UCZENIE MASZYNOWE

DRZEWA DECYZYJNE - REGRESJA

Dr hab. inż. Grzegorz Dudek
Wydział Elektryczny
Politechnika Częstochowska

Projekt finansowany w ramach programu Ministra Nauki i Szkolnictwa Wyższego pod nazwą „Regionalna Inicjatywa Doskonałości" w latach 2019 - 2022 nr projektu 020/RID/2018/19 kwota finansowania 12 000 000 PLN

Plan prezentacji

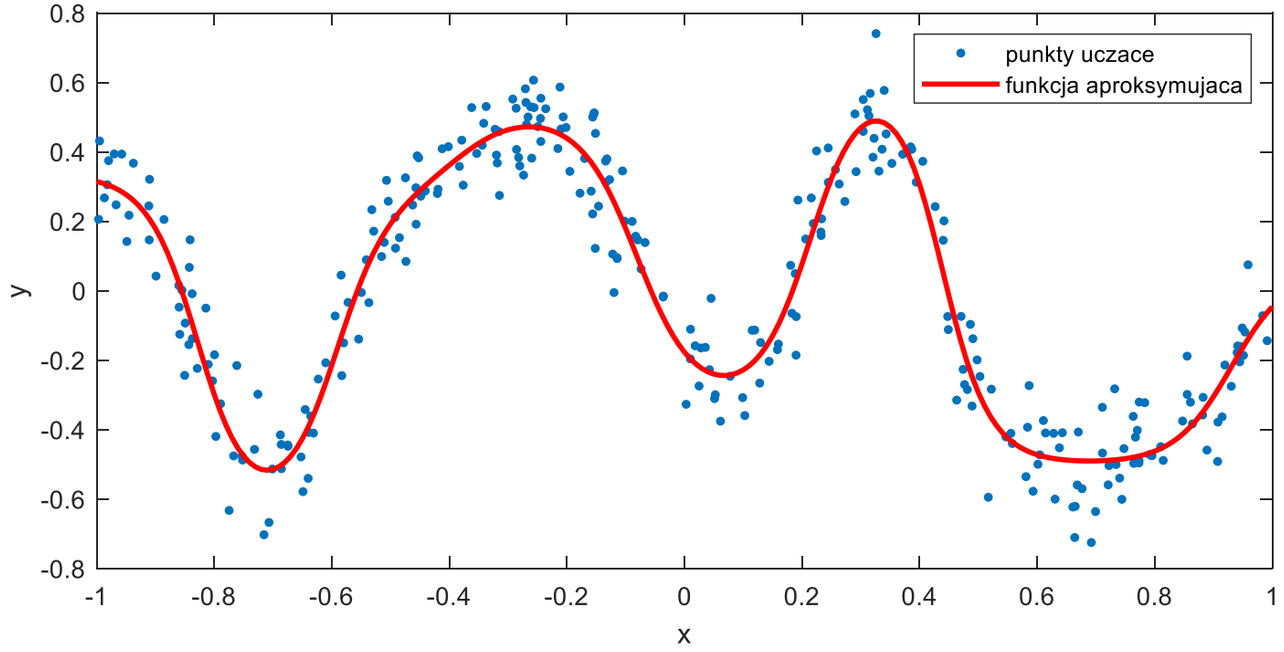
1. Informacje wstępne
2. Idea
3. Konstrukcja drzewa regresyjnego
4. Proces regresji za pomocą drzewa
5. Testy w węzłach drzewa
6. Przycinanie drzewa
7. Przykłady zastosowania drzewa regresyjnego
8. Podsumowanie

Informacje wstępne

Aproksymacja funkcji to procedura przybliżania jednej funkcji $f(x)$ za pomocą innej funkcji $h(x)$.

Przyczyny stosowania aproksymacji:

- funkcja aproksymowana dana jest w postaci punktów pomiarowych,
- funkcja aproksymowana wyrażona jest za pomocą skomplikowanej, niepraktycznej zależności analitycznej.



Informacje wstępne

System uczy się funkcji odwzorowującej przykłady **x** na zbiór liczb rzeczywistych *y*.

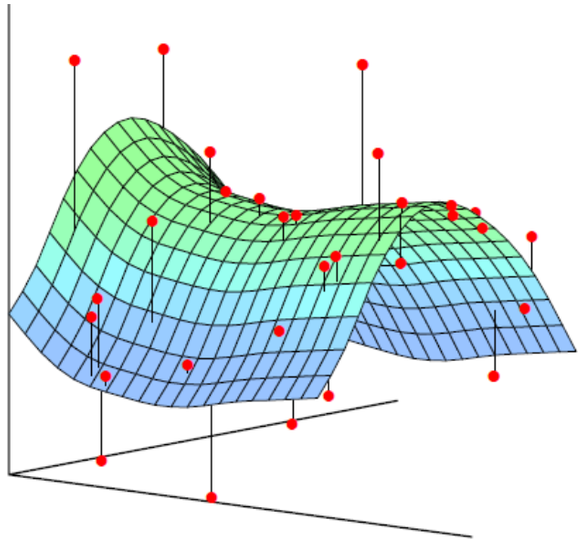
Przykłady trenujące:

argument funkcji + wartość funkcji: $\langle \mathbf{x}, y \rangle$

gdzie $y = f(\mathbf{x}) + \varepsilon$

Zbiór trenujący – $\Phi = \{\langle \mathbf{x}_1, y_1 \rangle, \langle \mathbf{x}_2, y_2 \rangle, \ldots, \langle \mathbf{x}_N, y_N \rangle\}$

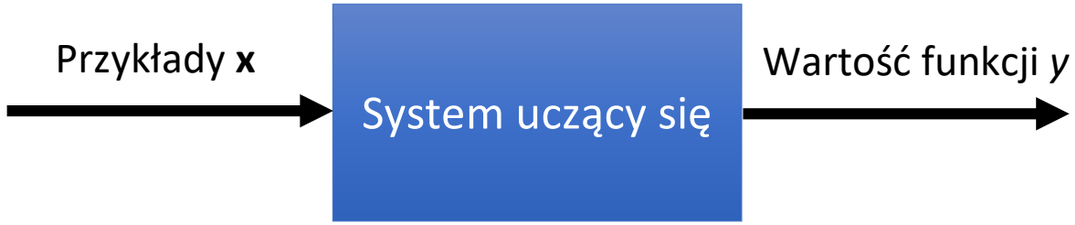
L.p.	**x**		*y*
	x_1	x_2	
1	5.5	6.0	275.375
2	–4.5	–7.0	–261.625
…	…	…	…
N	0.5	8.0	257.125



Informacje wstępne

Zadanie systemu uczącego się (uczenie z nadzorem) – znalezienie hipotezy dobrze przybliżającej nieznaną funkcję docelową $f : X \to \mathbb{R}$ dla przykładów trenujących i innych przykładów z dziedziny.

Hipoteza – funkcja przekształcająca przykłady w zbiór liczb rzeczywistych – $h : X \to \mathbb{R}$.



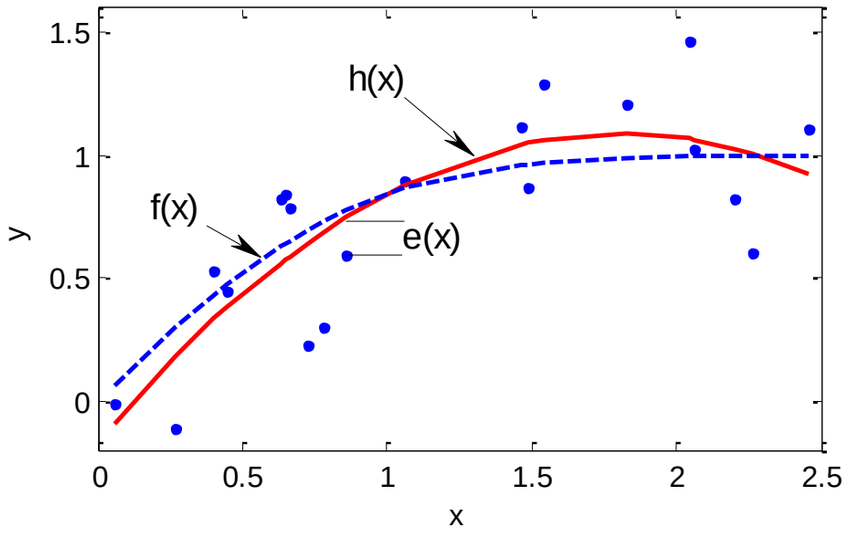
Informacje wstępne

Błąd aproksymacji

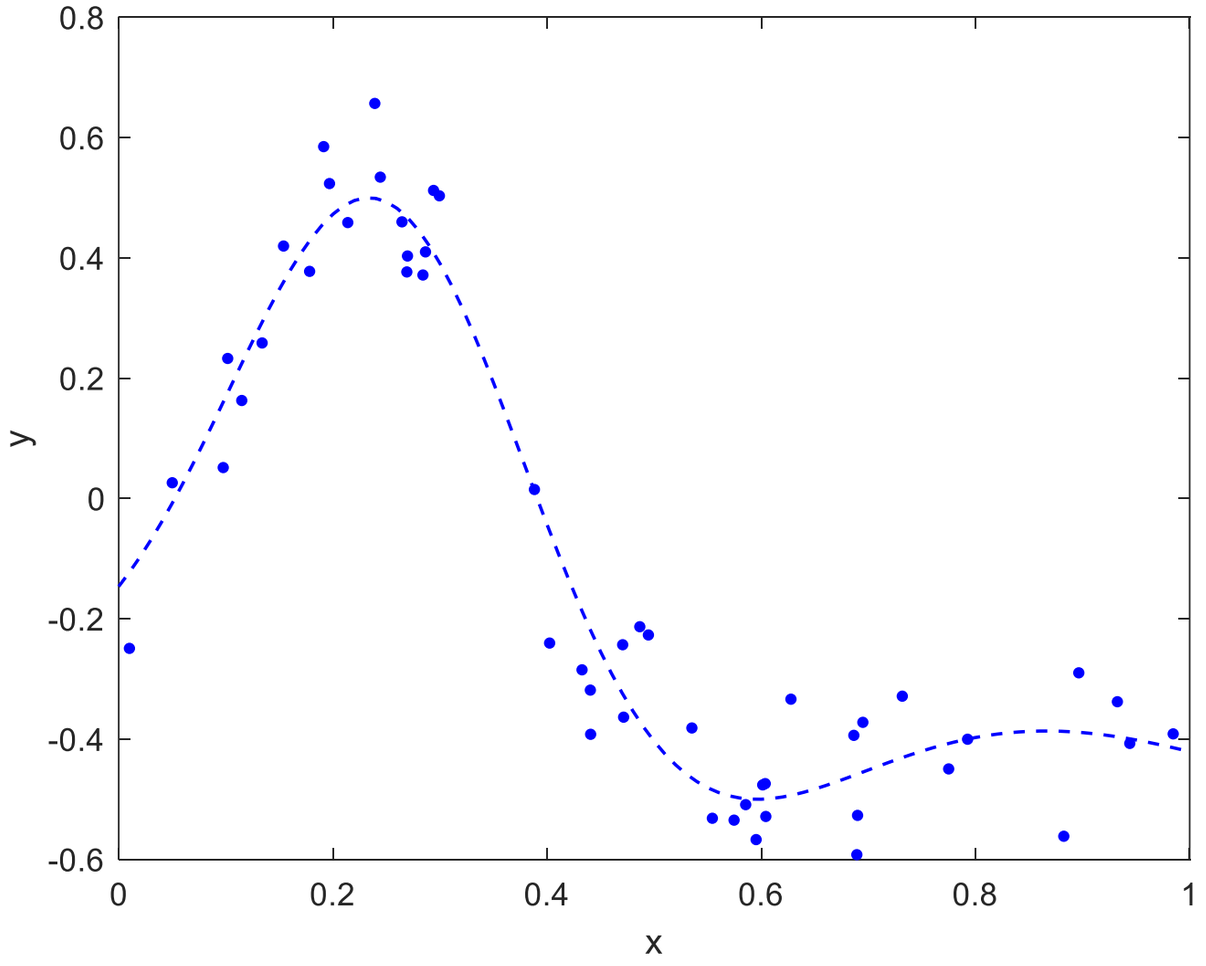
Błąd średniokwadratowy (MSE):

$$E(h) = \frac{1}{|\Phi|}\sum_{\mathbf{x}\in\Phi}(f(\mathbf{x}) - h(\mathbf{x}))^2 = \text{mean}(e^2(\mathbf{x}))$$

Ponieważ zwykle nie znamy prawdziwej wartości funkcji docelowej $f(\mathbf{x})$, w jej miejsce podstawiamy y ($y = f(\mathbf{x}) + \varepsilon$)

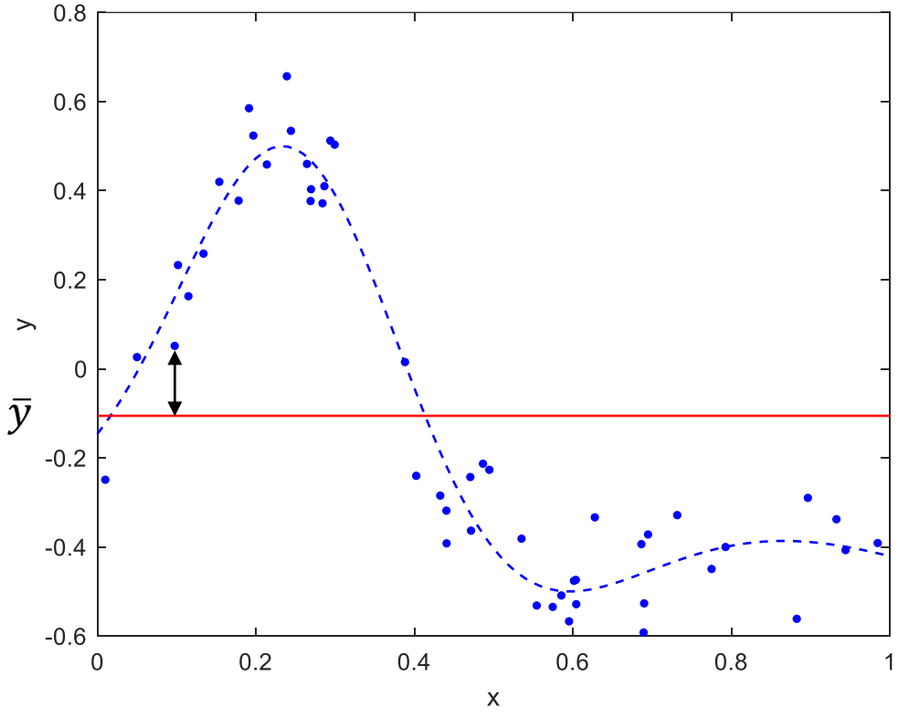


Idea



Idea

Miara zróżnicowania punktów:

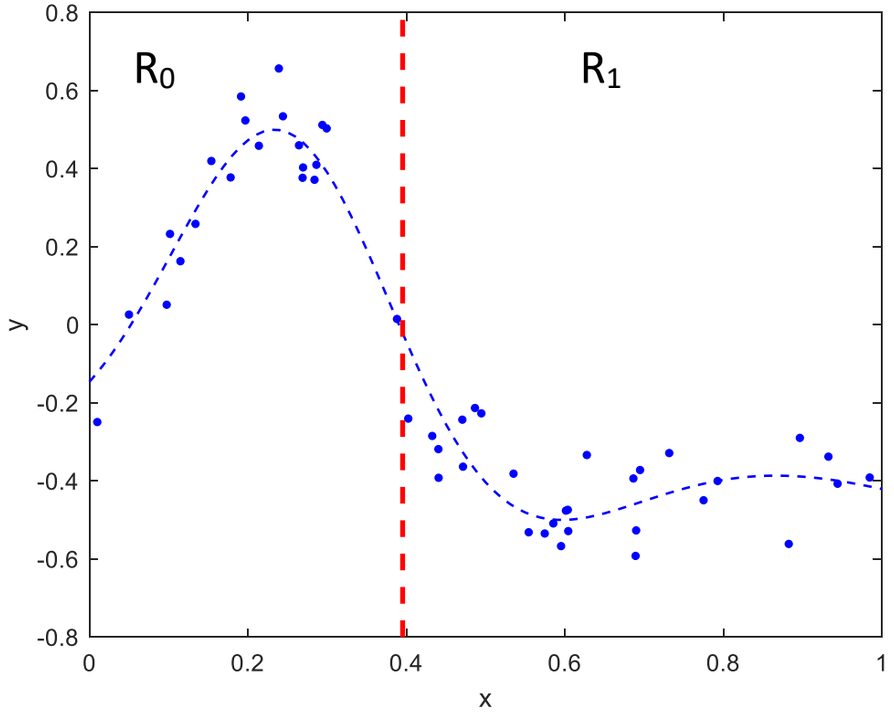


$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2$$

$$\varepsilon = 0{,}1607$$

Idea

Dzielimy zbiór punktów na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:

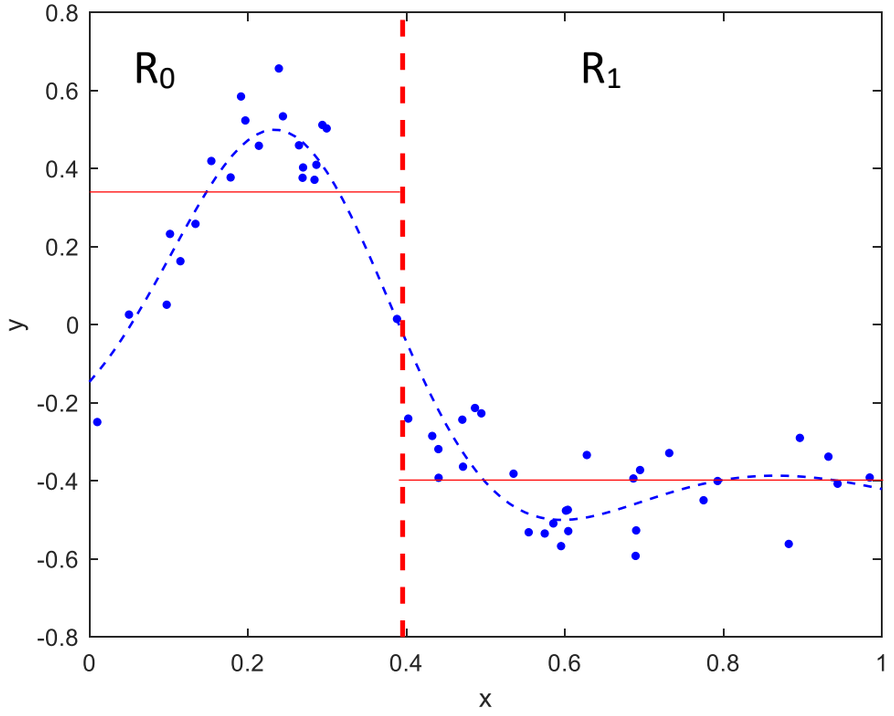


$$\varepsilon_{R_0} = \frac{1}{N_{R_0}} \sum_{i \in R_0} (y_i - \bar{y}_{R_0})^2 = 0{,}0485$$

$$\varepsilon_{R_1} = \frac{1}{N_{R_1}} \sum_{i \in R_1} (y_i - \bar{y}_{R_1})^2 = 0{,}0120$$

