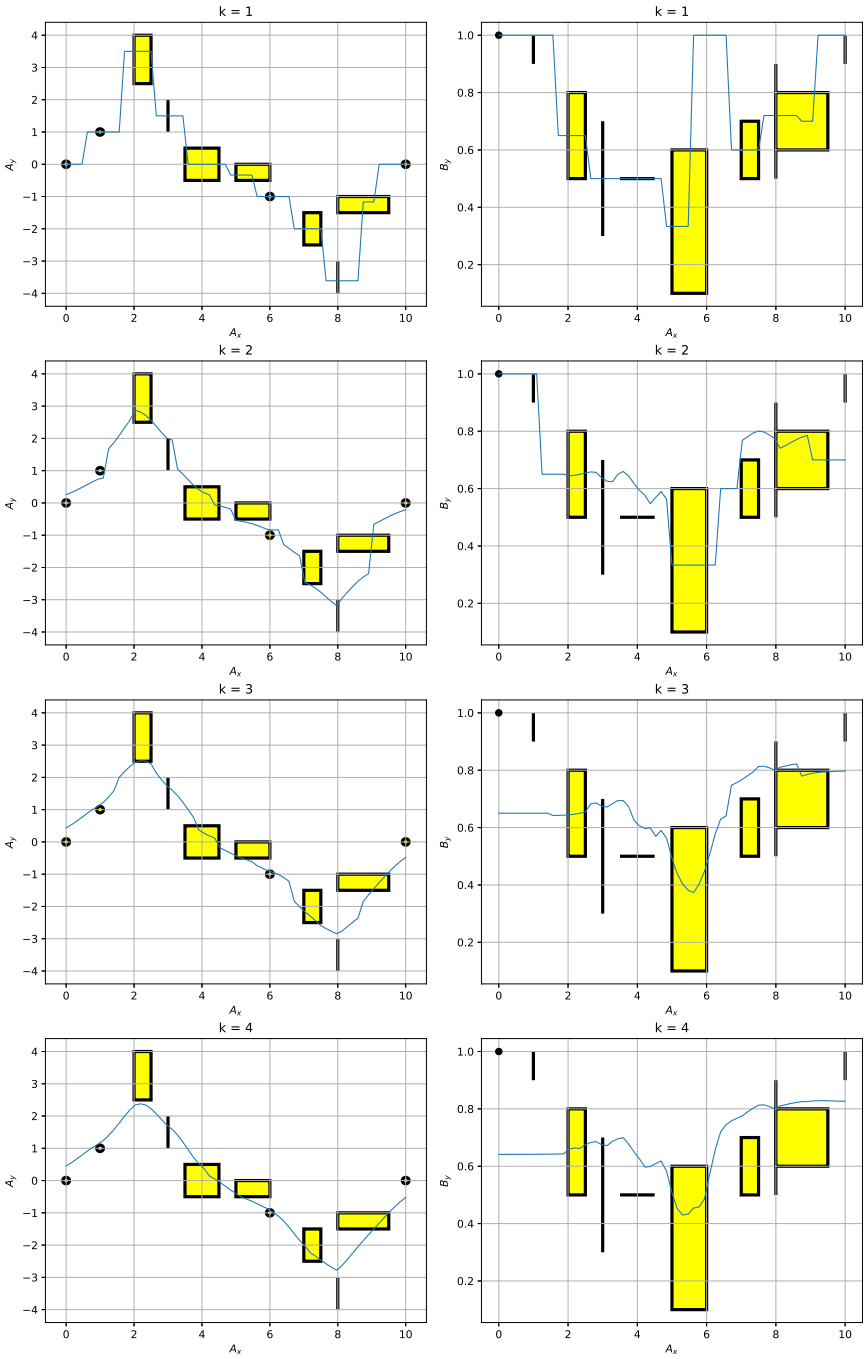
Na rys. 5.31 pokazano odpowiedź modelu utworzonego za pomocą ważonej metody kNN dla interwałowego wejścia $x^* = [4.5, 5]$ i $k = 2$. Możemy zauważyć, że funkcja przynależności liczby rozmytej opisującej prawdopodobieństwo nie jest odcinkowo-liniowa. Z kolei na rys. 5.32 widzimy odpowiedź modelu utworzonego za pomocą liniowych mini-modeli dla podobnego wejścia. Funkcja przynależności liczby rozmytej opisującej prawdopodobieństwo jest tu trójkątna i symetryczna.



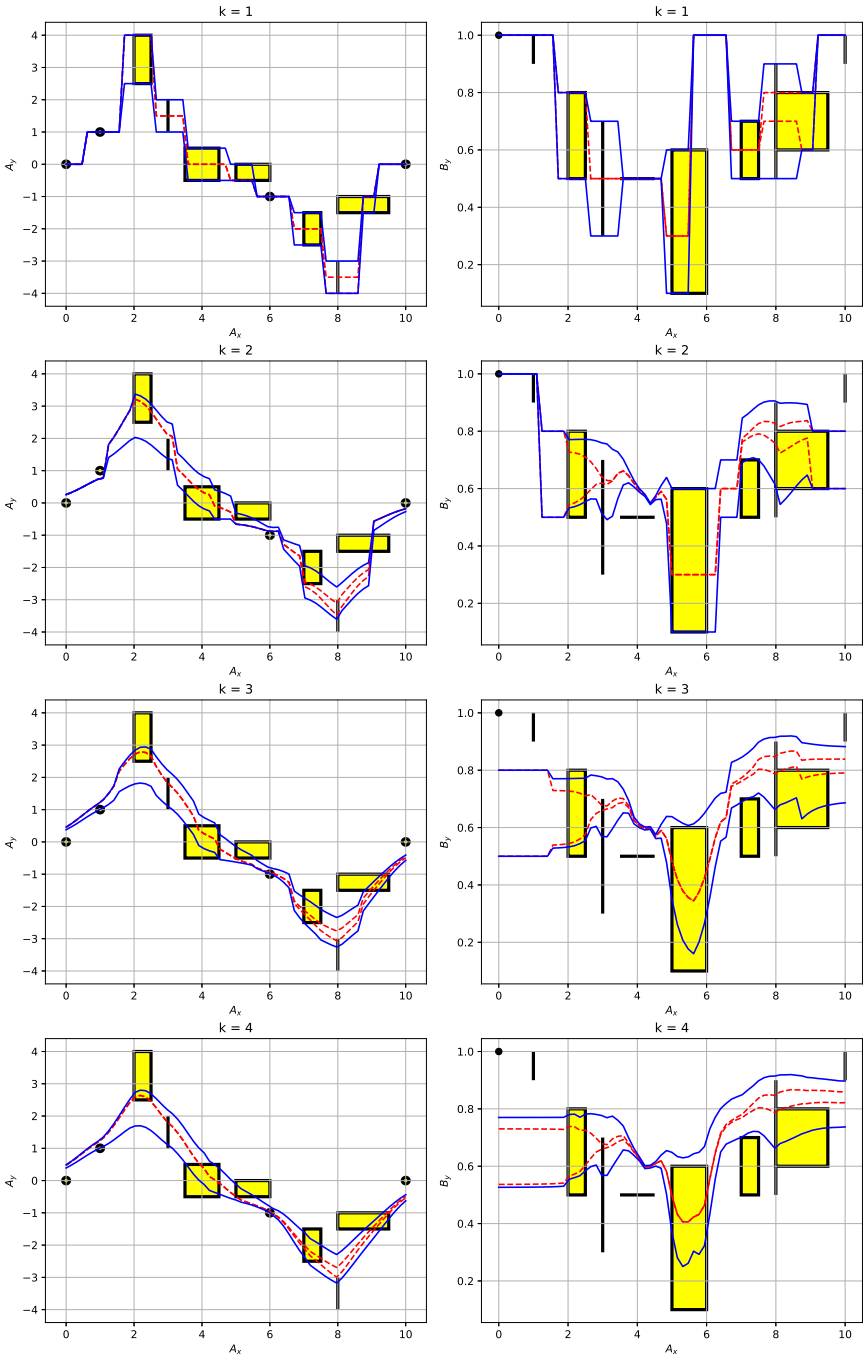
Rysunek 5.23. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