Idea

Dzielimy zbiór punktów na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:

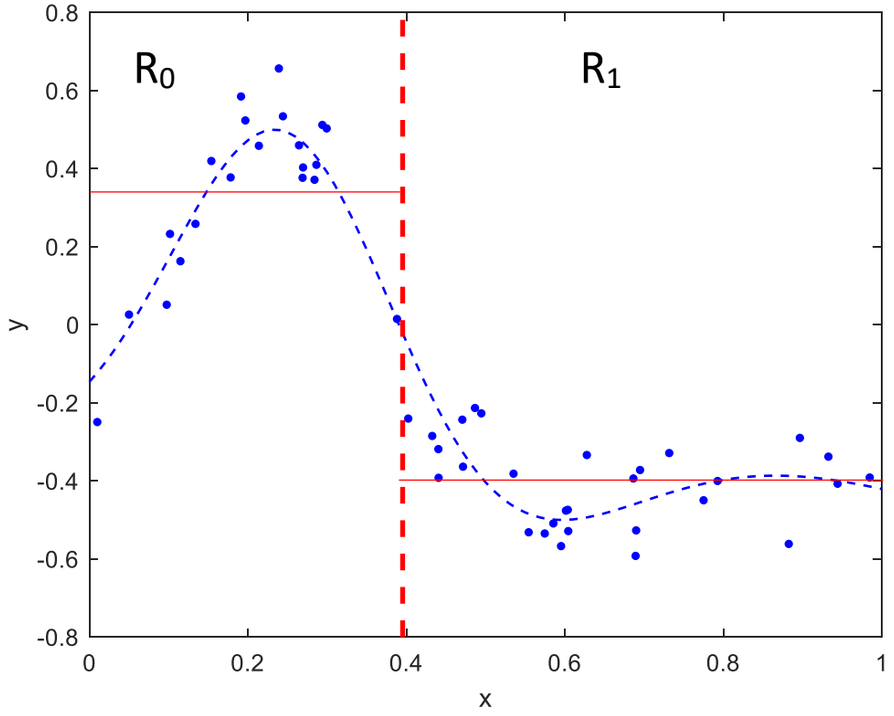


$$\varepsilon_{R_0} = \frac{1}{N_{R_0}} \sum_{i \in R_0} (y_i - \bar{y}_{R_0})^2 = 0{,}0485$$

$$\varepsilon_{R_1} = \frac{1}{N_{R_1}} \sum_{i \in R_1} (y_i - \bar{y}_{R_1})^2 = 0{,}0120$$

Idea

Dzielimy zbiór punktów na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:



$$\varepsilon_{R_0} = \frac{1}{N_{R_0}} \sum_{i \in R_0} (y_i - \bar{y}_{R_0})^2 = 0{,}0485$$

$$\varepsilon_{R_1} = \frac{1}{N_{R_1}} \sum_{i \in R_1} (y_i - \bar{y}_{R_1})^2 = 0{,}0120$$

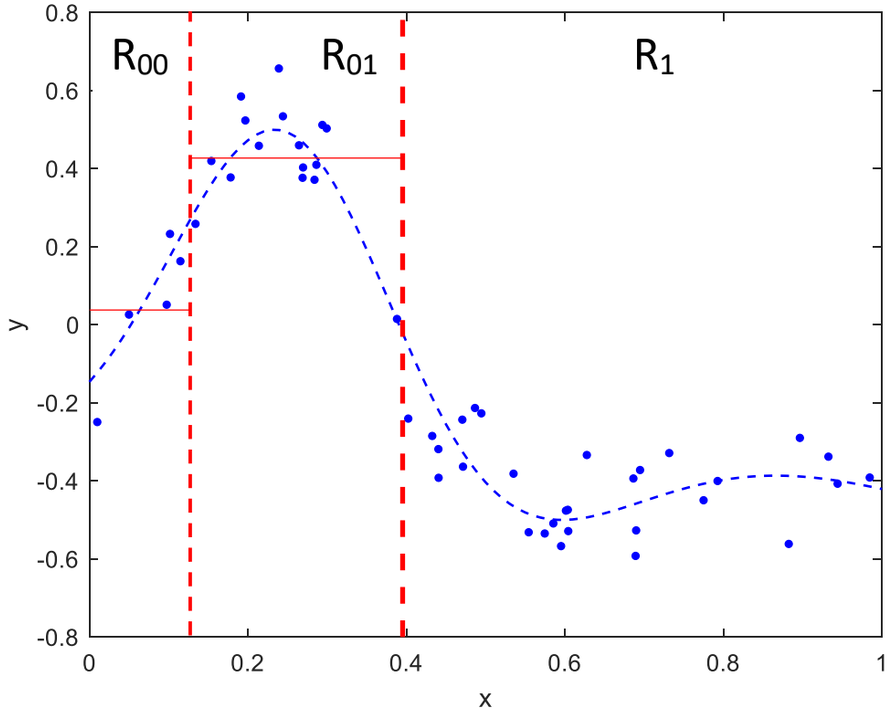
Redukcja zróżnicowania:

$$\Delta I = \varepsilon_R - \frac{N_{R_0}}{N_R} \varepsilon_{R_0} - \frac{N_{R_1}}{N_R} \varepsilon_{R_1} =$$

$$= 0{,}1607 - \frac{21}{50} 0{,}0485 - \frac{29}{50} 0{,}0120 = 0{,}1334$$

Idea

Dzielimy obszar R_0 na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:

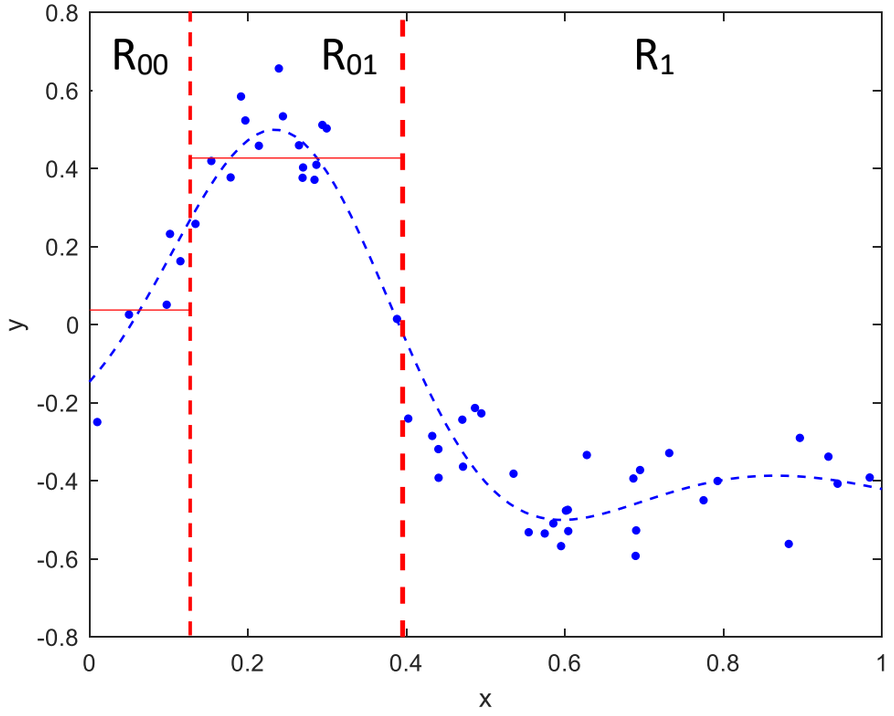


$$\varepsilon_{R_{00}} = \frac{1}{N_{R_{00}}} \sum_{i \in R_{00}} (y_i - \bar{y}_{R_{00}})^2 = 0{,}0272$$

$$\varepsilon_{R_{01}} = \frac{1}{N_{R_{01}}} \sum_{i \in R_{01}} (y_i - \bar{y}_{R_{01}})^2 = 0{,}0200$$

Idea

Dzielimy obszar R_0 na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:



$$\varepsilon_{R_{00}} = \frac{1}{N_{R_{00}}} \sum_{i \in R_{00}} (y_i - \bar{y}_{R_{00}})^2 = 0{,}0272$$

$$\varepsilon_{R_{01}} = \frac{1}{N_{R_{01}}} \sum_{i \in R_{01}} (y_i - \bar{y}_{R_{01}})^2 = 0{,}0200$$

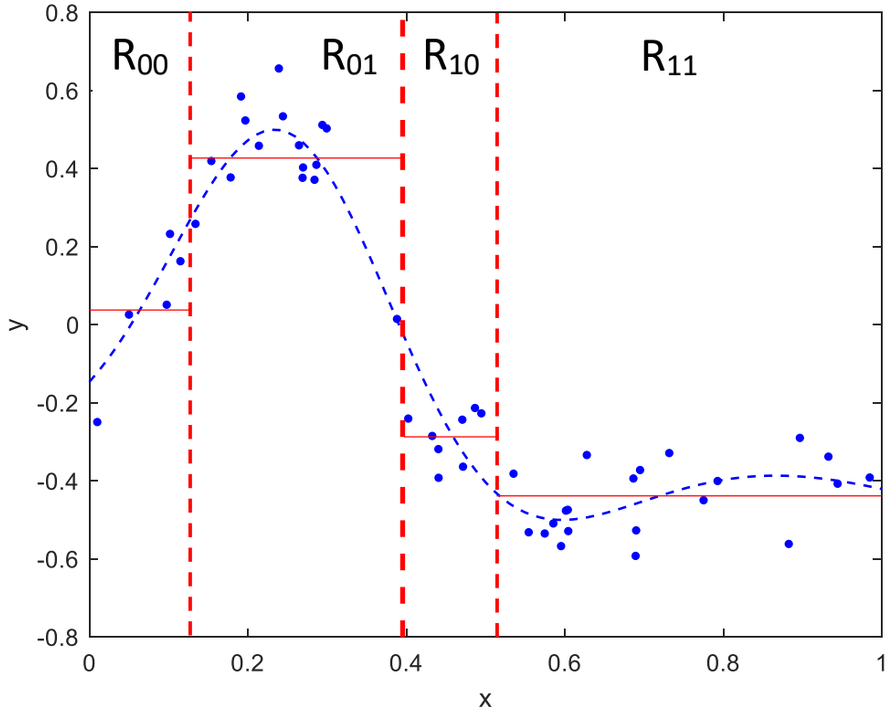
Redukcja zróżnicowania:

$$\Delta I = \varepsilon_{R_0} - \frac{N_{R_{00}}}{N_{R_0}} \varepsilon_{R_{00}} - \frac{N_{R_{01}}}{N_{R_0}} \varepsilon_{R_{01}} =$$

$$= 0{,}0485 - \frac{5}{21} 0{,}0272 - \frac{16}{21} 0{,}0200 = 0{,}0268$$

Idea

Dzielimy obszar R_1 na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:

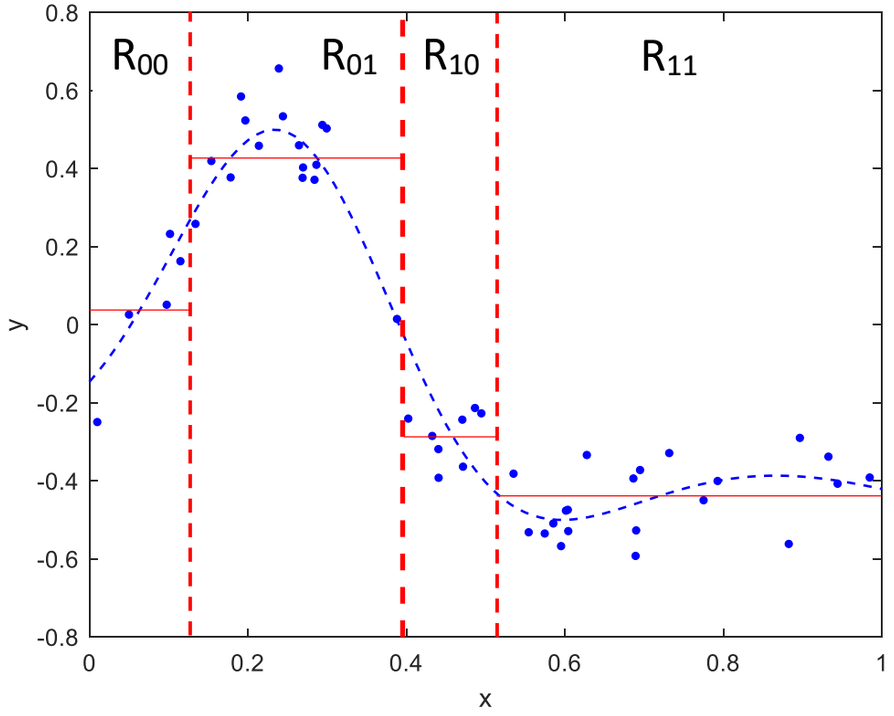


$$\varepsilon_{R_{10}} = \frac{1}{N_{R_{10}}} \sum_{i \in R_{10}} (y_i - \bar{y}_{R_{10}})^2 = 0{,}0039$$

$$\varepsilon_{R_{11}} = \frac{1}{N_{R_{11}}} \sum_{i \in R_{11}} (y_i - \bar{y}_{R_{11}})^2 = 0{,}0079$$

Idea

Dzielimy obszar R_1 na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:



$$\varepsilon_{R_{10}} = \frac{1}{N_{R_{10}}} \sum_{i \in R_{10}} (y_i - \bar{y}_{R_{10}})^2 = 0{,}0039$$

$$\varepsilon_{R_{11}} = \frac{1}{N_{R_{11}}} \sum_{i \in R_{11}} (y_i - \bar{y}_{R_{11}})^2 = 0{,}0079$$

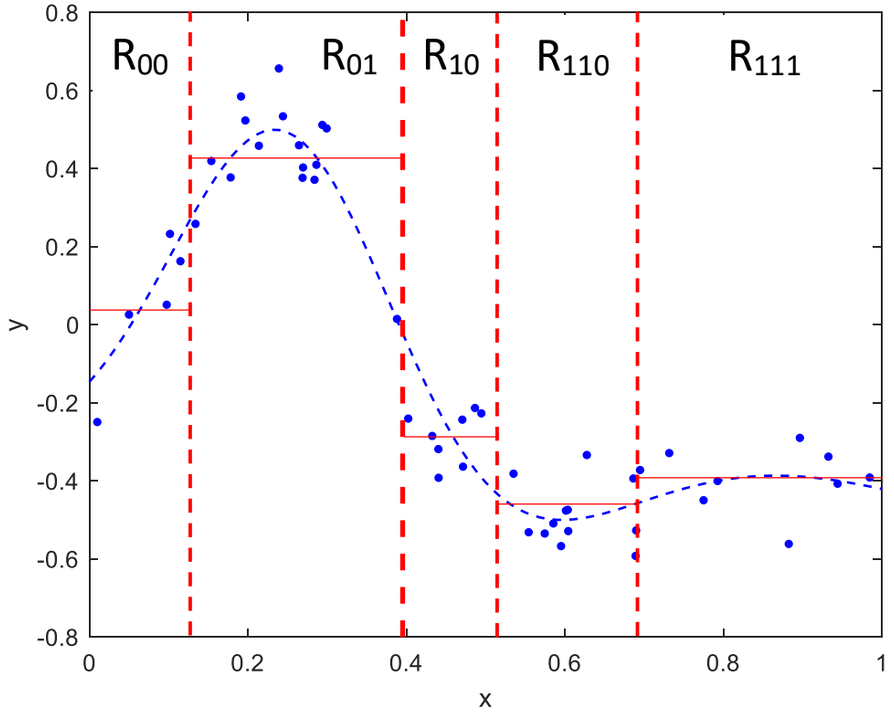
Redukcja zróżnicowania:

$$\Delta I = \varepsilon_{R_1} - \frac{N_{R_{10}}}{N_{R_1}} \varepsilon_{R_{10}} - \frac{N_{R_{11}}}{N_{R_1}} \varepsilon_{R_{11}} =$$

$$= 0{,}0120 - \frac{8}{29} 0{,}0039 - \frac{21}{29} 0{,}0079 = 0{,}0052$$

Idea

Dzielimy obszar R_{11} na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:

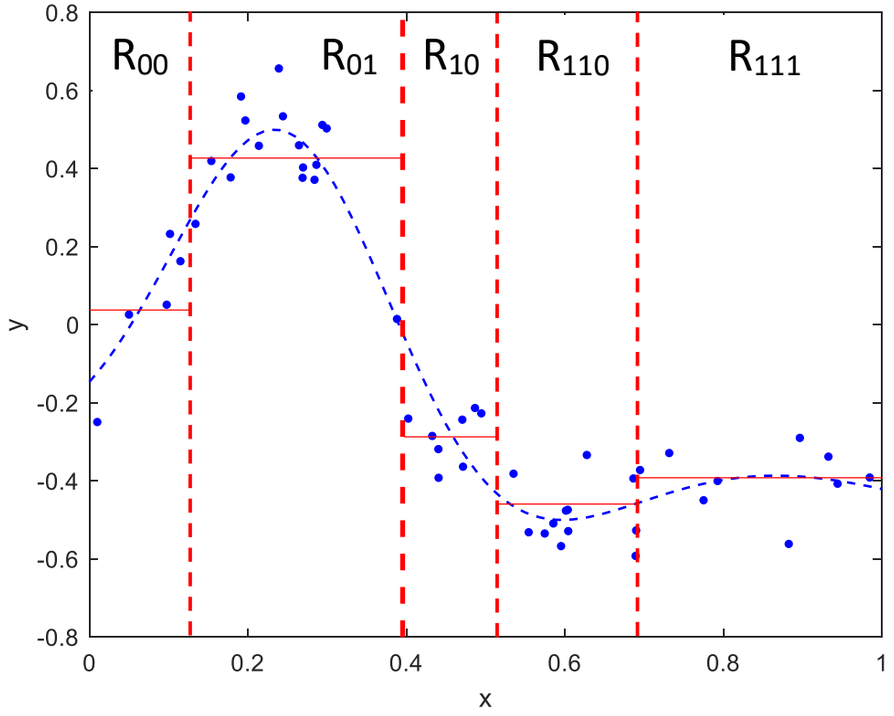


$$\varepsilon_{R_{110}} = \frac{1}{N_{R_{110}}} \sum_{i \in R_{110}} (y_i - \bar{y}_{R_{110}})^2 = 0{,}0058$$

$$\varepsilon_{R_{111}} = \frac{1}{N_{R_{111}}} \sum_{i \in R_{111}} (y_i - \bar{y}_{R_{111}})^2 = 0{,}0056$$

Idea

Dzielimy obszar R_{11} na dwie części względem *x*, tak aby po podziale uzyskać największą redukcję zróżnicowania:



$$\varepsilon_{R_{110}} = \frac{1}{N_{R_{110}}} \sum_{i \in R_{110}} (y_i - \bar{y}_{R_{110}})^2 = 0{,}0058$$

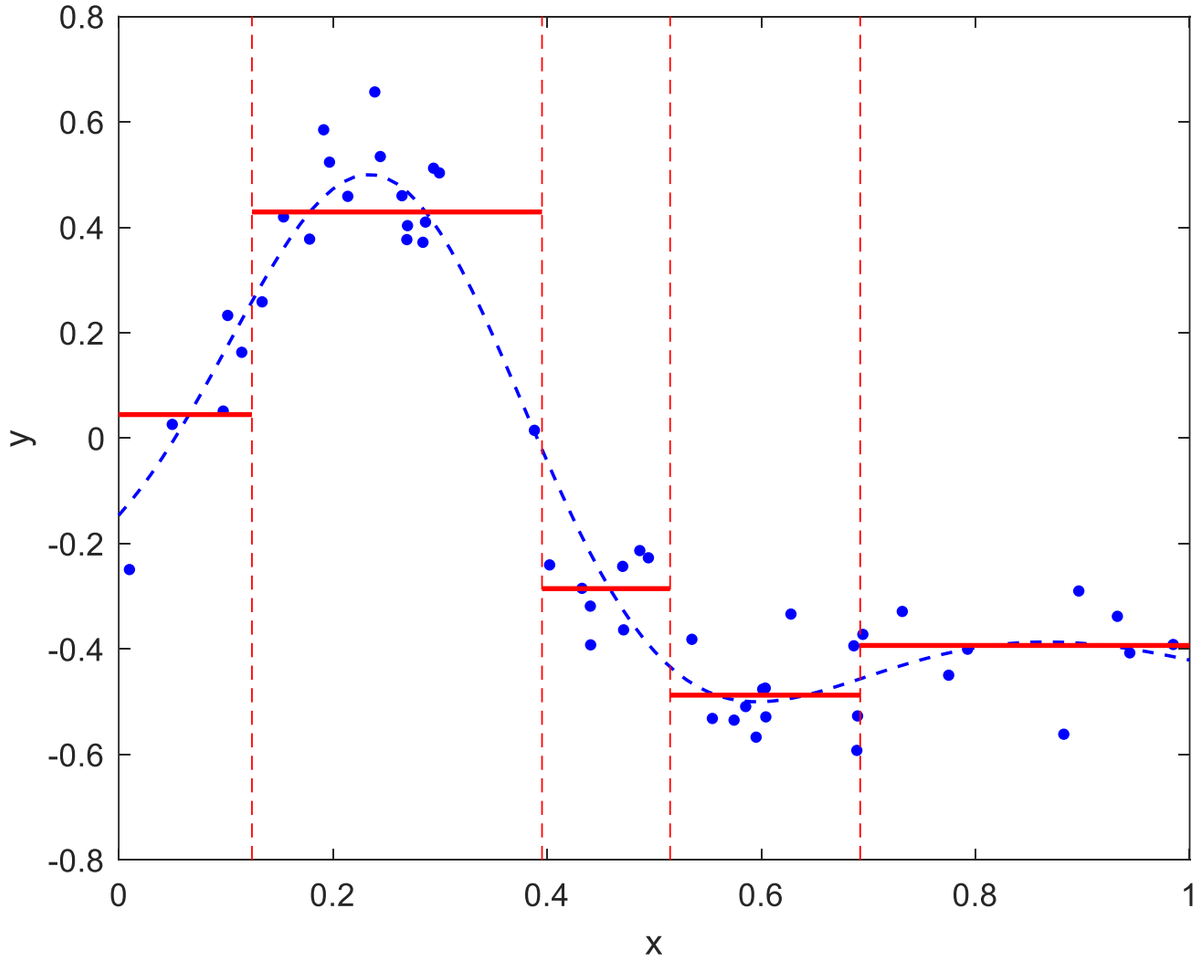
$$\varepsilon_{R_{111}} = \frac{1}{N_{R_{111}}} \sum_{i \in R_{111}} (y_i - \bar{y}_{R_{111}})^2 = 0{,}0056$$

Redukcja zróżnicowania:

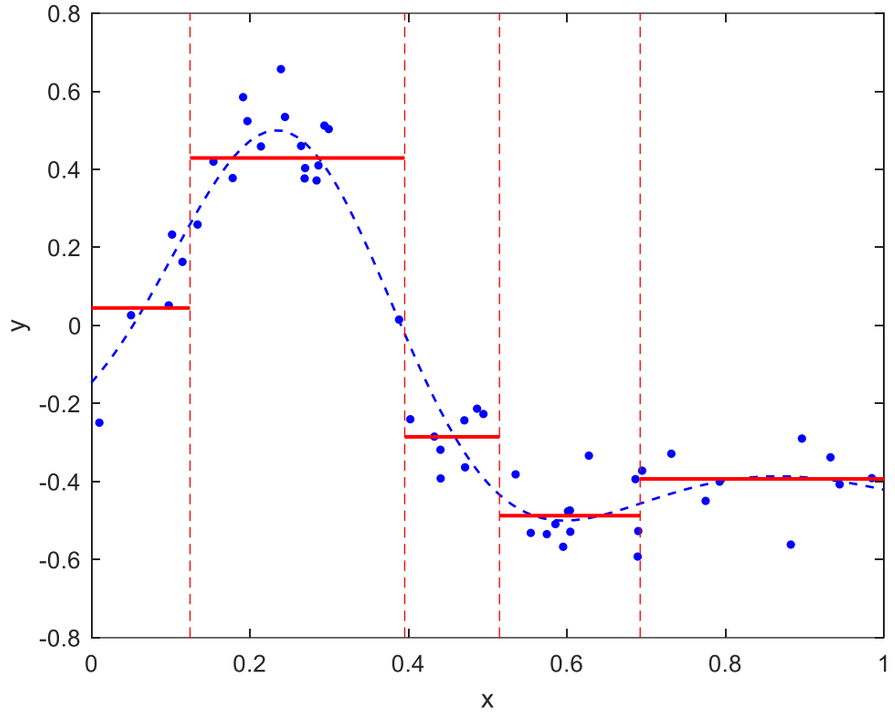
$$\Delta I = \varepsilon_{R_{11}} - \frac{N_{R_{110}}}{N_{R_{11}}} \varepsilon_{R_{110}} - \frac{N_{R_{111}}}{N_{R_{11}}} \varepsilon_{R_{111}} =$$