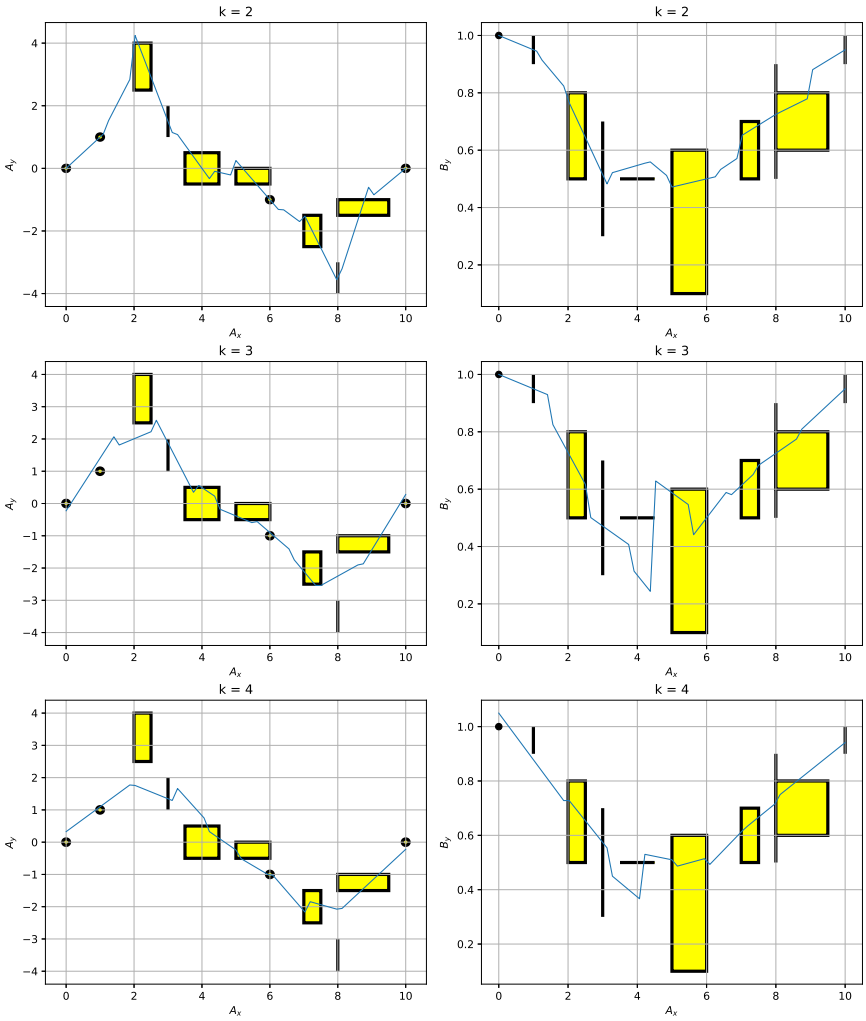
Rysunek 5.24. Charakterystyki modeli utworzonych z wykorzystaniem prostej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linie przerywane szerokość jej rdzenia



Rysunek 5.25. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



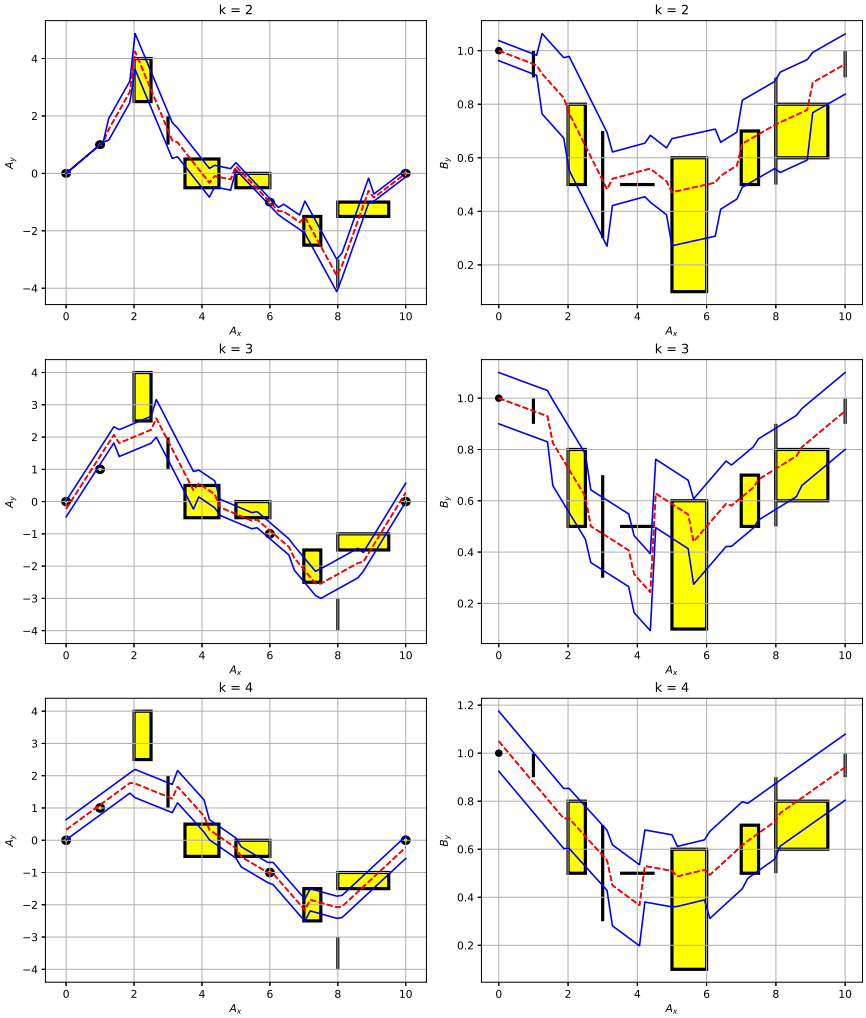
Rysunek 5.26. Charakterystyki modeli utworzonych z wykorzystaniem ważonej metody kNN dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linie przerywane szerokość jej rdzenia



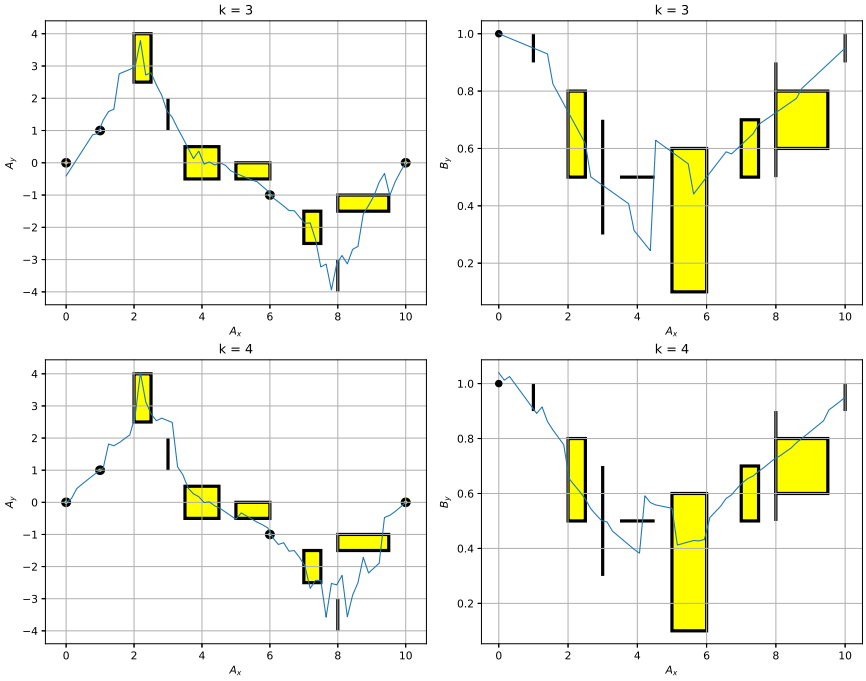
Rysunek 5.27. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej

Tabela 5.4. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 2

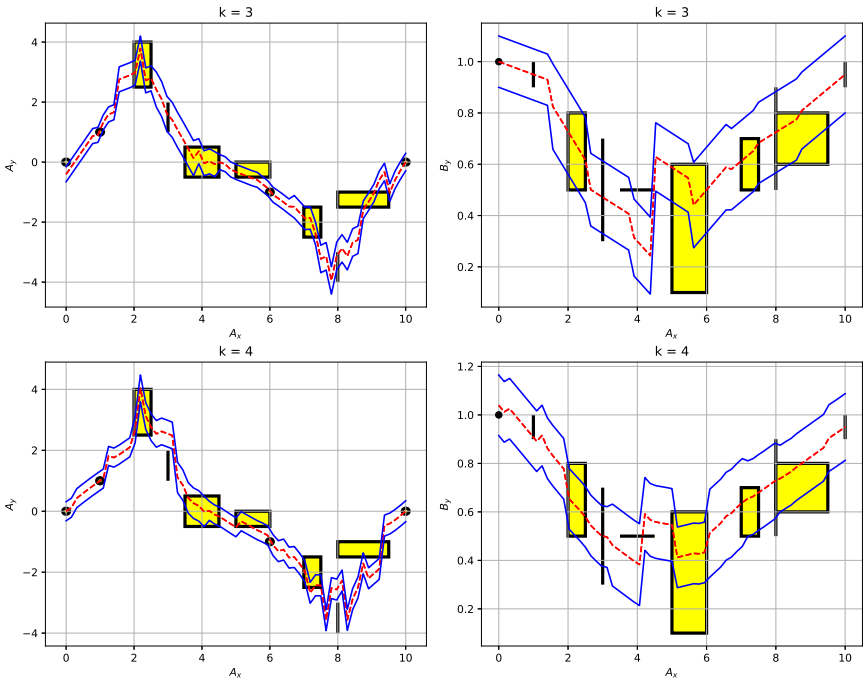
model	e_{MAE} dla $k = 2$	e_{MAE} dla $k = 3$
kNN – średnia arytmetyczna	1.159	1.350
kNN – średnia ważona	1.183	1.202
mini-model liniowy	1.196	1.201
mini-model nieliniowy	1.196	1.056