$$= 0{,}0079 - \frac{12}{21} 0{,}0058 - \frac{9}{21} 0{,}0056 = 0{,}0022$$

Idea



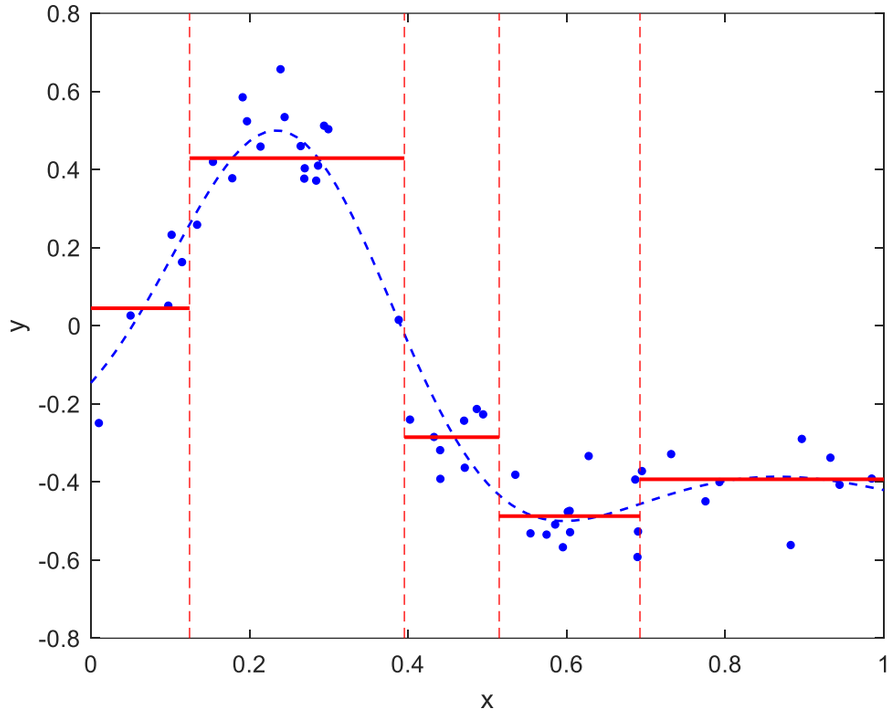
Konstrukcja drzewa

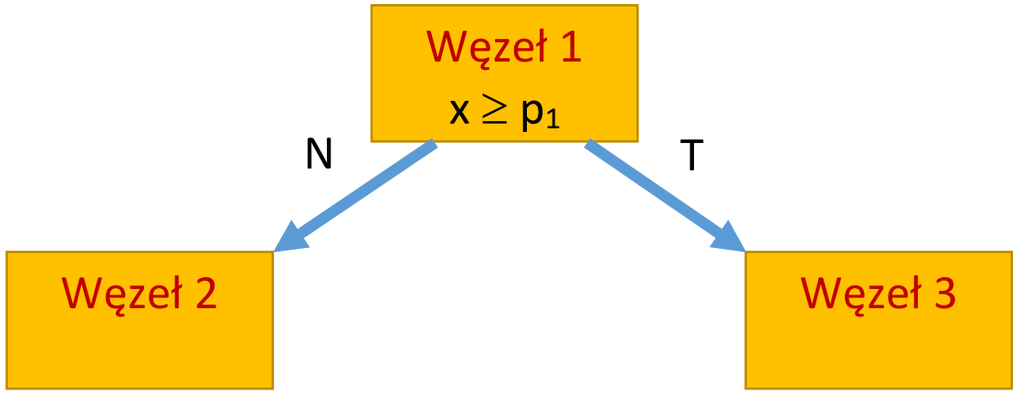


Węzeł 1

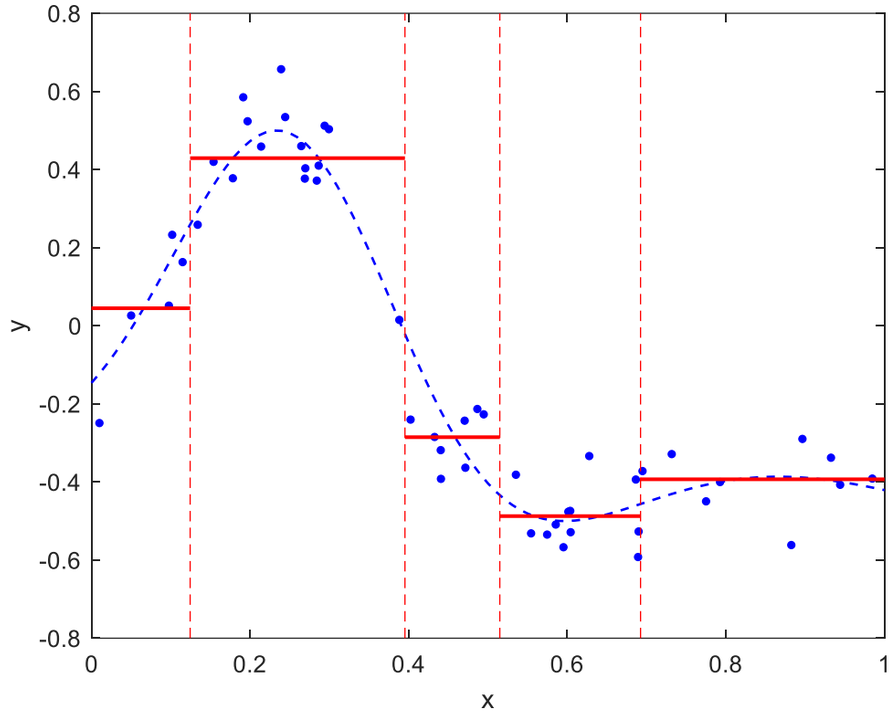
$x \geq p_1$

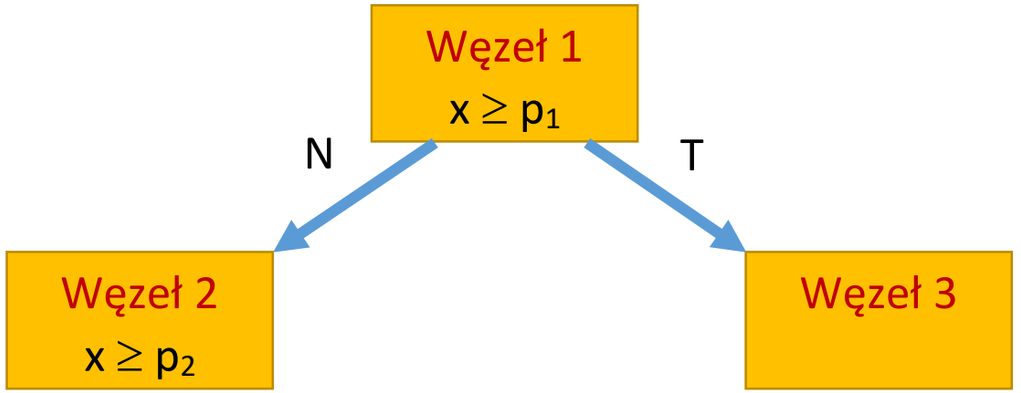
Konstrukcja drzewa



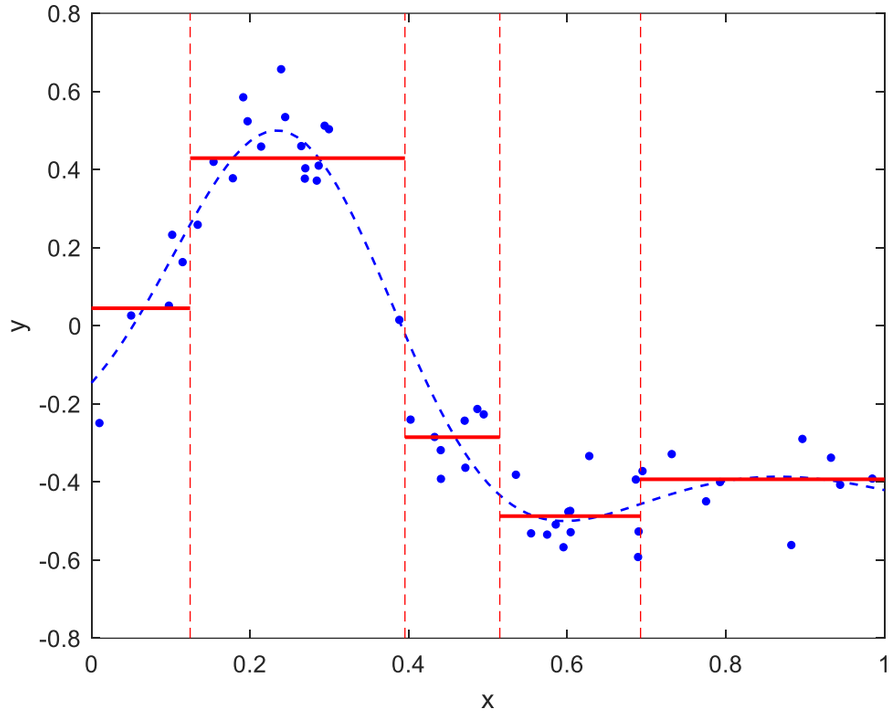


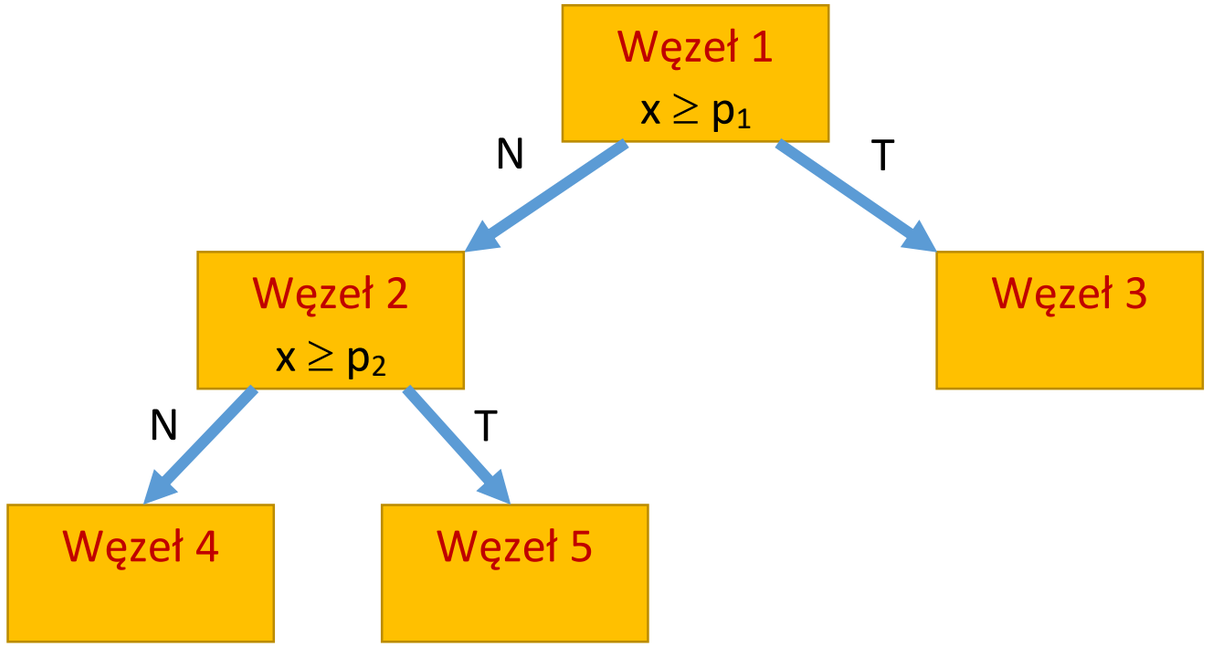
Konstrukcja drzewa



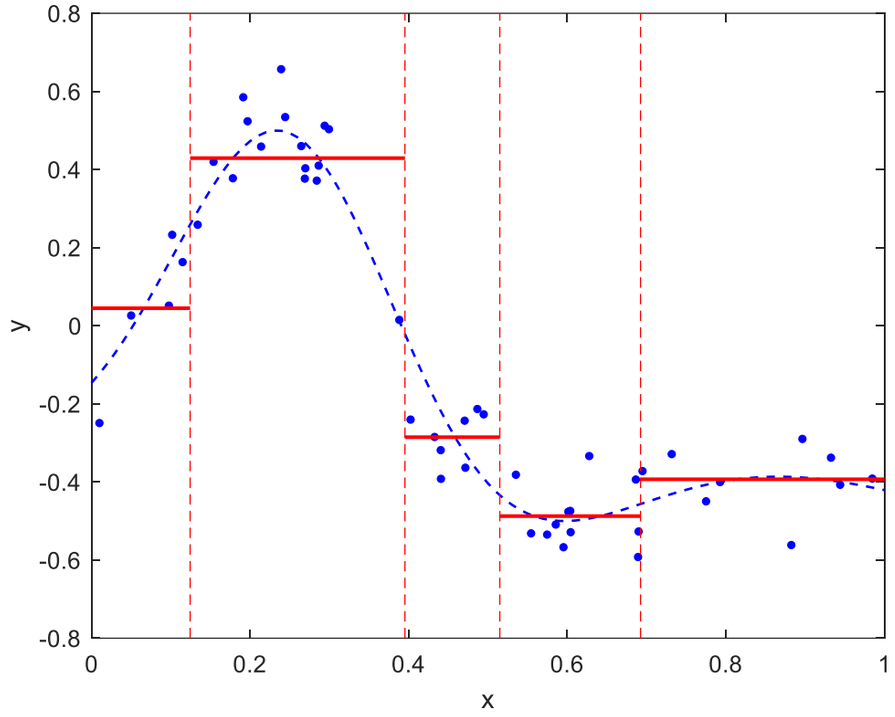


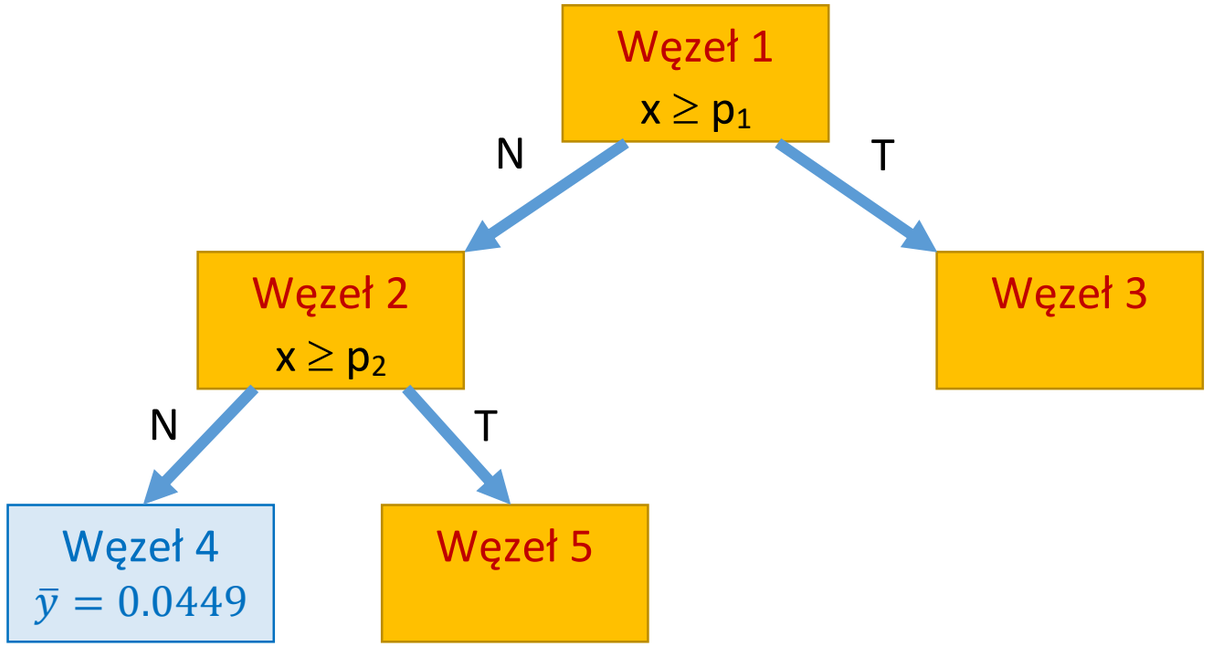
Konstrukcja drzewa



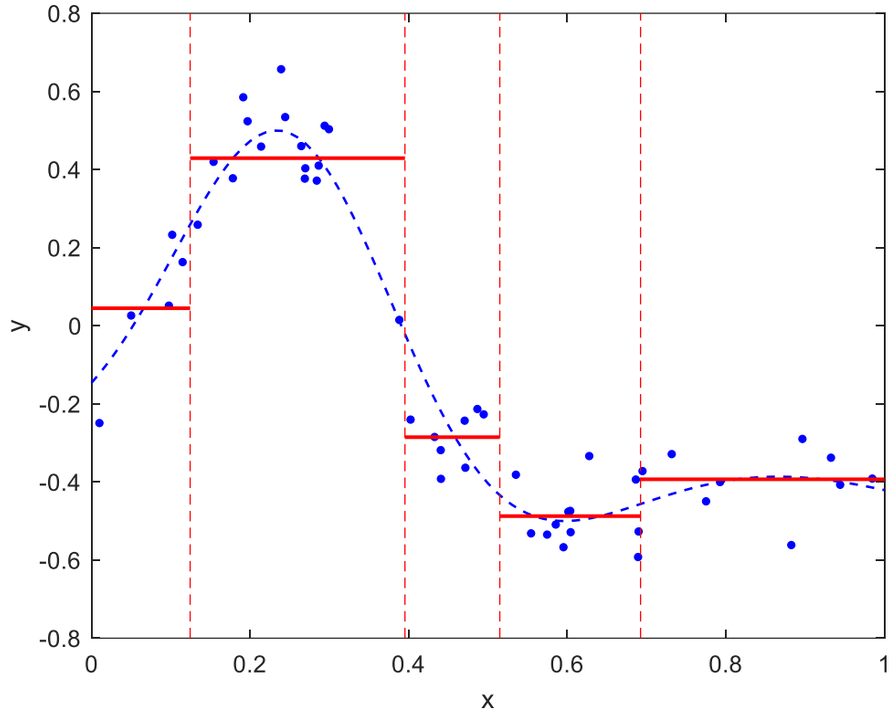


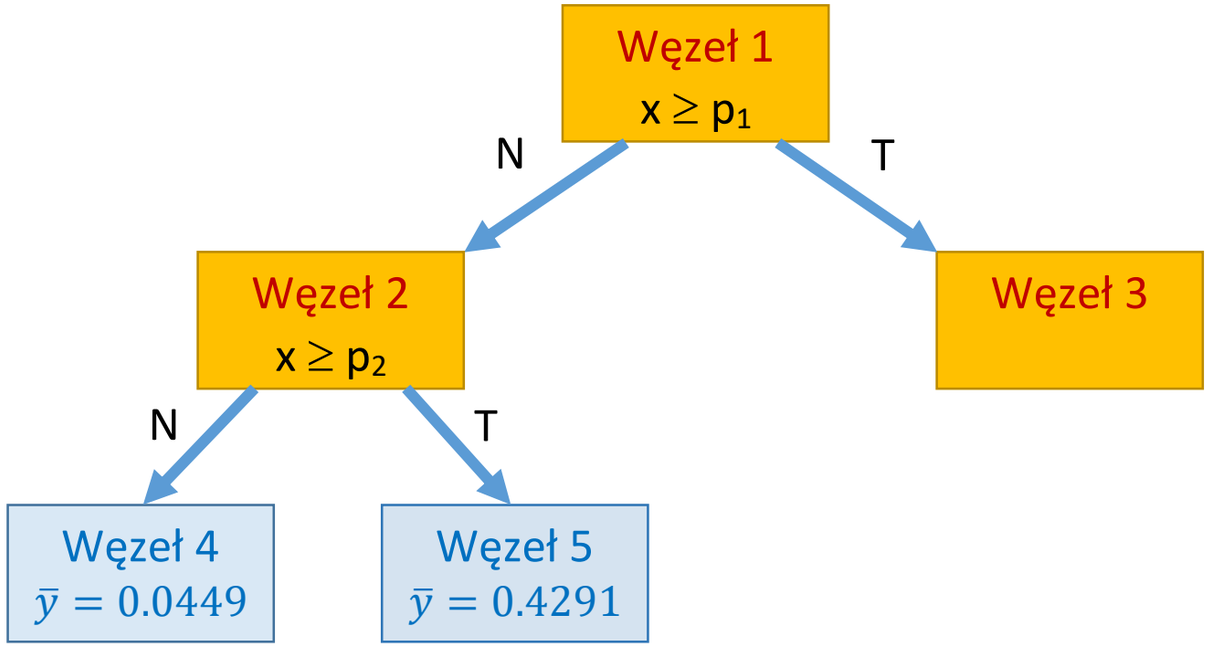
Konstrukcja drzewa



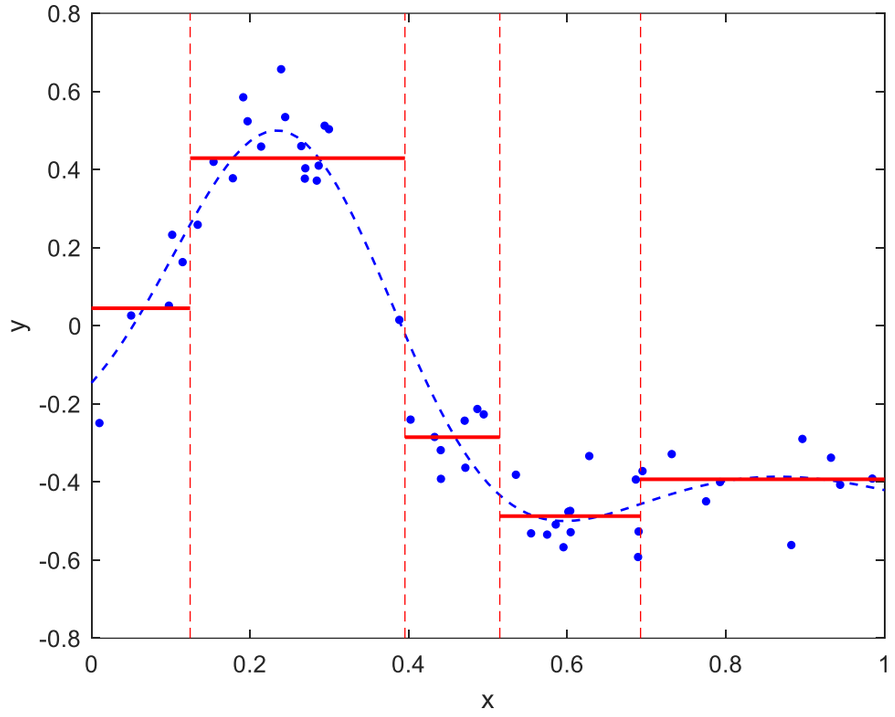


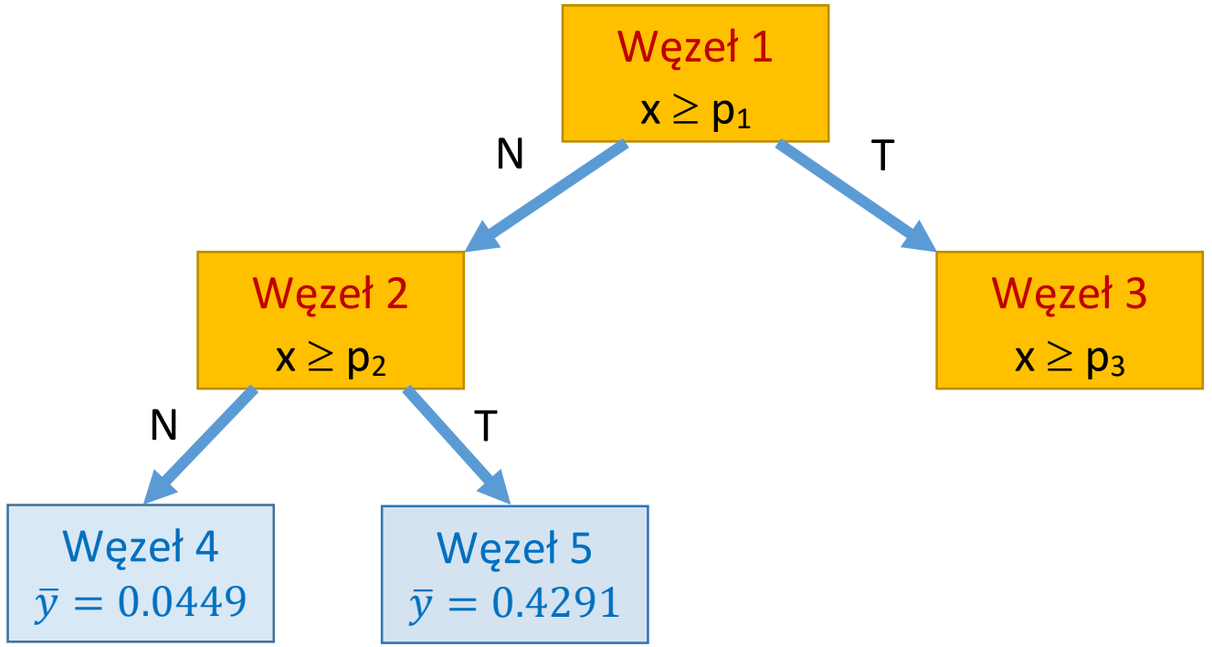
Konstrukcja drzewa



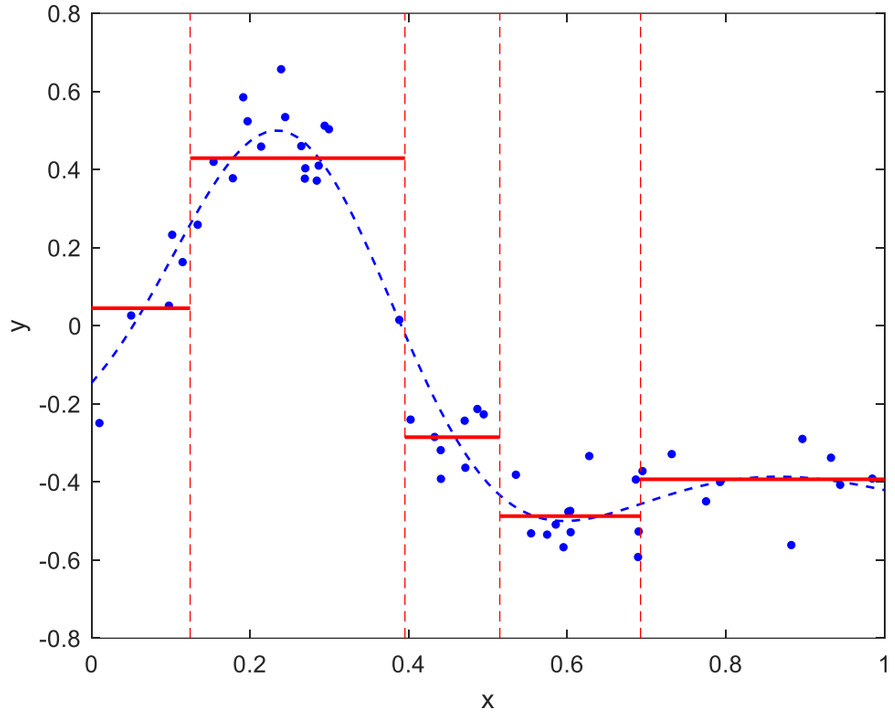


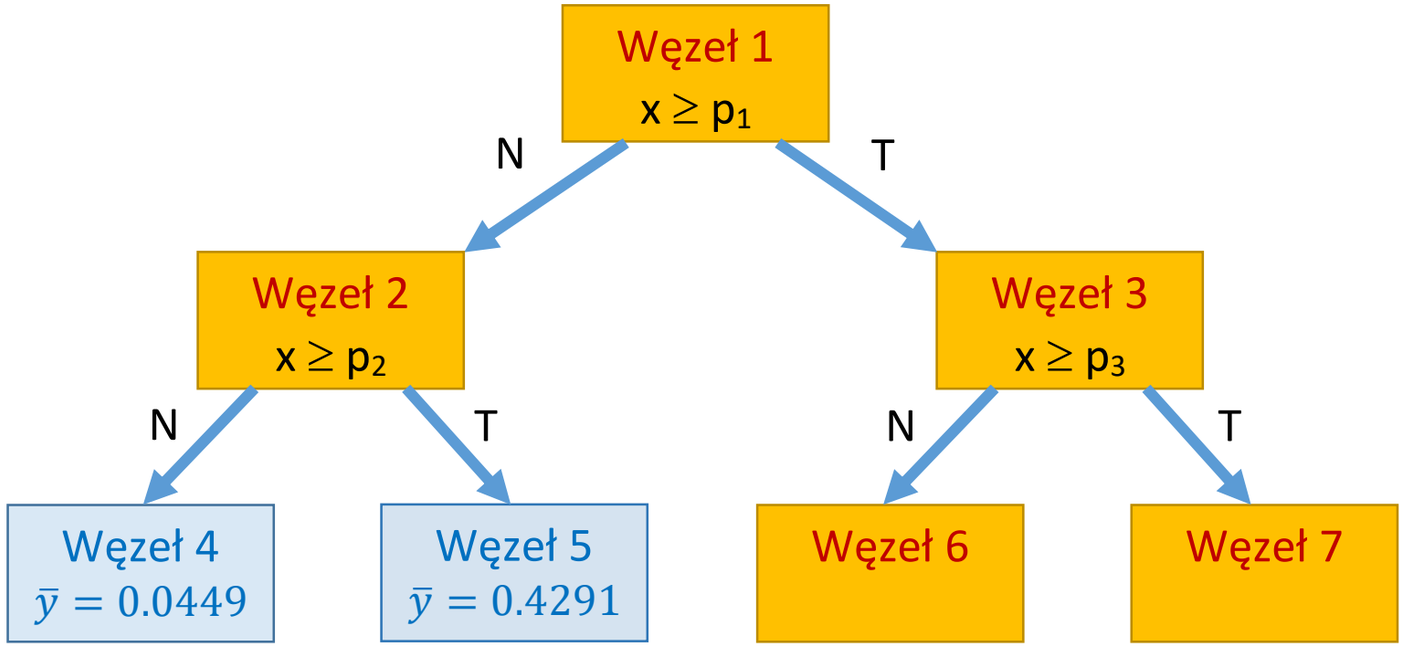
Konstrukcja drzewa



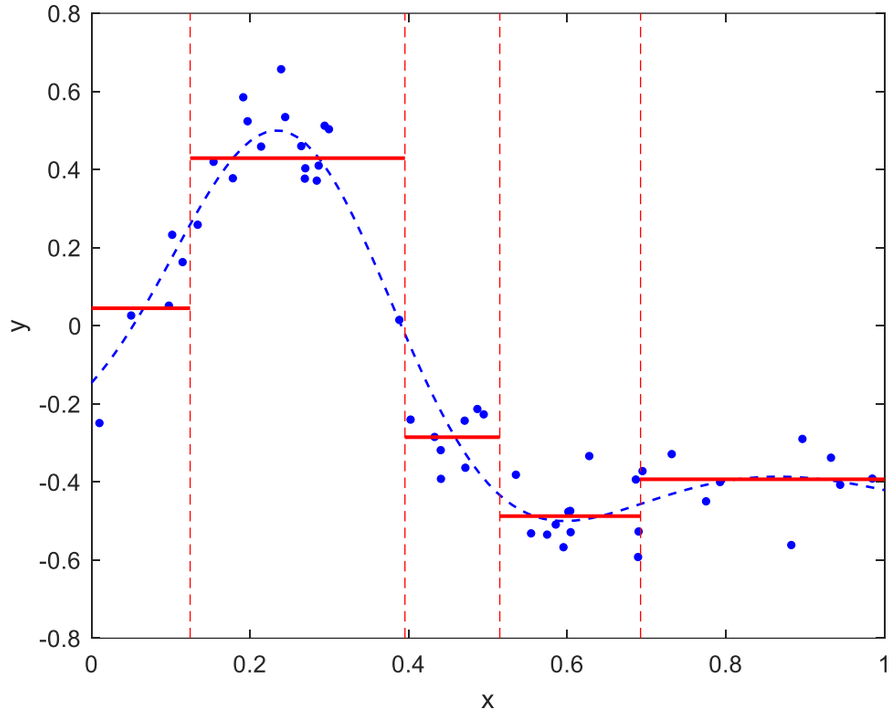


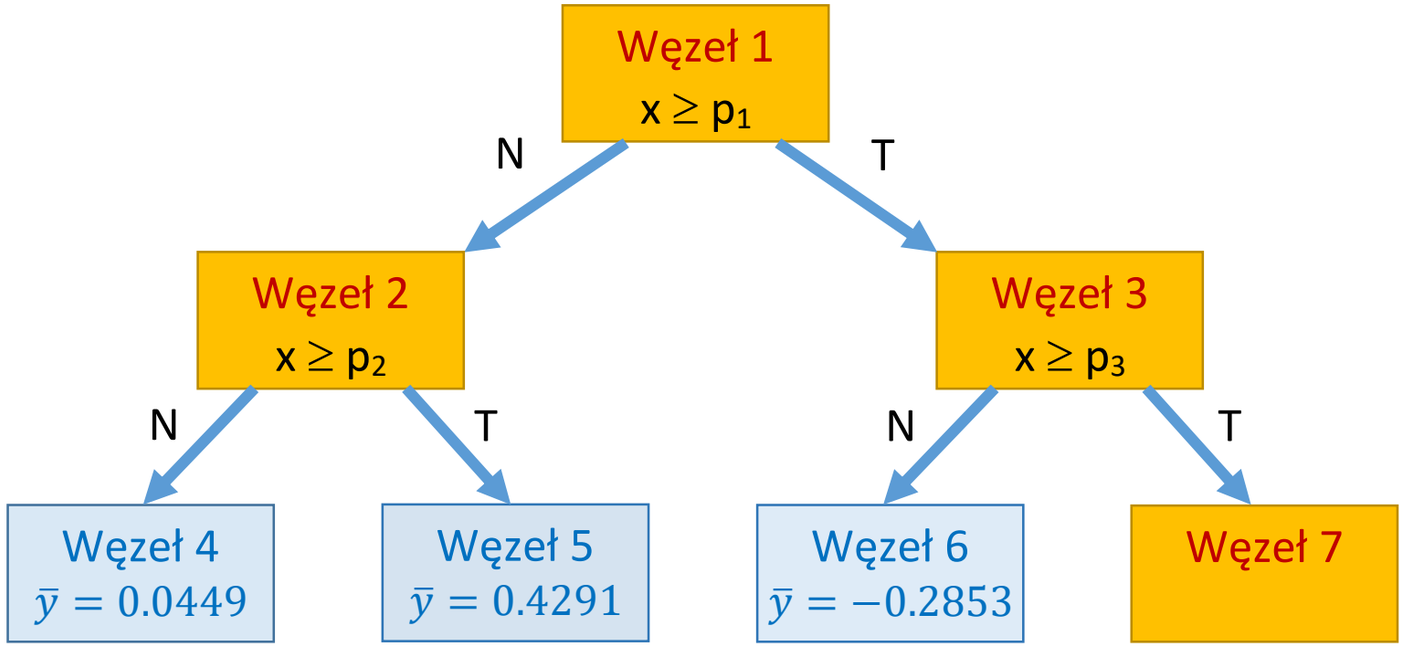
Konstrukcja drzewa



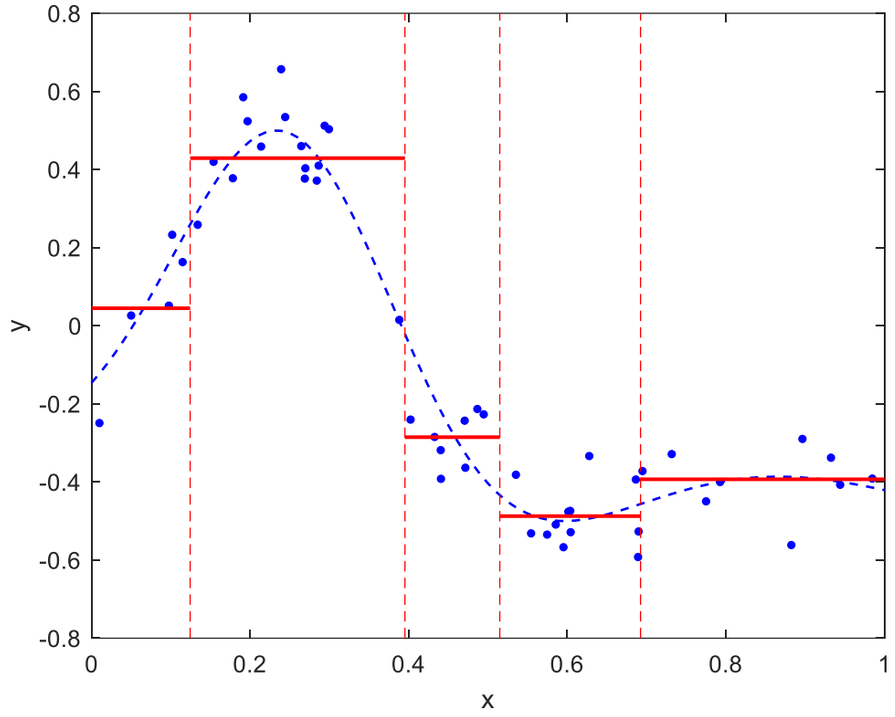


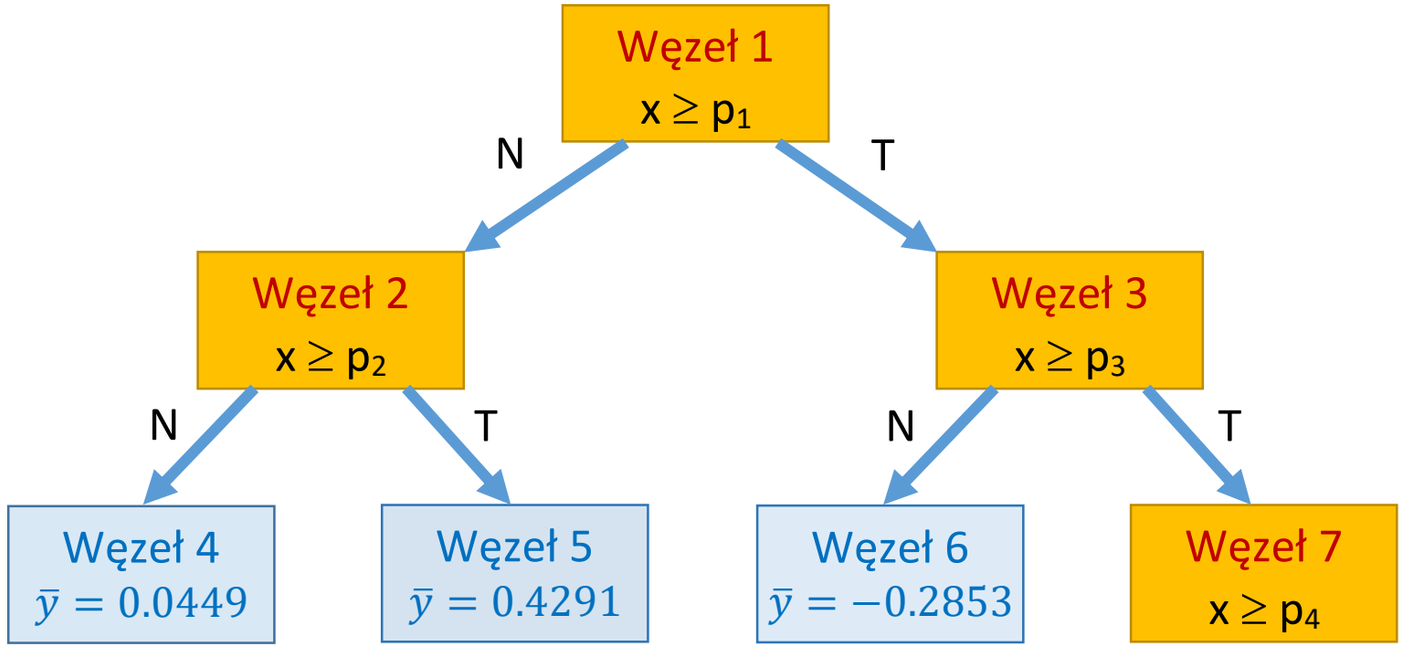
Konstrukcja drzewa



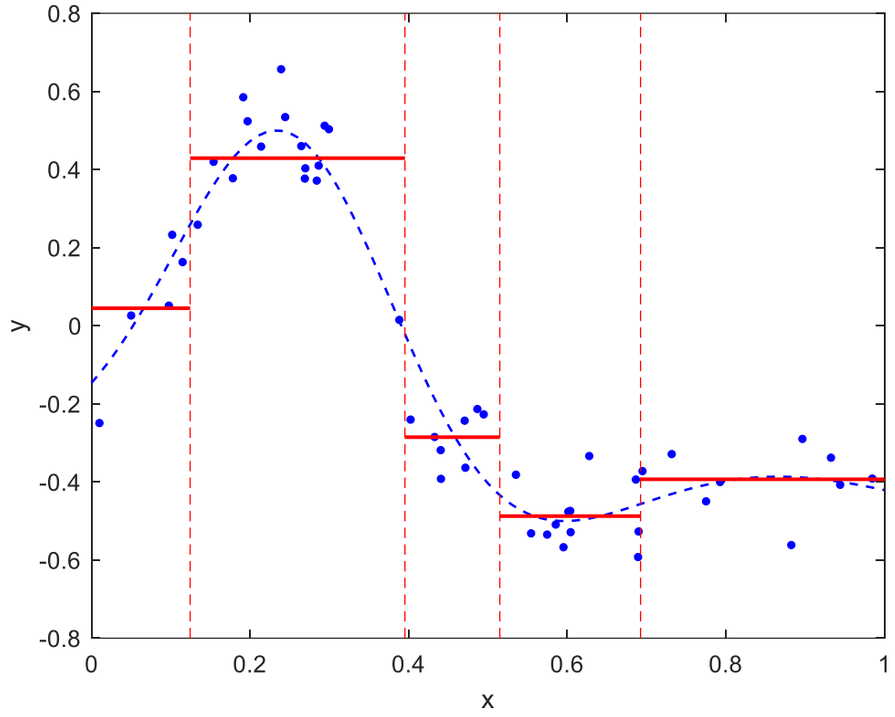


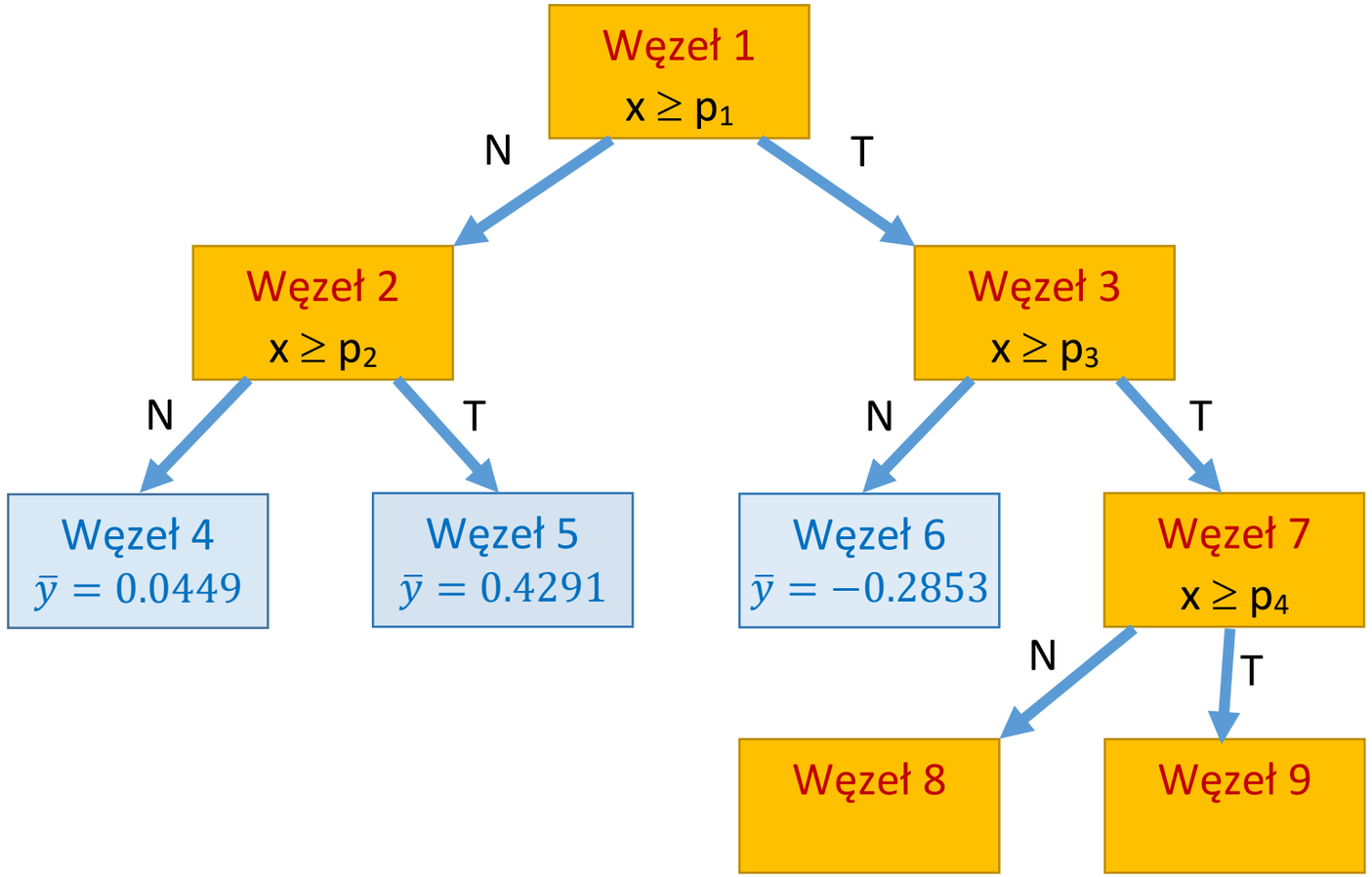
Konstrukcja drzewa



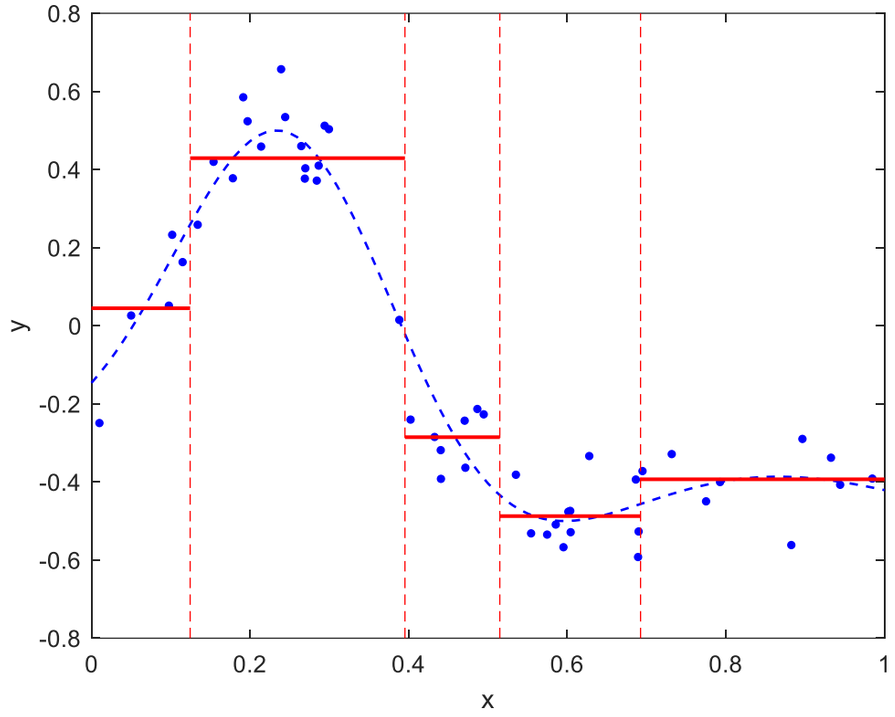


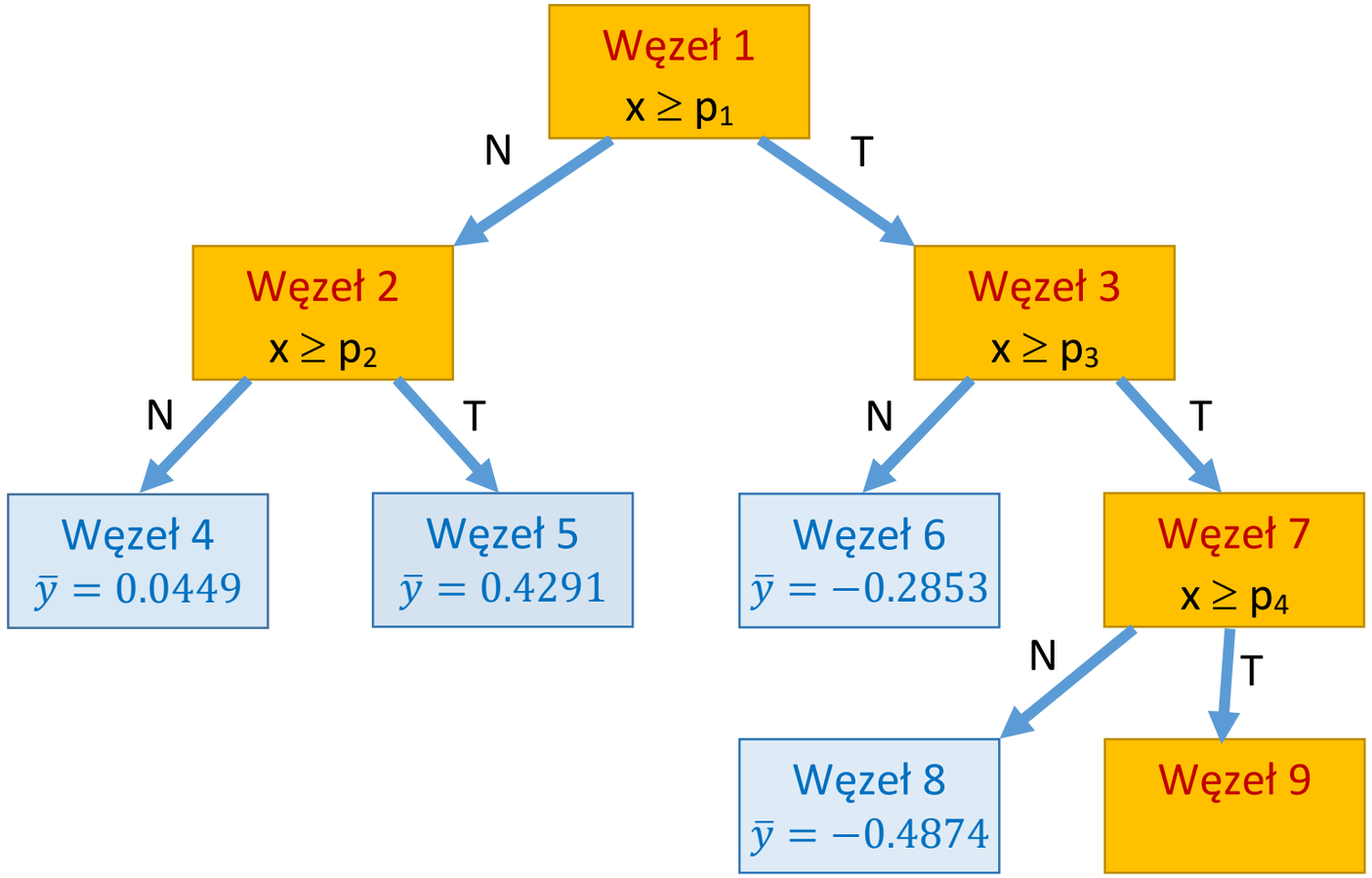
Konstrukcja drzewa



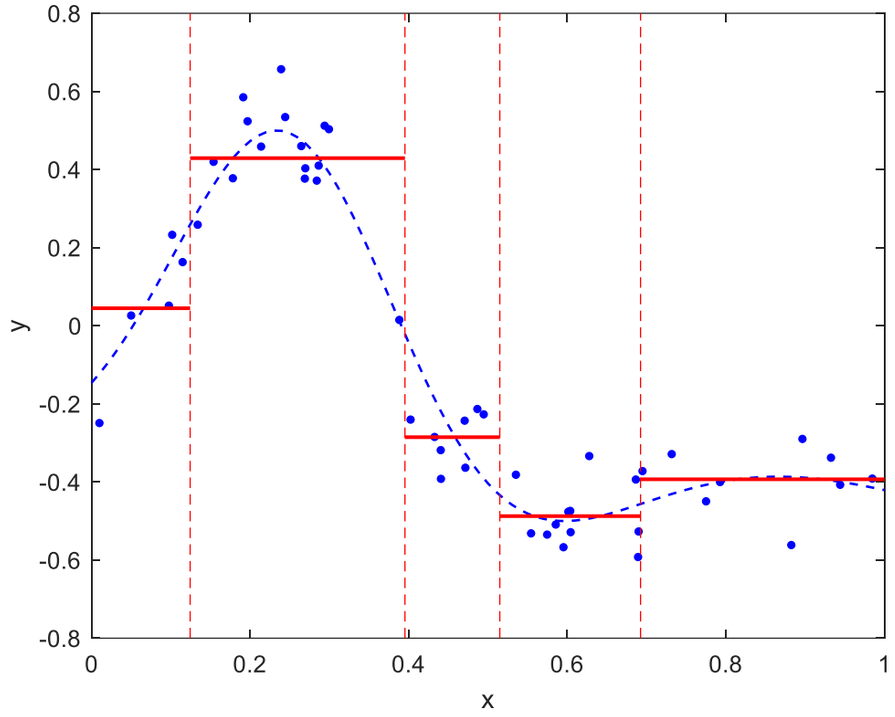


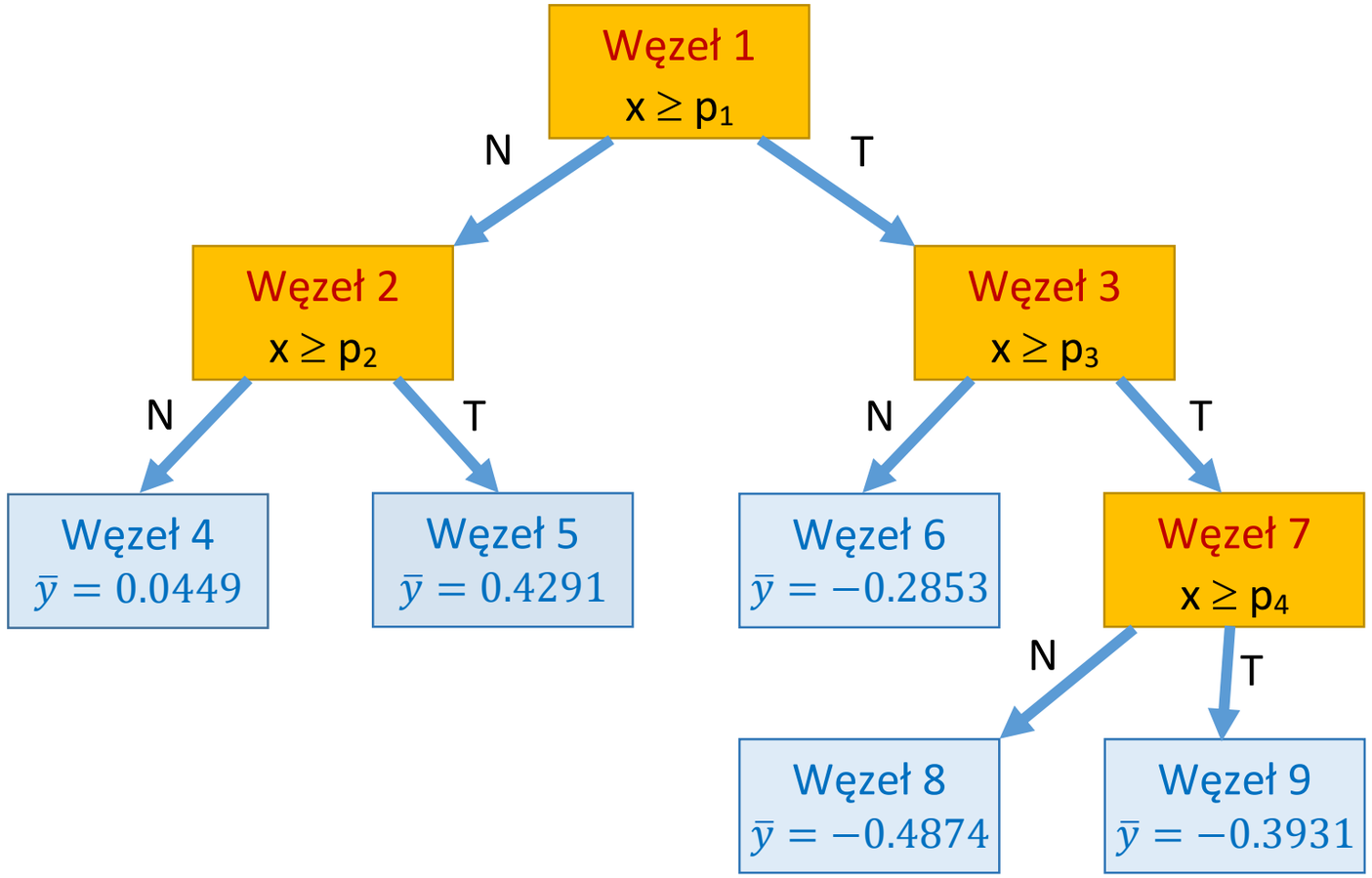
Konstrukcja drzewa





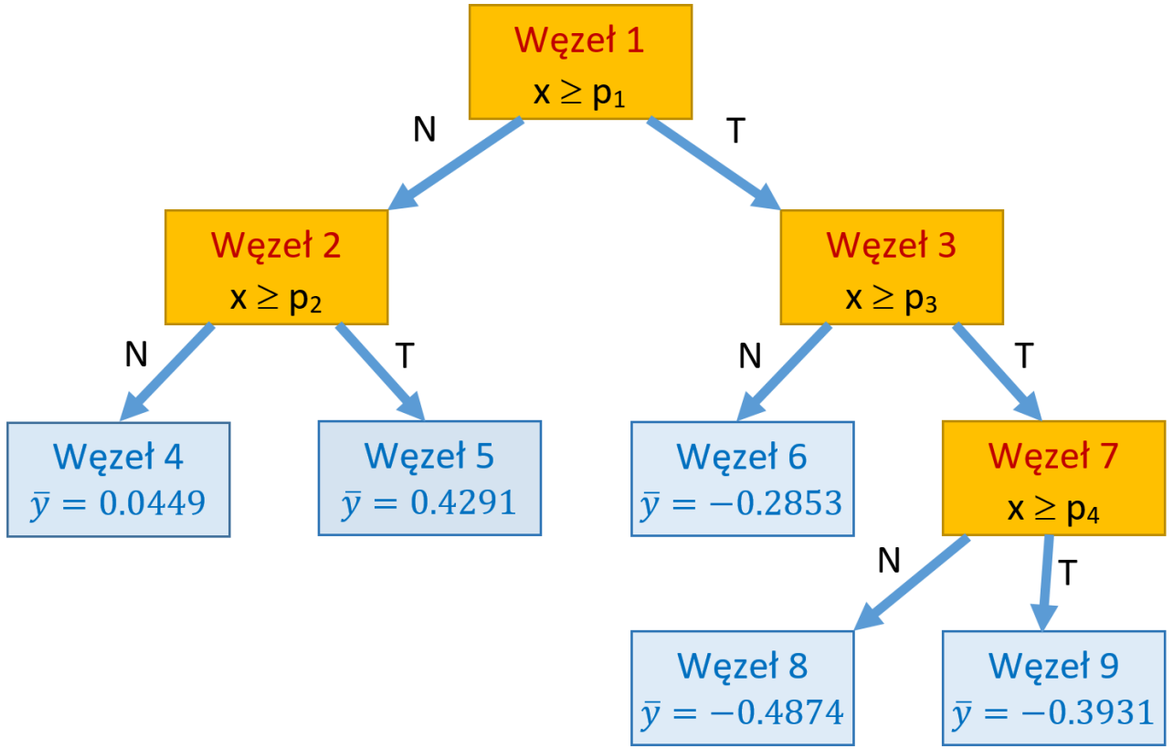