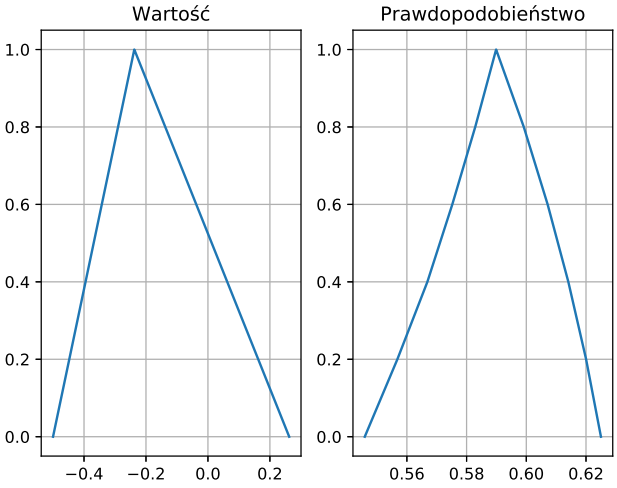
Rysunek 5.28. Charakterystyki modeli utworzonych z wykorzystaniem liniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenia jej rdzenia



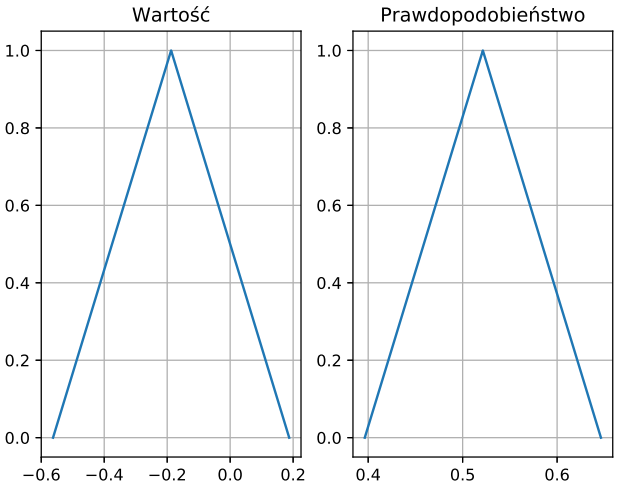
Rysunek 5.29. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linia przedstawia środek ciężkości wynikowej liczby rozmytej



Rysunek 5.30. Charakterystyki modeli utworzonych z wykorzystaniem nieliniowych mini-modeli dla różnych ilości najbliższych sąsiadów k. Linie ciągłe przedstawiają szerokość nośnika wynikowej liczby rozmytej, a linia przerywana położenie jej rdzenia



Rysunek 5.31. Odpowiedź modelu utworzonego za pomocą ważonej metody kNN dla interwałowego wejścia $x^* = [4.5, 5]$ i $k = 2$



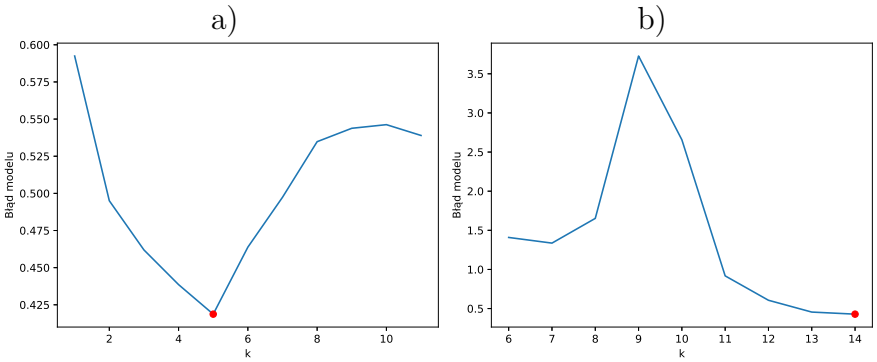
Rysunek 5.32. Odpowiedź modelu utworzonego za pomocą liniowych mini-modeli dla interwałowego wejścia $x^* = [4.5, 5]$ i $k = 2$

Badania nr 3 – mieszane dane 4-wejściowe

Podobnie jak w przypadku liczb rozmytych, eksperyment ostatni będzie miał na celu sprawdzenie, w jaki sposób metody poradzą sobie z mieszanymi danymi wielowejściowymi. Ponownie zajmiemy się przykładem wyceny wartości nieruchomości, jednak dla wartości posiadanych danych określimy jeszcze prawdopodobieństwa ich wystąpienia w postaci liczb rozmytych.

Załóżmy, że dla 15 obiektów posiadamy informacje o następujących cechach: wiek (w latach), standard wykończenia (w skali od 0 do 10 punktów), powierzchnia (w m^2), odległość od centrum (w kilometrach) i cenę jednego metra kwadratowego (w tys. zł za m^2). Część informacji znamy dokładnie, a część w przybliżeniu: w formie interwału lub formie pojęcia lingwistycznego, reprezentowanego przez liczbę rozmytą. Dodatkowo, znamy prawdopodobieństwa wystąpienia poszczególnych wartości: dokładnie albo nieprecyzyjnie, w postaci lingwistycznej (opisanej liczbą rozmytą). Na podstawie posiadanej wiedzy, będziemy próbowali wyceniać kolejne nieruchomości. Dane przedstawiono w tabelach 5.5 i 5.6. Funkcje przynależności pojęć wykorzystanych w tabeli 5.6 pokazano na rys. 5.34. W kolejnych eksperymentach, podobnie jak to miało miejsce wcześniej, wartości atrybutów wejściowych danych uczących były normalizowane z powodu znacznych różnic w szerokości ich dziedzin.

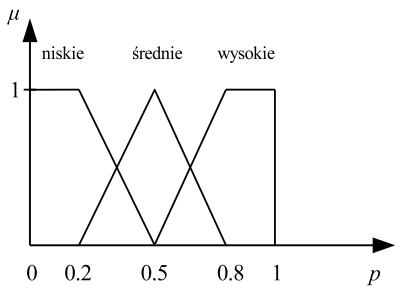
Na rys. 5.33 pokazano jak zmienia się bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla prostej metody kNN (podobny wykres otrzymano dla ważonej metody kNN) oraz dla modelu bazującego na mini-modelach nieliniowych (podobny wykres otrzymano dla modelu bazującego na mini-modelach liniowych). Wynika z niego, że optymalna liczba dla metod kNN wynosi $k = 5$, a dla modelu bazującego na mini-modelach wartość ta wynosi $k = 14$. Ponieważ w danych jest tylko 15 próbek, nie można było zbadać zachowania modelu dla większych wartości k.



Rysunek 5.33. Bezwzględny błąd kroswalidacji modelu w zależności od przyjętej liczby najbliższych sąsiadów k dla: prostej metody kNN (a) i dla modelu bazującego na mini-modelach nieliniowych (b)

Tabela 5.5. Dane wykorzystane w badaniach nr 3 – wartości

Lp.	Wiek [lata]	Standard	Powierzchnia [m^2]	Odległość [km]	Cena [tys. zł/m^2]
1	10	6	90	4	4.6
2	'około 20' TFN(15,20,25)	7	80	[5,6]	'około 4.6' TFN(4.4,4.6,4.8)
3	[50,60]	4	70	8	3.9
4	[70,100]	2	50	3	'około 3.5' TFN(3.4,3.5,3.6)
5	5	'wysoki' TFN(8,10,10)	40	3	4.9
6	12	4	30	'około 7' TFN(6,7,8)	'około 4.6' TFN(4.5,4.6,4.7)
7	20	7	65	6	'około 4.8' TFN(4.7,4.8,4.9)
8	[0,1]	10	75	0.0	5.5
9	'stary' TFN(80,100,100)	0	30	'duża' TFN(12,15,15)	'około 3.2' TFN(3.0,3.2,3.4)
10	[25,35]	2	55	1	4.6
11	[50,70]	[8,10]	100	4	4.2
12	25	'średni' TFN(4,5,6)	110	'około 7' TFN(6,7,8)	3.9
13	10	'wysoki' TFN(8,10,10)	85	11	'około 4.5' TFN(4.4,4.5,4.6)
14	5	3	45	5	4.7
15	[45,55]	6	70	'około 8' TFN(7,8,9)	'około 4.1' TFN(4,4.1,4.2)