Konstrukcja drzewa





Drzewo regresyjne

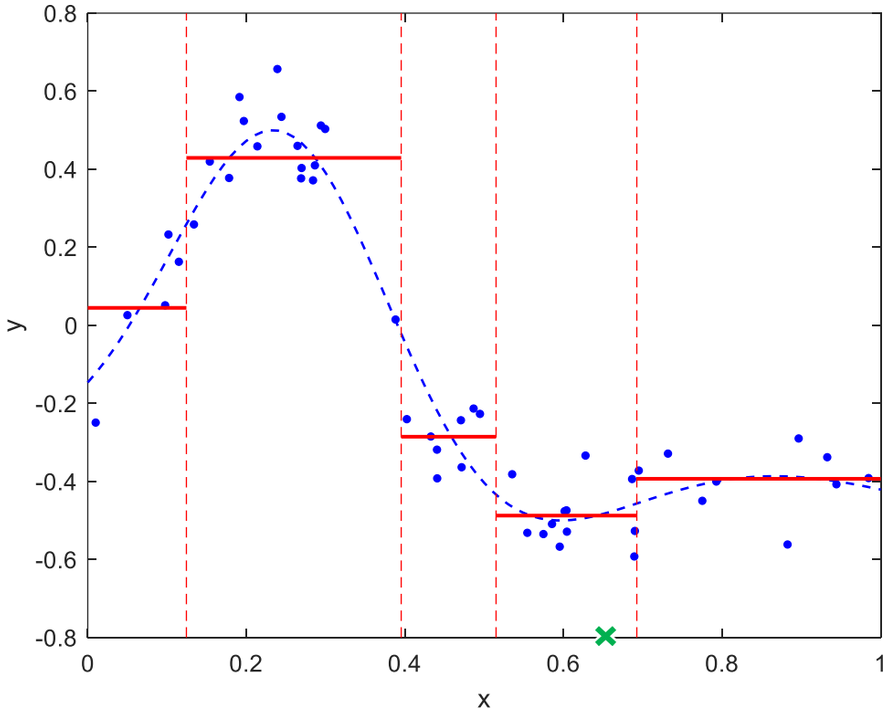
- Drzewo regresyjne jest strukturą złożona z węzłów, gałęzi i liści (węzłów terminalnych)
- Węzły odpowiadają testom przeprowadzanym na wartościach atrybutów, gałęzie odpowiadają wynikom tych testów, a liście – przewidywanym wartościom funkcji
- Drzewo konstruowane jest iteracyjnie. W każdej iteracji zachodzi podział zbioru danych na dwa podzbiory
- Drzewo binarne – z każdego węzła wychodzą tylko dwie gałęzie (test daje dwa wyniki)

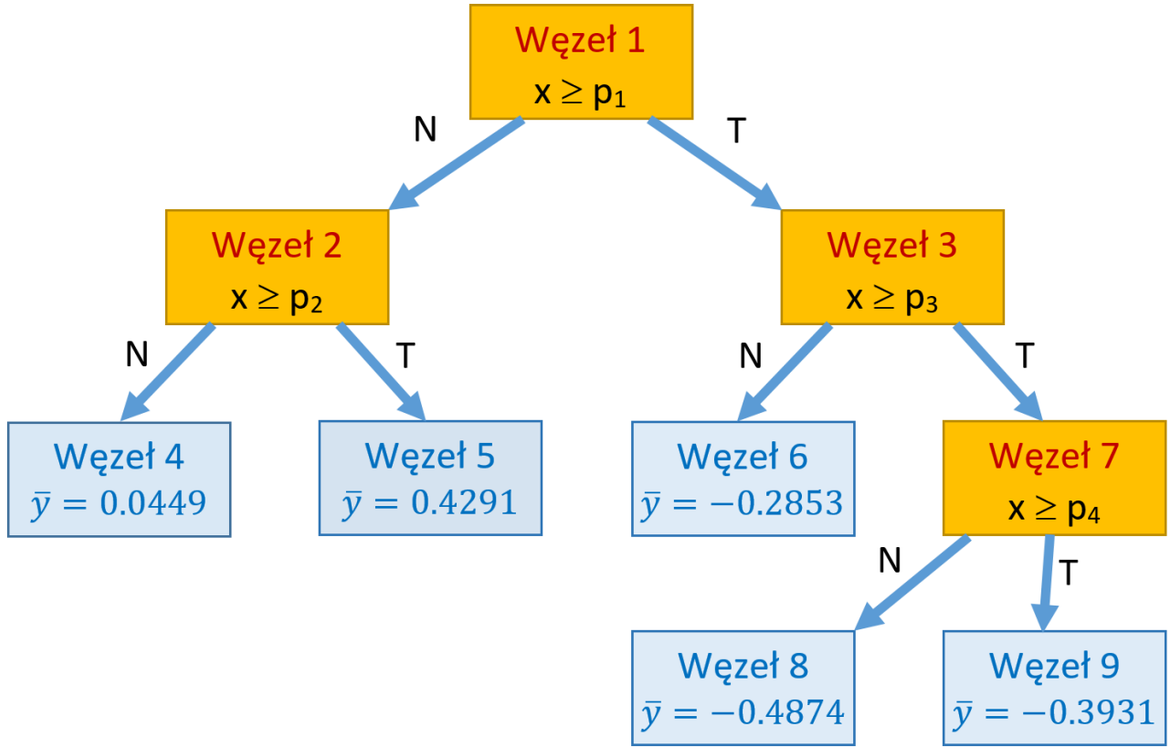


Wyznaczenie wartości funkcji dla nowego przykładu

Wyznaczenie wartości funkcji dla przykładu x polega na przejściu od korzenia do jednego z liści.

W odwiedzanych węzłach wykonujemy związane z nimi testy i poruszamy się po gałęziach odpowiadających uzyskiwanym wynikom.





Rodzaje i liczba testów

Test dla atrybutów:
N – nominalnych,
P – porządkowych,
C – ciągłych

- testy tożsamościowe – sprawdzamy wartość atrybutu; wynikiem testu jest wartość atrybutu, np. dla atrybutu *wiatr* wynikiem testu może być *słaby* lub *silny* (N, P)

  Liczba możliwych testów – tyle ile jest atrybutów

- testy równościowe (N, P)

  $$test(x) = \begin{cases} 1 & \text{jeśli } a(x) = v \\ 0 & \text{jeśli } a(x) \neq v \end{cases},$$ gdzie *a* jest atrybutem przykładu *x*, *v* jest jedną z możliwych wartości atrybutu

  Liczba możliwych testów – możliwych testów dla jednego atrybutu jest tyle ile ten atrybut przyjmuje wartości

- testy przynależnościowe (N, P, C)

  $$test(x) = \begin{cases} 1 & \text{jeśli } a(x) \in V \\ 0 & \text{jeśli } a(x) \notin V \end{cases},$$ gdzie *V* jest podzbiorem wartości atrybutu

  Liczba możliwych testów – możliwych testów dla jednego atrybutu jest tyle ile podzbiorów *V* zdefiniujemy

Rodzaje i liczba testów

Test dla atrybutów:
N – nominalnych,
P – porządkowych,
C – ciągłych

- testy podziałowe (N, P, C)

$$test(x) = \begin{cases} 1 & \text{jeśli } a(x) \in V_1 \\ 2 & \text{jeśli } a(x) \in V_2 \\ \dots & \\ m & \text{jeśli } a(x) \in V_m \end{cases}$$

, gdzie parami rozłączne zbiory V_1, V_2, …, V_m stanowią wyczerpujący podział przeciwdziedziny atrybutu

Liczba możliwych testów – możliwych testów dla jednego atrybutu jest tyle ile podzbiorów *V* zdefiniujemy

- testy nierównościowe (P, C)

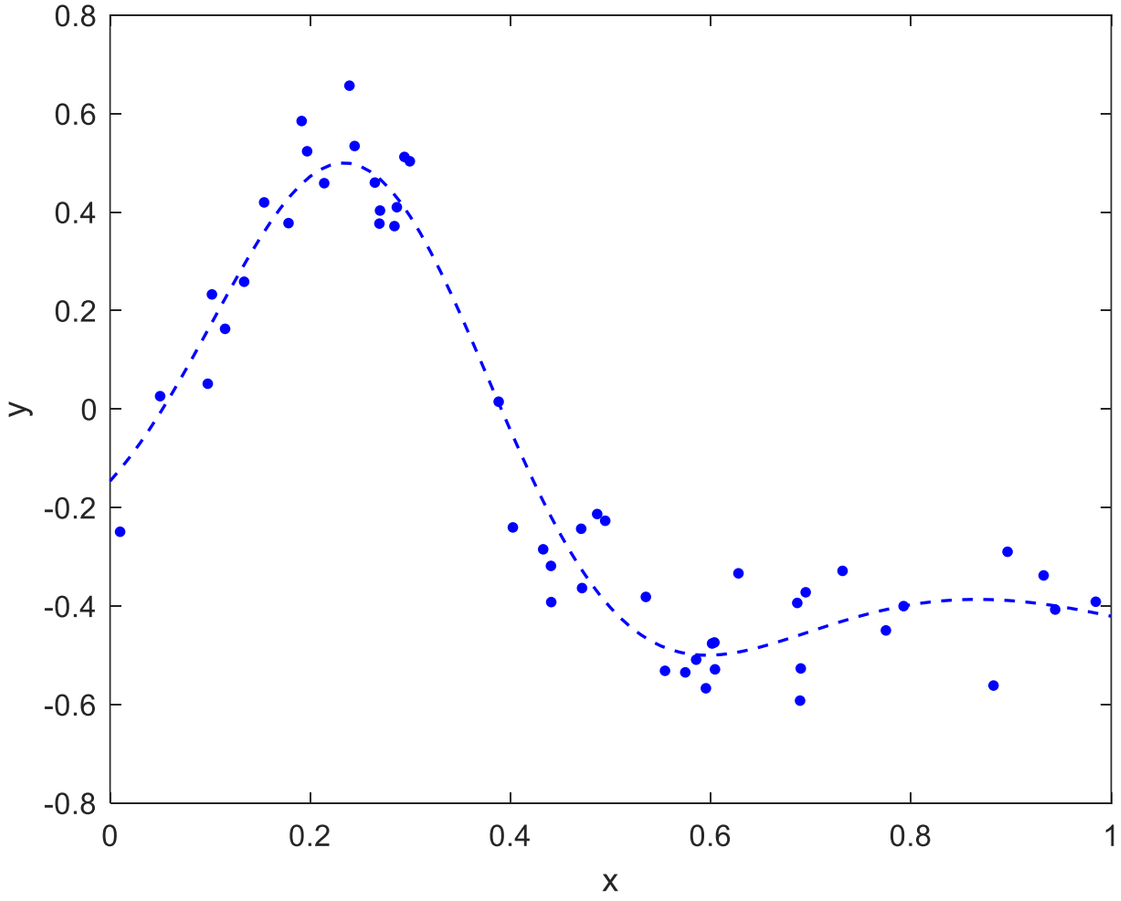
$$test(x) = \begin{cases} 1 & \text{jeśli } a(x) \geq p \\ 0 & \text{jeśli } a(x) < p \end{cases}$$

, gdzie *p* jest wartością progową z przeciwdziedziny atrybutu

Liczba możliwych testów – możliwych testów dla jednego atrybutu jest tyle ile przyjmuje on różnych wartości w zbiorze uczącym minus 1