Rysunek 5.34. Funkcje przynależności pojęć wykorzystanych w tabeli 5.6

Tabela 5.6. Dane wykorzystane w badaniach nr 3 – prawdopodobieństwa wystąpienia

Lp.	Wiek [lata]	Standard	Powierzchnia [m^2]	Odległość [km]	Cena [tys. zł/m^2]
1	1	średnie	1	wysokie	wysokie
2	wysokie	wysokie	wysokie	średnie	wysokie
3	średnie	niskie	wysokie	wysokie	wysokie
4	średnie	średnie	1	wysokie	średnie
5	1	1	1	wysokie	wysokie
6	1	średnie	wysokie	średnie	niskie
7	wysokie	niskie	średnie	1	średnie
8	1	wysokie	1	wysokie	wysokie
9	średnie	niskie	średnie	wysokie	średnie
10	wysokie	średnie	wysokie	średnie	niskie
11	wysokie	wysokie	1	wysokie	średnie
12	średnie	średnie	wysokie	średnie	wysokie
13	wysokie	średnie	wysokie	średnie	średnie
14	1	średnie	1	wysokie	wysokie
15	średnie	wysokie	wysokie	średnie	niskie

W tabeli 5.7 podany jest średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej za pomocą wzoru (5.40) przy założeniu $k = 5$ i $k = 14$.

Tabela 5.7. Średni błąd bezwzględny wyznaczony metodą walidacji krzyżowej w badaniach nr 3

model	e_{MAE} dla $k = 5$	e_{MAE} dla $k = 14$
kNN – średnia arytmetyczna	0.419	0.626
kNN – średnia ważona	0.411	0.675
mini-model liniowy	0.465	0.415
mini-model nieliniowy	0.465	0.428

W tabeli 5.8 pokazano środki ciężkości wartości i prawdopodobieństwa odpowiedzi modeli dla przykładowego wejścia rzeczywistego $\mathbf{x}^* = [50, 6, 70, 4]^T$. Odpowiedzi wyznaczone zostały dla optymalnej liczby najbliższych sąsiadów wyznaczonej dla poszczególnych modeli. Dodatkowo na rys. 5.35 pokazane zostały kształty funkcji przynależności odpowiedzi.

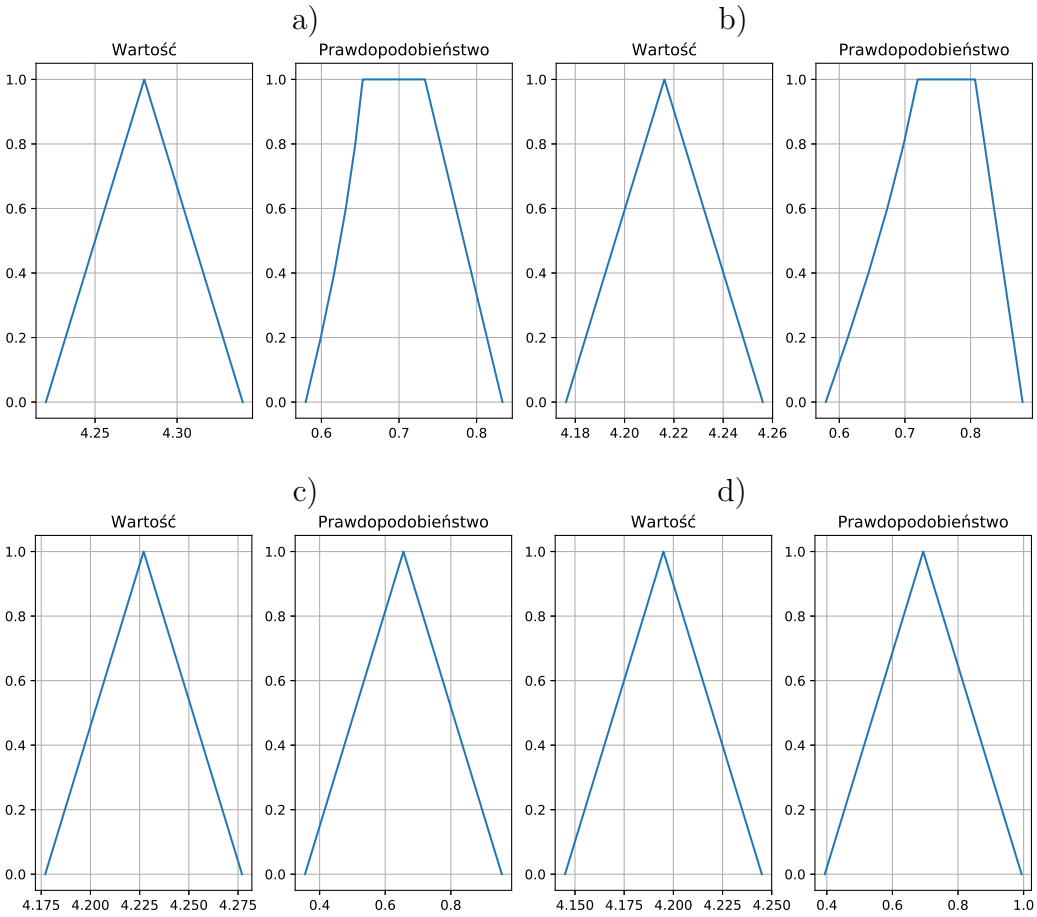
Tabela 5.8. Wartości odpowiedzi modeli dla przykładowego wejścia $\mathbf{x}^* = [50, 6, 70, 4]^T$

model	wartość wyjścia	prawdopodobieństwo wyjścia
kNN – średnia arytmetyczna ($k = 5$)	4.280	0.703
kNN – średnia ważona ($k = 5$)	4.216	0.746
mini-model liniowy ($k = 14$)	4.227	0.655
mini-model nieliniowy ($k = 14$)	4.195	0.694

W tabeli 5.9 pokazano środki ciężkości wartości i prawdopodobieństwa odpowiedzi modeli dla przykładowego wejścia rozmytego $\mathbf{x}^* = [\mathrm{RFN}(20, 30),\ \mathrm{TFN}(2.5, 5, 7.5),\ 45,\ \mathrm{RFN}(0, 2)]^T$. Odpowiedzi wyznaczone zostały dla optymalnej liczby najbliższych sąsiadów wyznaczonej dla poszczególnych modeli. Dodatkowo na rys. 5.36 pokazane zostały kształty funkcji przynależności wybranych odpowiedzi.

5.7. Wnioski

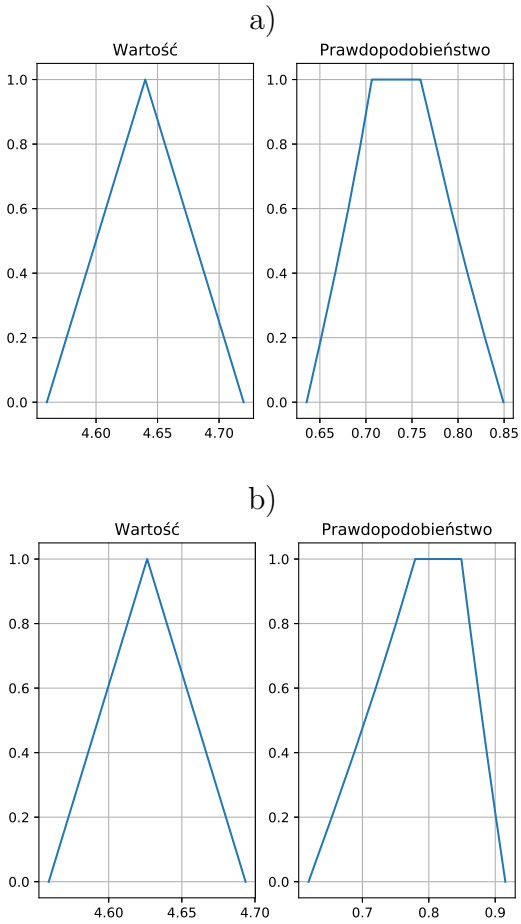
Jako podsumowanie badań można wskazać, że metody modelowania oparte na modelach lokalnych, dobrze nadają się do pracy na danych niepewnych. Dane mogą być jednorodne bądź mieszanego typu, przy czym zarówno dane uczące, jak i dane wejściowe modelu mogą mieć postać liczb Z, bądź liczb rozmytych, interwałów lub liczb rzeczywistych. W każdej sytuacji opisanej w badaniach, modele zwracały poprawne i wiarygodne



Rysunek 5.35. Funkcje przynależności odpowiedzi modeli dla przykładowego wejścia rzeczywistego $\mathbf{x}^* = [50, 6, 70, 4]^T$: (a) prosta metoda kNN, (b) ważona metoda kNN, (c) model bazujący na mini-modelach liniowych, (d) model bazujący na mini-modelach nieliniowych

Tabela 5.9. Wartości odpowiedzi modeli dla przykładowego wejścia $\mathbf{x}^* = [\mathrm{RFN}(20, 30), \mathrm{TFN}(2.5, 5, 7.5), 45, \mathrm{RFN}(0, 2)]^T$

model	wartość wyjścia	prawdopodobieństwo wyjścia
kNN – średnia arytmetyczna ($k = 5$)	4.640	0.739
kNN – średnia ważona ($k = 5$)	4.626	0.787
mini-model liniowy ($k = 14$)	4.771	0.824
mini-model nieliniowy ($k = 14$)	4.753	0.858



Rysunek 5.36. Funkcje przynależności odpowiedzi modeli dla przykładowego wejścia rozmytego $\mathbf{x}^* = [\text{RFN}(20, 30),\ \text{TFN}(2.5, 5, 7.5),\ 45, \text{RFN}(0, 2)]^T$: (a) prosta metoda kNN, (b) ważona metoda kNN

wyniki. Dokładność modeli opartych na mini-modelach była tu, podobnie jak wcześniej, porównywalna bądź nieznacznie lepsza od dokładności modeli opartych o techniki kNN.

W przypadku pracy na liczbach Z, na uwagę zasługuje czas wykonywanych obliczeń. Arytmetyka liczb Z jest złożona obliczeniowo. Najwięcej czasu w wykonywanych obliczeniach przypada na wyznaczanie wielu rozkładów prawdopodobieństwa (wzory (5.9) i (5.10) s. 100) poprzez rozwiązanie zadania programowania liniowego oraz wykonanie na nich działań arytmetycznych (wzory (5.11) – (5.13)). Skutek tych dość złożonych operacji jest taki, że modele działające na liczbach Z pracują kilkadziesiąt razy wolniej od modeli pracujących na liczbach rzeczywistych. Spowolnienie działania jest tu mniejsze dla modeli bazujących na mini-modelach, bowiem w tym przypadku duża część obliczeń (np. wyznaczanie regresji liniowej) wykonywana jest na środkach liczb rozmytych, czyli na liczbach rzeczywistych.

6. Zakończenie

W niniejszej monografii przedstawiono metodę modelowania opartą na lokalnych modelach regresyjnych (mini-modelach) i pokazano w jaki sposób zastosować ją do przetwarzania informacji różnego typu. Oryginalnie, metoda ta opracowana była dla liczb rzeczywistych, jednak można ją także zaadaptować dla danych niepewnych, takich jak np. interwały, liczby rozmyte i liczby Z. Dzięki temu modele mogą być tworzone na podstawie informacji, która nie jest precyzyjna, a ponadto mogą taką informację przetwarzać.

Jedną z największych zalet modeli bazujących na lokalnej regresji jest ich elastyczność w odniesieniu do typu przetwarzanej informacji. Dane mogą być jednorodne bądź mieszanego typu, przy czym zarówno dane uczące, jak i dane wejściowe modelu mogą mieć postać liczb Z, bądź liczb rozmytych, interwałów lub liczb rzeczywistych. W każdej sytuacji opisanej w badaniach, modele zwracały poprawne i wiarygodne wyniki. Dokładność modeli opartych na mini-modelach była zwykle porównywalna bądź nieznacznie lepsza od dokładności innych modeli pamięciowych (np. opartych o techniki kNN).

Modelowanie bazujące na lokalnej regresji może być bardzo atrakcyjną alternatywą dla klasycznych systemów rozmytych. Najważniejsze zalety takiego modelowania, wykorzystanego do wyznaczania odpowiedzi systemu rozmytego wymienione są poniżej.

- Po pierwsze, baza reguł nie musi być kompletna. Konkluzje niektórych reguł w tabeli mogą być trudne do określenia, jednak nawet w przypadku braku części reguł w modelu, metody bazujące na regresji lokalnej skutecznie potrafią wyznaczyć jego wyjście.
- Baza reguł nie musi być spójna. W przypadku reguł o sprzecznych konkluzjach, model uśredni je wyznaczając odpowiedź.
- Użytkownik może sam założyć ile reguł modelu wykorzystywanych jest przy wyznaczaniu wyjścia, przy czym łatwo może wyznaczyć ilość optymalną dla danej metody modelowania. W klasycznym modelowaniu rozmytym, liczba wykorzystywanych reguł jest stała i zależna od przyjętej metody wyznaczania wyjścia.

Adaptacja metod modelowania pamięciowego do nowych typów danych jest stosunkowo prosta, stąd najbardziej oczywistym kierunkiem dalszym prac będzie opracowanie metod bazujących na regresji lokalnej do innych typów danych niepewnych, takich jak np. liczby rozmyte typu II, liczby szare, liczby w formie zbioru możliwych wartości i inne.

Kolejnym kierunkiem dalszych badań, może być opracowanie metod wykorzystujących w regresji wielomiany wyższych rzędów lub regresję nieliniową.

Literatura

[1] Aliev, R., Huseynov, O., Aliyev, R., Alizadeh, A.: *The Arithmetic of Z-numbers: Theory and Applications*. World Scientific (2015)

[2] Aliev, R., Pedrycz, W., Fazlollahi, B., Huseynov, O., Alizadeh, A., Guirimov, B.: Fuzzy logic-based generalized decision theory with imperfect information. *Information Sciences* **189**, s. 18–42 (2012)

[3] Aliev, R., Pedrycz, W., Huseynov, O.: Functions defined on a set of Z-numbers. *Information Sciences* **423**, s. 353–375 (2018)

[4] Andersen, E., Andersen, K.: The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. W: *High performance optimization*, s. 197–232. Springer (2000)

[5] Andersen, E., Gondzio, J., Mészáros, C., Xu, X.: *Implementation of interior point methods for large scale linear programming*. HEC/Université de Geneve (1996)

[6] Arabpour, A., Tata, M.: Estimating the parameters of a fuzzy linear regression model. *Iranian Journal of Fuzzy Systems* **5**(2), s. 1–19 (2008)

[7] Atkeson, C., Moore, A., Schaal, S.: Locally weighted learning. *Artificial Intelligence Review* **11**, s. 11–73 (1997)

[8] Bargiela, A., Pedrycz, W., Nakashima, T.: Multiple regression with fuzzy data. *Fuzzy Sets and Systems* **158**, s. 2169–2188 (2007)

[9] Bertsimas, D., Tsitsiklis, J.: *Introduction to linear optimization*. Athena Scientific Belmont, MA (1997)

[10] Bierens, H.: The Nadaraya-Watson kernel regression function estimator. W: *Topics in Advanced Econometrics*, s. 212–247. Cambridge University Press, New York (1988)

[11] Billard, L., Diday, E.: Regression analysis for interval-valued data. W: *Data Analysis, Classification, and Related Methods*, s. 369–374. Springer (2000)

[12] Brown, R.: Building a balanced k-d tree in O(kn log n) time. *Journal of Computer Graphics Techniques* **4**(1), s. 50–68 (2015)

[13] Cichosz, P.: *Systemy uczące się*. WNT, Warszawa (2000)

[14] Cleveland, W.: Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association* **74**(368), s. 829–836 (1979)

[15] Cleveland, W.: LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *American Statistician* **35**(1), 54 (1981)

[16] Cleveland, W., Devlin, S.: Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association* **83**(403), s. 596–610 (1988)

[17] Dantzig, G.: *Linear programming and extensions.* Princeton University Press (1998)

[18] Diamond, P.: Fuzzy least squares. *Information Sciences* **46**, s. 141–157 (1988)

[19] Diamond, P., Kloeden, P.: *Metric spaces of fuzzy sets, theory and applications.* World Scientific, Singapore (1994)

[20] Diamond, P., Rosenfeld, A.: Metric spaces of fuzzy sets. *Fuzzy Sets and Systems* **35**, s. 241–249 (1990)

[21] Dinagar, D., Anbalagan, A.: A new type-2 fuzzy number arithmetic using extension principle. W: *Proceedings of the International Conference on Advances in Engineering, Science and Management (ICAESM)*, s. 113–118 (2012)

[22] Domingues, M., de Souza, R., Cysneiros, F.: A robust method for linear regression of symbolic interval data. *Pattern Recognition Letters* **31**(13), s. 1991–1996 (2010)

[23] Dubois, D., Fargier, H., Fortin, J.: A generalized vertex method for computing with fuzzy intervals. W: *Proceedings of the IEEE International Conference on Fuzzy Systems*, vol. 1, s. 541–546 (2004)

[24] Dubois, D., Prade, H.: Operations on fuzzy numbers. *International Journal of Systems Science* **9**(6), s. 613–626 (1978)

[25] Duch, W.: Uncertainty of data, fuzzy membership functions, and multilayer perceptrons. *IEEE Transactions on Neural Networks* **16**(1), s. 10–23 (2005)

[26] Dutta, P., Boruah, H., Ali, T.: Fuzzy arithmetic with and without using α-cut method: a comparative study. *International Journal of Latest Trends in Computing* **2**(1), s. 99–107 (2011)

[27] Dymova, L.: *Soft computing in economics and finance.* Springer Verlag, Berlin, Heidelberg (2011)

[28] Ezadi, S., Allahviranloo, T.: Numerical solution of linear regression based on Z-numbers by improved neural network. *Intelligent Automation & Soft Computing*, s. 1–11 (2017)

[29] Fodor, J., Bede, B.: Arithmetics with fuzzy numbers: a comparative overview. W: *Proceeding of 4th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence*, s. 54–68. Herlany, Slovakia (2006)

[30] Friedman, J.: *A variable span smoother.* LCS Technical Report 5, SLAC PUB-3466, Stanford Univerity, Laboratory for Computational Statistics (1984)

[31] Gacek, A.: Granular modelling of signals: a framework of granular computing. *Information Sciences* **221**, s. 1–11 (2013)

[32] Garimella, R.: *A simple introduction to moving least squares and local regression estimation.* Technical Report No. LA-UR-17-24975, Los Alamos National Laboratory, United States (2017)

[33] Gauger, U., Turrin, S., Hanss, M., Gaul, L.: A new uncertainty analysis for the transformation method. *Fuzzy Sets and Systems* **159**(11), s. 1273–1291 (2008)

[34] Geladi, P., Kowalski, B.: Partial least-squares regression: a tutorial. *Analytica chimica acta* **185**, s. 1–17 (1986)

[35] Giannini, O., Hanss, M.: An interdependency index for the outputs of uncertain systems. *Fuzzy Sets and Systems* **159**(11), s. 1292–1308 (2008)

[36] Goguen, J.: L-fuzzy sets. *Journal of mathematical analysis and applications* **18**(1), s. 145–174 (1967)

[37] Goodman, J., O'Rourke, J., Indyk, P.: Chapter 39: Nearest neighbours in high-dimensional spaces. W: *Handbook of Discrete and Computational Geometry*. CRC Press (2004)

[38] Grzegorzewski, P.: Metrics and orders in space of fuzzy numbers. *Fuzzy Sets and Systems* **97**, s. 83–94 (1998)

[39] Gutowski, M.: Interval experimental data fitting. W: *Focus on Numerical Analysis*, s. 27–70. New York: Nova Science Publishers (2006)

[40] Gutowski, M.: Breakthrough in interval data fitting I. The role of hausdorff distance. *Prace Naukowe Politechniki Warszawskiej. Elektronika* **169**, s. 63–72 (2009)

[41] Gutowski, M.: Breakthrough in interval data fitting II. From ranges to means and standard deviations. *Prace Naukowe Politechniki Warszawskiej. Elektronika* **169**, s. 73–78 (2009)

[42] Haag, T., Hanss, M.: Comprehensive modeling of uncertain systems using fuzzy set theory. W: *Nondeterministic Mechanics*, s. 193–226. Springer, Vienna (2012)

[43] Hamrawi, H., Coupland, S.: Type-2 fuzzy arithmetic using alpha-planes. W: *Proceedings of the IFSA-EUSFLAT Conference*, s. 606–611. Portugal (2009)

[44] Hand, D., Mannila, H., Smyth, P.: *Principles of data mining*. The MIT Press (2000)

[45] Hansen, E.: A generalized interval arithmetic. W: K. Nickel (red.) *Interval matematics, Lecture Notes in Computer Science* 29, s. 7–18. Springer Verlag, Berlin, Heidelberg (1975)

[46] Hanss, M.: *Applied fuzzy arithmetic*. Springer Verlag, Berlin, Heidelberg (2005)

[47] Hanss, M.: Fuzzy arithmetic for uncertainty analysis. W: R. Seising, E. Trillas, C. Moraga, S. Termini (red.) *On Fuzziness: A Homage to Lotfi A. Zadeh* – vol. 1, s. 235–240. Springer (2013)

[48] Hanss, M., Turrin, S.: "The future is fuzzy" – An approach to comprehensive modeling and analysis of systems with epistemic uncertainties. W: *Proceedings of the Leuven Symposium on Applied Mechanics in Engineering – LSAME'08*. Leuven, Belgium (2008)

[49] Harrell Jr., F.: *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer (2015)

[50] Hillier, F., Lieberman, G.: *Introduction to mathematical programming*. McGraw-Hill (1977)

[51] Horová, I., Kolácek, J., Zelinka, J.: *Kernel Smoothing in MATLAB: theory and practice of kernel smoothing*. World Scientific Publishigng, Singapore (2012)

[52] Jabbarova, K., Huseynov, O.: A Z-valued regression model for a catalytic cracking process. W: *Proceedings of the 11th International Conference on Applications of Fuzzy Systems and Soft Computing*, s. 165–174 (2014)

[53] Jaroszewicz, S., Korzeń, M.: Arithmetic operations on independent random variables: a numerical approach. *SIAM Journal on Scientific Computing* **34**(3), s. 1251–1265 (2012)

[54] Kaufmann, A., Gupta, M.: *Introduction to fuzzy arithmetic*. Van Nostrand Reinhold, New York (1991)

[55] Klir, G.: Fuzzy arithmetic with requisite constraints. *Fuzzy Sets and Systems* **91**(2), s. 165–175 (1997)

[56] Klir, G., Pan, Y.: Constrained fuzzy arithmetic: basic questions and some answers. *Soft Computing* **2**(2), s. 100–108 (1998)

[57] Kloke, J., McKean, J.: *Nonparametric statistical methods using R.* Chapman and Hall/CRC (2014)

[58] Kordos, M., Blachnik, M., Strzempa, D.: Do we need whatever more than k-NN? W: L. Rutkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, J. Zurada (red.) *Proceedings of 10-th International Conference on Artificial Intelligence and Soft Computing*, LNCS, vol. 6113, s. 414–421. Springer, Heidelberg (2010)

[59] Koronacki, J., Ćwik, J.: *Statystyczne systemy uczące się.* Wydawnictwa Naukowo-Techniczne, Warszawa (2005)

[60] Korzeń, M., Klęsk, P.: Sets of approximating functions with finite Vapnik-Czervonenkis dimension for nearest-neighbors algorithm. *Pattern Recognition Letters* **32**, s. 1882–1893 (2011)

[61] Kramosil, I., Michalek, J.: Fuzzy metrics and statistical metric spaces. *Kybernetika* **11**(5), s. 336–344 (1975)

[62] Landowski, M.: Differences between Moore's and RDM interval arithmetic. W: *Proceedings of Thirteenth International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets.* Warsaw, Poland (2014)

[63] Lerman, P.: Fitting segmented regression models by grid search. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **29**(1), s. 77–84 (1980)

[64] Li, Q., Liu, S.: The foundation of the grey matrix and the grey input-output analysis. *Applied Mathematical Modelling* **32**(3), s. 267–291 (2008)

[65] Lima Neto, E., Carvalho, F.: Centre and range method for fitting a linear regression model to symbolic interval data. *Computational Statistics & Data Analysis* **52**(3), s. 1500–1515 (2008)

[66] Lima Neto, E., Carvalho, F.: Constrained linear regression models for symbolic interval-valued variables. *Computational Statistics & Data Analysis* **54**(2), s. 333–347 (2010)

[67] Liu, B.: *Uncertainty theory.* Springer Verlag, Berlin, Heidelberg (2010)

[68] Liu, S., Lin, Y.: *Grey systems, theory and applications.* Springer Verlag, Berlin, Heidelberg (2010)

[69] Mendel, J.: Type-2 fuzzy sets and systems: An overview. *IEEE Computational Intelligence Magazine* **2**(2), s. 20–29 (2007)

[70] Ming, M., Friedman, M., Kandel, A.: General fuzzy least squares. *Fuzzy Sets and Systems* **88**, s. 107–118 (1997)

[71] Modarres, M., Nasrabadi, M., Nasrabadi, E., Mohtashmi, G.: Evaluation of fuzzy linear regression models: a mathematical programming approach. W: *Proceedings of 4th Seminar on Fuzzy Sets and its Applications*, s. 129–135. Iran (2003)

[72] Moore, A., Atkeson, C., Schaal, S.: *Memory-based learning for control.* Technical report CMU-RI-TR-95-18, Carnegie-Mellon University, Robotics Institute (1995)

[73] Moore, R.: *Interval analysis.* Prentice Hall, Englewood Cliffs, New Jork (1966)

[74] Moore, R., Kearfott, R., Cloud, M.: *Introduction to interval analysis.* Society for Industrial and Applied Mathematics, Philadelphia (2009)

[75] Muggeo, V.: Testing with a nuisance parameter present only under the alternative: a score-based approach with application to segmented modelling. *Journal of Statistical Computation and Simulation* **86**(15), s. 3059–3067 (2016)

[76] Myers, R., Myers, R.: *Classical and modern regression with applications*, vol. 2. Duxbury Press Belmont, CA (1990)

[77] Nadaraya, E.: On estimating regression. *Theory of Probability & Its Applications* **9**(1), s. 141–142 (1964)

[78] Nasrabadi, M., Nasrabadi, E.: A mathematical-programming approach to fuzzy linear regression analysis. *Applied Mathematics and Computation* **155**(3), s. 873–881 (2004)

[79] Nasrabadi, M., Nasrabadi, E., Nasrabady, A.: Fuzzy linear regression analysis: a multi-objective programming approach. *Applied mathematics and computation* **163**(1), s. 245–251 (2005)

[80] Oosterbaan, R., Sharma, D., Singh, K., Rao, K.: Crop production and soil salinity: evaluation of field data from India by segmented linear regression with breakpoint. W: *Proceedings of the symposium on land drainage for salinity control in arid and semi-arid regions*, vol. 3, s. 373–383. Cairo, Egypt (1990)

[81] Özelkan, E., Duckstein, L.: Multi-objective fuzzy regression: a general framework. *Computers & Operations Research* **27**(7–8), s. 635–652 (2000)

[82] Pawlak, Z.: *Rough sets: theoretical aspects of reasoning about data.* Dordrecht: Kluwer Academic Publishers (1991)

[83] Pedrycz, W.: From fuzzy data analysis and fuzzy regression to granular fuzzy data analysis. *Fuzzy Sets and Systems* **274**, s. 12–17 (2015)

[84] Pedrycz, W., Skowron, A., Kreinovich, V.: *Handbook of granular computing.* John Wiley & Sons, Chichester (2008)

[85] Peters, G.: Fuzzy linear regression with fuzzy intervals. *Fuzzy Sets and Systems* **63**, s. 45–55 (1994)

[86] Piegat, A.: *Fuzzy modeling and control.* Physica Verlag, Heidelberg-New York (2001)

[87] Piegat, A., Landowski, M.: Is the conventional interval-arithmetic correct? *Journal of Theoretical and Applied Computer Science* **6**(2), s. 27–44 (2012)

[88] Piegat, A., Landowski, M.: Multidimensional approach to interval-uncertainty calculations. W: *New trends in Fuzzy Sets, Intuitionistic Fuzzy Sets, Generalized Nets and Related Topics*, s. 137–152. System Research Institute of Polish Academy of Sciences, Warsaw, Poland (2013)

[89] Piegat, A., Landowski, M.: Two interpretations of multidimensional RDM interval arithmetic: multiplication and division. *International Journal of Fuzzy Systems* **15**(4), s. 486–496 (2013)

[90] Piegat, A., Landowski, M.: Horizontal membership function and examples of its applications. *International Journal of Fuzzy Systems* **17**(1), s. 22–30 (2015)

[91] Piegat, A., Landowski, M.: Aggregation of inconsistent expert opinions with use of horizontal intuitionistic membership functions. W: K. Atanassov, O. Castillo, J. Kacprzyk (red.) *Novel Developments in Uncertainty Representation and Processing*, s. 215–224. Springer (2016)

[92] Piegat, A., Pluciński, M.: Computing with Words with the use of inverse RDM models of membership functions. *International Journal of Applied Mathematics and Computer Science* **25**(3), s. 675–688 (2015)

[93] Piegat, A., Pluciński, M.: Fuzzy number addition with the application of horizontal membership functions. *The Scientific World Journal*, Article ID: 367214, s. 1–16 (2015)

[94] Piegat, A., Pluciński, M.: Some advantages of the RDM-arithmetic of Intervally-Precisiated Values. *International Journal of Computational Intelligence Systems* **8**(6), s. 1192–1209 (2015)

[95] Piegat, A., Pluciński, M.: Fuzzy number division and the multi-granularity phenomenon. *Bulletin of the Polish Academy of Sciences, Technical Sciences* **65**(4), s. 497–511 (2017)

[96] Piegat, A., Wąsikowska, B., Korzeń, M.: Application of the self-learning, 3-point mini-model for modeling of unemployment rate in poland. *Studia Informatica* **27**, s. 59–69 (2010) [in Polish]

[97] Piegat, A., Wąsikowska, B., Korzeń, M.: Differences between the method of mini-models and the k-nearest neighbors on example of modeling of unemployment rate in Poland. W: *Proceedings of 9th Conference on Information Systems in Management*, s. 34–43. WULS Press, Warsaw, Poland (2011)

[98] Pietrzykowski, M.: Effectiveness of mini-models method when data modelling within a 2D-space in an information deficiency situation. *Journal of Theoretical and Applied Computer Science* **6**(3), s. 21–27 (2012)

[99] Pietrzykowski, M.: Comparison between mini-models based on multidimensional polytopes and k-nearest neighbor method: case study of 4D and 5D problems. W: *Soft Computing in Computer and Information Science*, s. 107–118. Springer (2015)

[100] Pietrzykowski, M.: Local regression algorithms based on centroid clustering methods. *Procedia Computer Science* **112**, s. 2363–2371 (2017)

[101] Pietrzykowski, M., Piegat, A.: Geometric approach in local modeling: learning of mini-models based on n-dimensional simplex. W: *Proceedings of International Conference on Artificial Intelligence and Soft Computing*, s. 460–470. Springer (2015)

[102] Pietrzykowski, M., Pluciński, M.: Mini-model method based on k-means clustering. *Przegląd elektrotechniczny* **93**(1), s. 73–76 (2017)

[103] Pluciński, M.: Application of data with missing attributes in the probability RBF neural network learning and classification. W: J. Sołdek, L. Drobiazgiewicz (red.) *Artificial Intelligence and Security in Computing Systems: Proceedings of the 9-th International Conference ACS'2002*, s. 63–72. Boston/Dordrecht/London: Kluwer Academic Publishers (2003)

[104] Pluciński, M.: Application of the information-gap theory for evaluation of nearest neighbours method robustness to data uncertainty. *Przegląd elektrotechniczny* **88**(10b), s. 272–275 (2012)

[105] Pluciński, M.: Mini-models – local regression models for the function approximation learning. W: *Artificial Intelligence and Soft Computing* (LNCS, vol. 7268, part II), s. 160–167. Springer, Berlin Heidelberg (2012)

[106] Pluciński, M.: Nonlinear ellipsoidal mini-models – application for the function approximation task. *Przegląd elektrotechniczny* **88**(10b), s. 247–251 (2012)

[107] Pluciński, M.: Evaluation of the mini-models robustness to data uncertainty with the application of the information-gap theory. W: *Artificial Intelligence and Soft Computing* (Lecture Notes in Artificial Intelligence, vol. 7895, part II), s. 230–241. Springer, Berlin Heidelberg (2013)

[108] Pluciński, M.: Application of mini-models to the interval information granules processing. W: A. Wiliński, I. El Fray, J. Pejaś (red.) *Soft Computing in Computer and Information Science*, s. 37–48. Springer International Publishing (2015)

[109] Pluciński, M.: Solving Zadeh's challenge problems with the application of RDM-arithmetic. W: *Artificial Intelligence and Soft Computing* (Lecture Notes in Artificial Intelligence, vol. 9119, part I), s. 239–248. Springer International Publishing, Switzerland (2015)

[110] Pluciński, M.: Processing of Z^+-numbers using the k nearest neighbors method. W: J. Pejaś, I. El Fray, T. Hyla, J. Kacprzyk (red.) *Advances in Soft and Hard Computing*, s. 76–85. Springer, Cham (2019)

[111] Pluciński, M., Pietrzykowski, M.: Application of the k nearest neighbors method to fuzzy data processing. *Przegląd elektrotechniczny* **93**(1), s. 77–81 (2017)

[112] Rachev, S., Klebanov, L., Stoyanov, S., Fabozzi, F.: *The Methods of Distances in the Theory of Probability and Statistics*. Springer Science & Business Media (2013)

[113] Rattanalertnusorn, A., Phannarook, S., Jaroengeratikun, U.: A method for formulating fuzzy linear regression model and estimating the model parameters. W: *Proceedings of International Conference on Applied Statistics*. Pattaya, Thailand (2015)

[114] Redden, D., Woodall, W.: Properties of certain fuzzy linear regression methods. *Fuzzy Sets and Systems* **64**, s. 361–375 (1994)

[115] Redden, D., Woodall, W.: Further examination of fuzzy linear regression. *Fuzzy Sets and Systems* **79**, s. 203–211 (1996)

[116] Ruoning, X.: S-curve regression model in fuzzy environment. *Fuzzy Sets and Systems* **90**, s. 317–326 (1997)

[117] Sadikoglu, F., Huseynov, O., Memmedova, K.: Z-regression analysis in psychological and educational researches. *Procedia Computer Science* **102**, s. 385–389 (2016)

[118] Sakawa, M., Yano, H.: Multiobjective fuzzy linear regression analysis for fuzzy input–output data. *Fuzzy Sets and Systems* **47**, s. 173–181 (1992)

[119] Sevastjanov, P., Dymova, L.: A new method for solving interval and fuzzy equations: linear case. *Information Sciences* **17**, s. 925–937 (2009)

[120] Shary, S.: On controlled solution set of interval algebraic systems. *Interval Computations* **6**(6), s. 66–75 (1992)

[121] Shary, S.: Solving the tolerance problem for interval linear systems. *Interval Computations* **2**, s. 6–26 (1994)

[122] Shary, S.: On optimal solution of interval linear equations. *SIAM Journal on Numerical Analysis* **32**(2), s. 610–630 (1995)

[123] Shary, S.: *Finite-dimensional interval analysis.* Książka dostępna w formie elektronicznej po rosyjsku: http://www.nsc.ru/interval/Library/InteBooks (2015)

[124] Simonoff, J.: *Smoothing methods in statistics.* Springer Science & Business Media (2012)

[125] Springer, M.: *The Algebra of Random Variables.* John Wiley & Sons, New York, Chichester, Brisbane (1979)

[126] Stefanini, L.: A generalization of Hukuhara difference and division for interval and fuzzy arithmetic. *Fuzzy sets and systems* **161**(11), s. 1564–1584 (2010)

[127] Stefanini, L.: New tools in fuzzy arithmetic with fuzzy numbers. W: *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications*, s. 471–480. Springer (2010)

[128] Stefanini, L., Sorini, L., Guerra, M.: Parametric representation of fuzzy numbers and application to fuzzy calculus. *Fuzzy Sets and Systems* **157**(18), s. 2423–2455 (2006)

[129] Stefanini, L., Sorini, L., Guerra, M.: Fuzzy numbers and fuzzy arithmetic. W: W. Pedrycz, A. Skowron, V. Kreinovich (red.) *Handbook of Granular Computing*, s. 249–283. John Wiley & Sons, Chichester (2008)

[130] Tanaka, H.: Fuzzy data analysis by possibilistic linear models. *Fuzzy Sets and Systems* **24**, s. 363–375 (1987)

[131] Tanaka, H., Hayashi, I., Watada, J.: Possibilistic linear regression analysis for fuzzy data. *European Journal of Operational Research* **40**(3), s. 389–396 (1989)

[132] Tanaka, H., Uejima, S., Asai, K.: Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man and Cybernetics* **12**, s. 903–907 (1982)

[133] Tanaka, H., Watada, J.: Possibilistic linear systems and their application to the linear regression model. *Fuzzy Sets and Systems* **27**, s. 275–289 (1988)

[134] Tang, W., Li, X., Zhao, R.: Metric spaces of fuzzy variables. *Computers & Industrial Engineering* **57**, s. 1268–1273 (2009)

[135] Wang, H., Tsaur, R.: Resolution of fuzzy regression model. *European Journal of Operational Research* **126**, s. 637–650 (2000)

[136] Wasserman, P.: *Advanced methods in neural computing.* Van Nostrand Reinhold, New York (1993)

[137] Watson, S.: Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics*, Series A **26**(4), s. 359–372 (1964)

[138] Wu, C., Chang, N.: Grey input-output analysis and its application for environmental cost allocation. *European Journal of Operational Research* **145**(1), s. 175–201 (2003)

[139] Yoon, J., Jung, H., Choi, S.: Equivalence in alpha-level linear regression. *Communications of the Korean Statistical Society* **17**(4), s. 611–624 (2010)

[140] Zadeh, L.: Fuzzy sets. *Information and Control* **8**, s. 338–353 (1965)

[141] Zadeh, L.: Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems* **4**(2), s. 103–111 (1996)

[142] Zadeh, L.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy sets and systems* **90**(2), s. 111–127 (1997)

[143] Zadeh, L.: From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. *International Journal of Applied Mathematics and Computer Science* **12**(3), s. 307–324 (2002)

[144] Zadeh, L.: Toward a generalized theory of uncertainty (GTU) – an outline. *Information sciences* **172**(1), s. 1–40 (2005)

[145] Zadeh, L.: A note on Z-numbers. *Information Sciences* **181**, s. 2923–2932 (2011)

[146] Zhilin, S.: Simple method for outlier detection in fitting experimental data under interval error. *Chemometrics and Intelligent Laboratory Systems* **88**(1), s. 60–68 (2007)

Skorowidz

α-przekrój zbioru rozmytego, 54

arytmetyka
- interwałowa, 21
- interwałowa RDM, 22
  - korzyści, 25
  - własności, 24
- liczb losowych, 99
- liczb rozmytych, 55
  - RDM, 58
  - własności, 61
  - wielowymiarowa RDM, 56
- liczb Z, 99, 100
- liczb Z^+, 99

baza mini-modelu, 11
brzeg liczby rozmytej, 53

funkcja
- gęstości prawdopodobieństwa, 97
- przynależności, 53
  - pionowa, 56
  - pozioma, 57

granula informacyjna, 7

Hessian, 33, 37, 69, 74

interwał, 21
- zdegenerowany, 21

liczba
- rozmyta, 53
  - jednoelementowa, 65
  - normalna, 53
  - prostokątna, 65
  - trójkątna, 54
  - trapezowa, 54
  - wypukła, 53
- Z, 97
- Z^+, 97

metoda
- k najbliższych sąsiadów, 9
  - prosta, 9
  - ważona, 9
- najmniejszych kwadratów, 9
  - dla liczb rozmytych, 66, 71
  - dla liczb Z, 106, 114
  - interwałowa, 31, 35

metryka
- Hausdorffa, 29

mini-model
- liniowy, 9
- nieliniowy, 11
- z bazą hiper-elipsoidalną, 13

niepewność informacji, 7
nośnik liczby rozmytej, 53, 100

odległość
- interwałów, 29
- liczb rozmytych, 63
- liczb Z, 105

poziom niepewności, 7

regresja
- jądrowa, 7
- liniowa, 9
- lokalna, 6, 9

odcinkowa, 6
rozmyte ograniczenie, 97

walidacja krzyżowa, 12

Summary

Application of local regression models (mini-models) in the processing of uncertain information

The monograph describes first of all models based on mini-models, which belong to memory methods and are an example of the application of the local regression. In the following chapters of the book an attempt was made to adapt methods – which were originally developed for real data – to uncertain data (intervals, fuzzy numbers and Z numbers). Methods for creating local models based on imprecise information were developed as well as ways to determine the response of such models.

The dissertation has the following layout. The second chapter presents the concept of linear and non-linear mini-models using a hyper-spherical and hyper-ellipsoidal base. The chapter ends with the results of experiments in which the quality of operation and the accuracy of the models were tested.

The third chapter describes the first type of uncertainty, i.e. intervals. This part of the work discusses the basic concepts and arithmetic of intervals, and also presents the interval version of the least squares method, which is used in the creation of mini-models. This chapter, like the previous one, finishes the presentation of experimental research results. The fourth chapter presents the most important information about fuzzy numbers and their arithmetic. The main part of it is the discussion of the local modeling method for fuzzy data, based on the fuzzy version of the least squares method.

In the penultimate, fifth chapter, Z and Z^+ numbers and their arithmetic are characterized. Also here, the least squares method for data in the form of Z numbers is described and the local modeling methods that use it. Both the fourth and fifth chapter end the results of the research, whose main task was to demonstrate the reliability and good quality of the models work. The final part of the dissertation is the ending in which the benefits of the proposed solutions were summarized and the scope of further work was discussed.

Modeling based on local regression can be a very attractive alternative to classic fuzzy systems. The most important advantages of such modeling used to determine the fuzzy system response are listed below.

- First of all, the rule base does not have to be complete. Conclusions of some rules in the base can be difficult to determine, but even if there are no parts of the rules in the model, methods based on local regression can effectively determine its output.
- The rule base does not have to be consistent. In the case of rules with conflicting conclusions, the model averages them by determining the answer.
- The user can choose how many model rules are used to determine the output, and can easily determine the optimal quantity for a given modeling method. In classic fuzzy modeling, the number of rules in calculations is constant and depends on the method used to determine the output.

One of the greatest advantages of models based on local regression is their flexibility with regard to the type of information processed. The data can be homogeneous or of mixed type, whereby both the training data and the input data of the model can be in the form of Z or fuzzy numbers, intervals or real numbers. In every situation described in the research, the models returned correct and reliable results. The accuracy of models based on mini-models was usually comparable or slightly better than the accuracy of other memory models (eg based on kNN techniques).