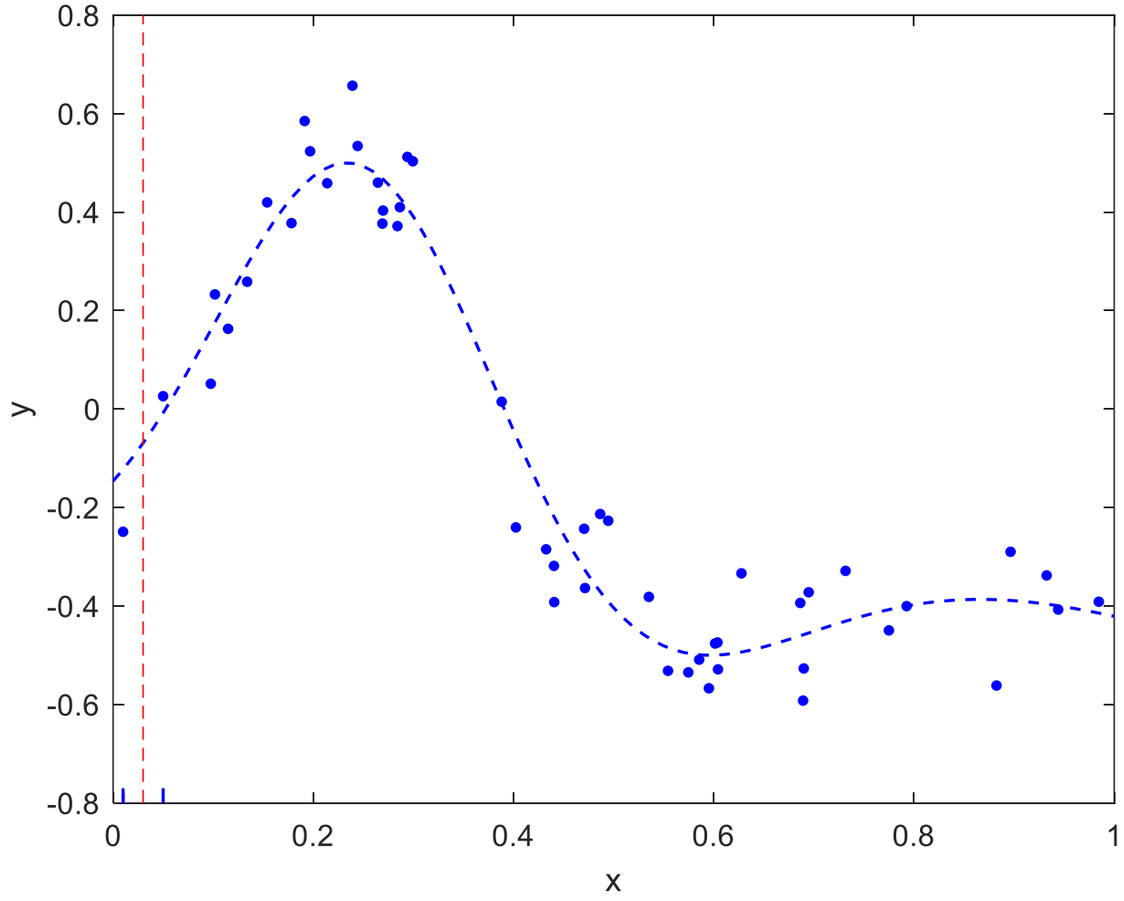
Rodzaje i liczba testów

Liczba możliwych testów nierównościowych



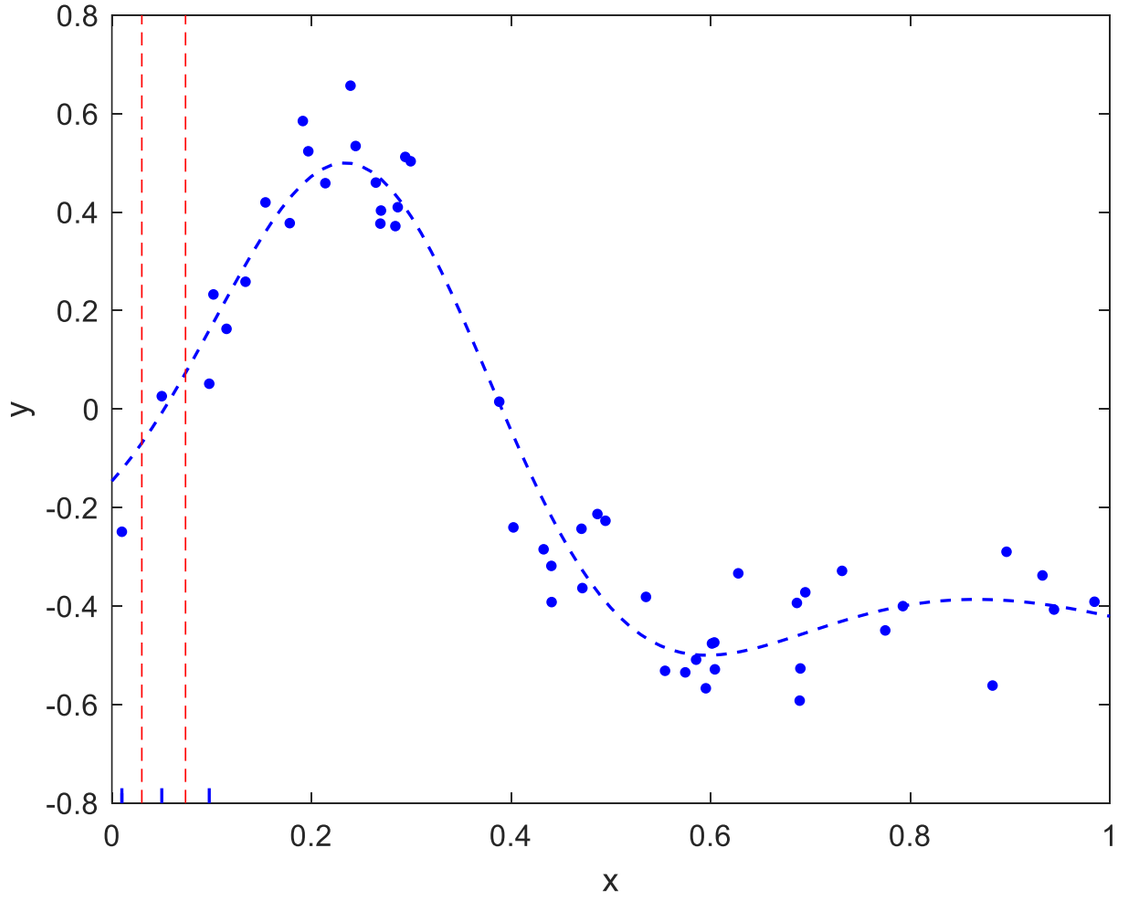
Rodzaje i liczba testów

Liczba możliwych testów nierównościowych



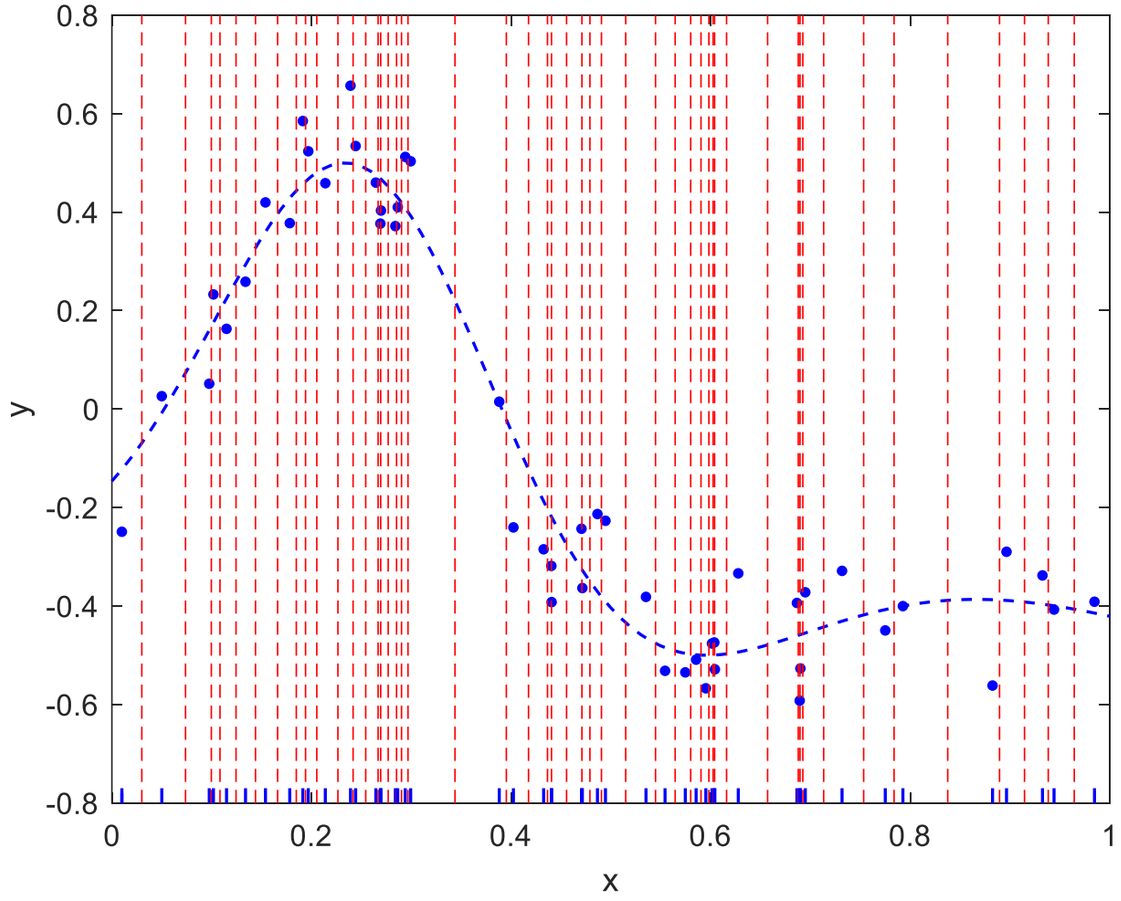
Rodzaje i liczba testów

Liczba możliwych testów nierównościowych



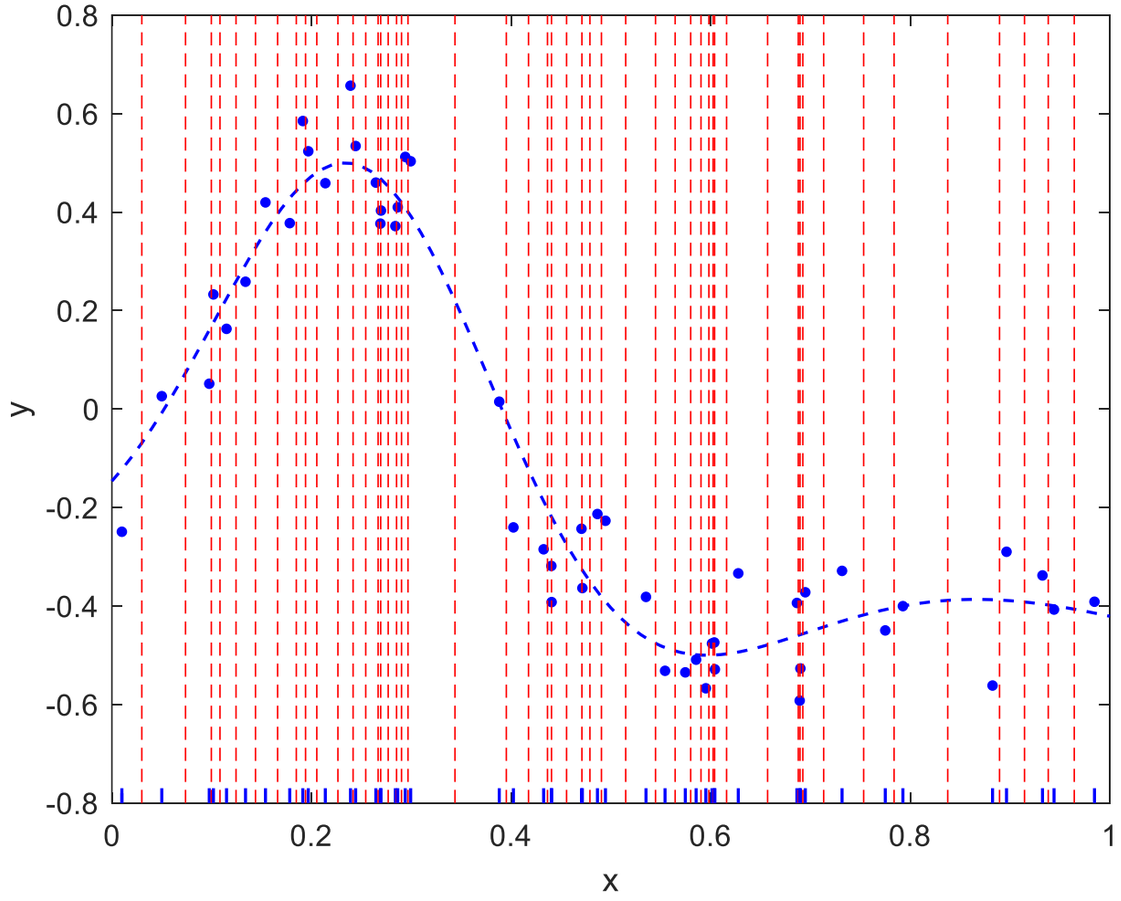
Rodzaje i liczba testów

Liczba możliwych testów nierównościowych



Rodzaje i liczba testów

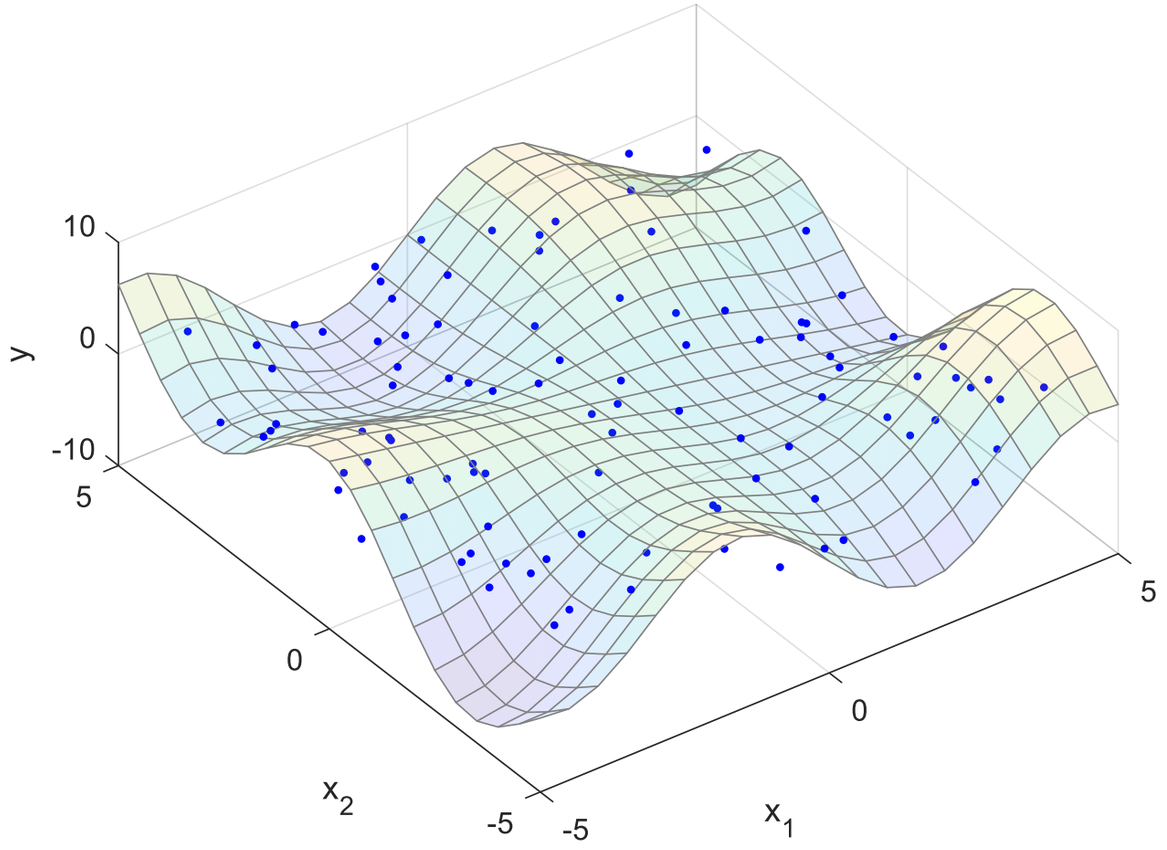
Liczba możliwych testów nierównościowych



Liczba testów dla x: $N - 1$

Rodzaje i liczba testów

Liczba możliwych testów nierównościowych – przypadek funkcji dwóch zmiennych

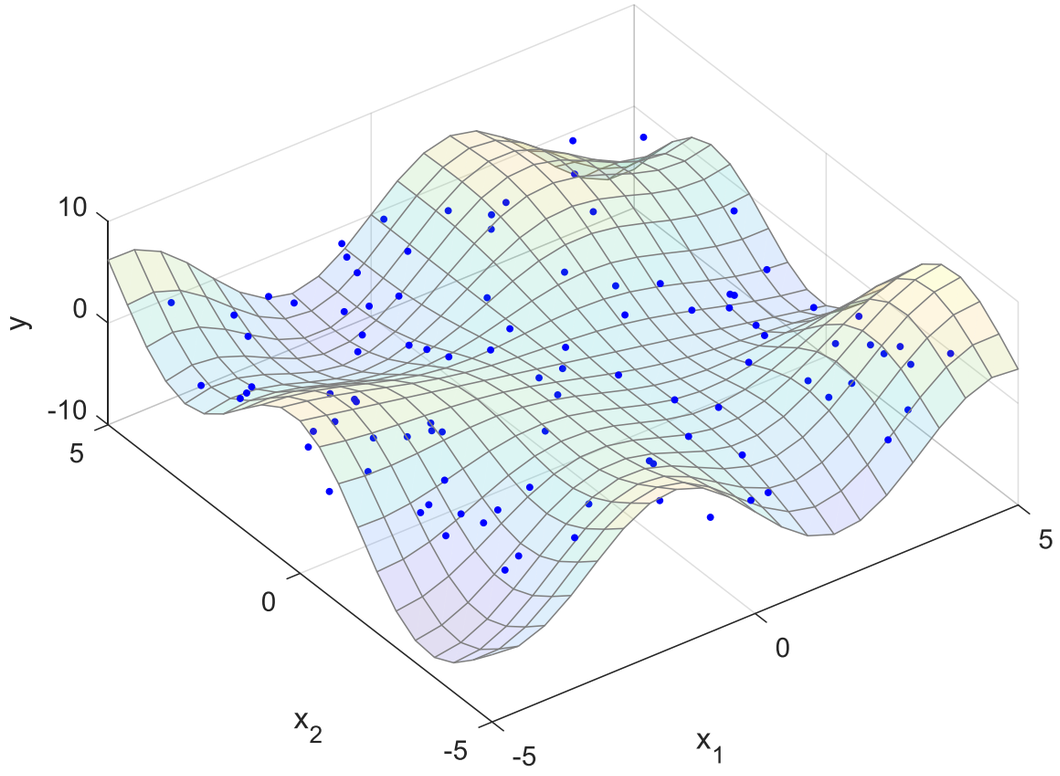


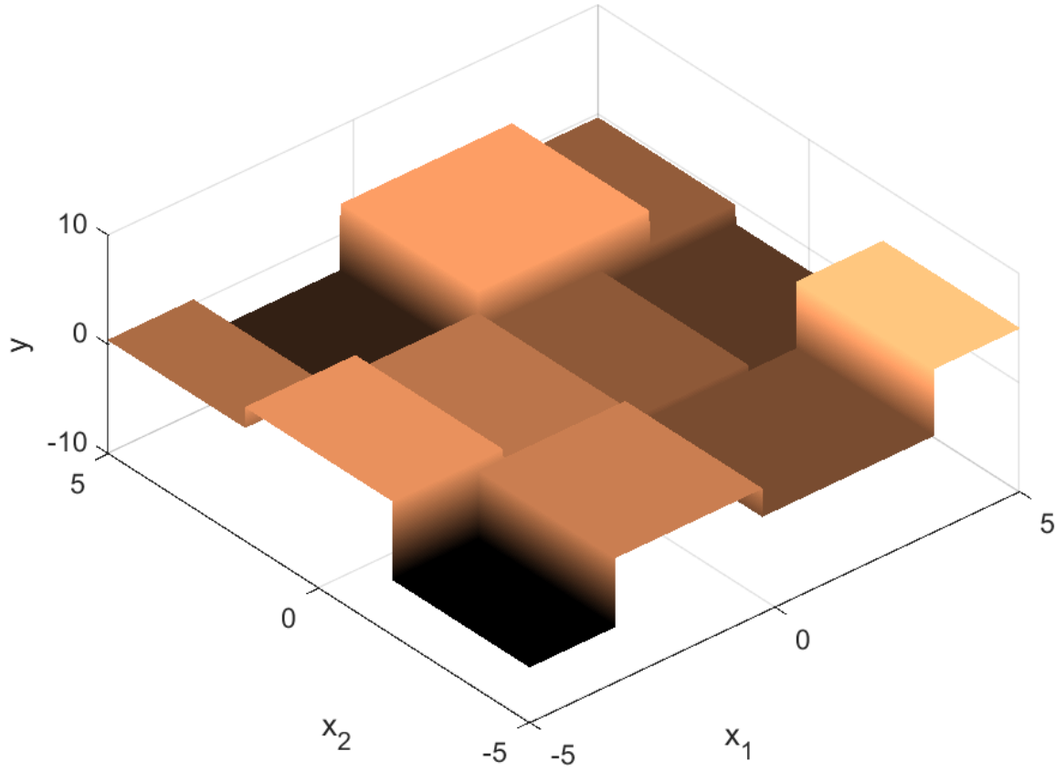
Liczba testów dla n atrybutów $x_1, x_2, \ldots, x_n$:

$(N - 1)\, n$

Rodzaje i liczba testów

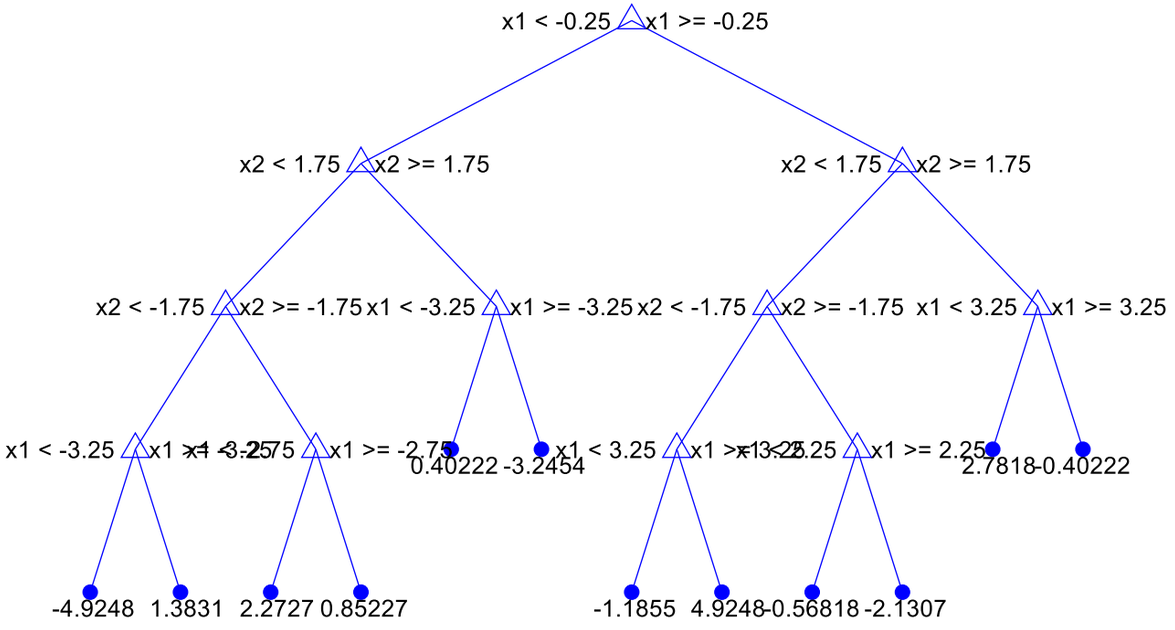
Liczba możliwych testów nierównościowych – przypadek funkcji dwóch zmiennych

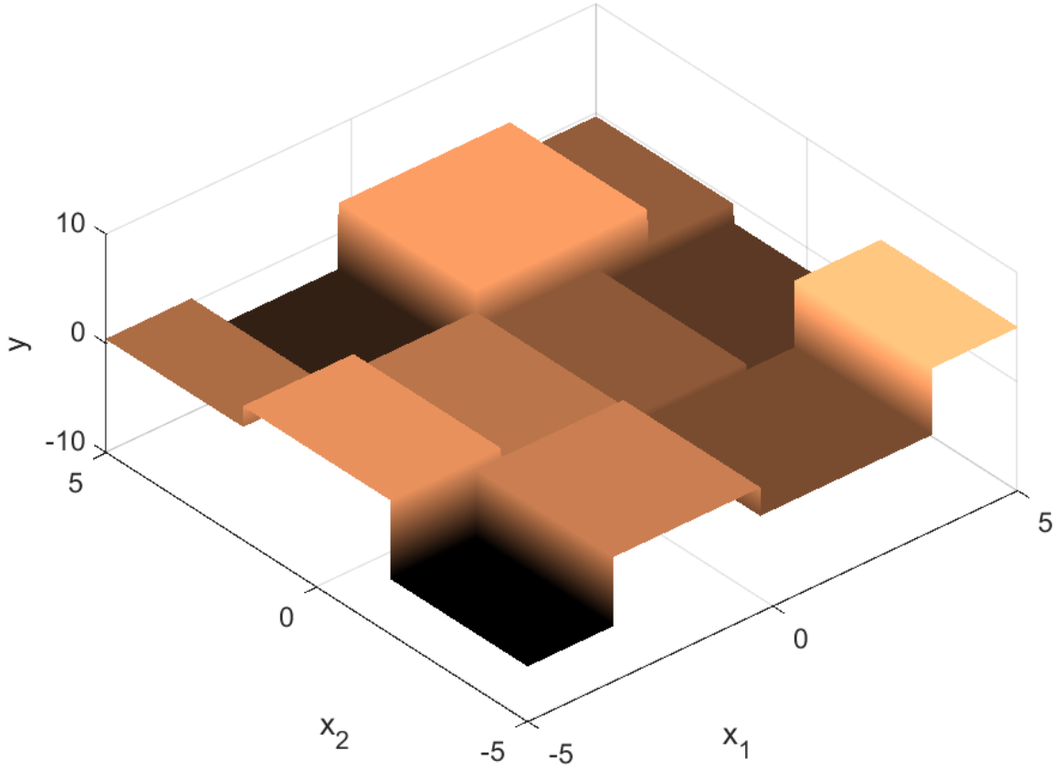




Rodzaje i liczba testów

Liczba możliwych testów nierównościowych – przypadek funkcji dwóch zmiennych





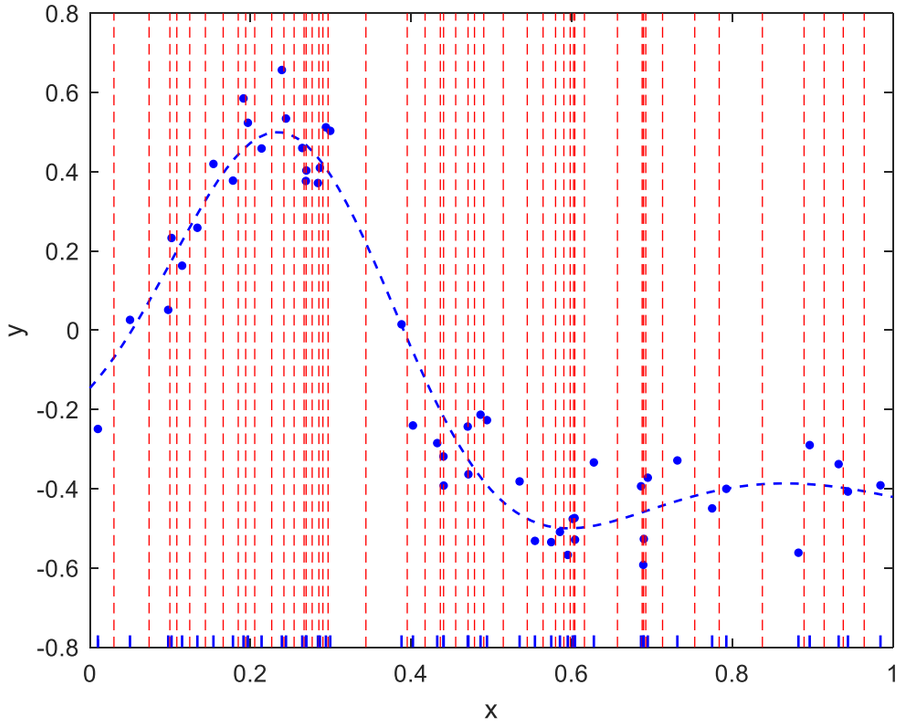
Kryterium wyboru testu

Test powinien dzielić zbiór przykładów tak, aby uzyskać maksymalną redukcję zróżnicowania danych po podziale:

$$\max_{x_i, p_i} \Delta I$$

$$\Delta I = \varepsilon_R - \frac{N_{R_0}}{N_R}\varepsilon_{R_0} - \frac{N_{R_1}}{N_R}\varepsilon_{R_1}$$

$$\varepsilon_R = \frac{1}{N_R}\sum_{i \in R}(y_i - \bar{y}_R)^2$$



Zstępujący schemat konstrukcji drzew decyzyjnych

- Zaczynając od korzenia (poziom 0) przemieszczamy się do kolejnych poziomów podejmując decyzje o utworzeniu liścia bądź węzła.
- Jeśli kryterium stopu jest spełnione tworzony jest liść i na podstawie podzbioru przykładów, które do niego dotarły ustalana jest jego etykieta.
- W przeciwnym przypadku tworzony jest węzeł i wybierany jest dla niego test.

Kryterium stopu i ustalenie etykiety liścia

Liść tworzony jest w przypadkach gdy:

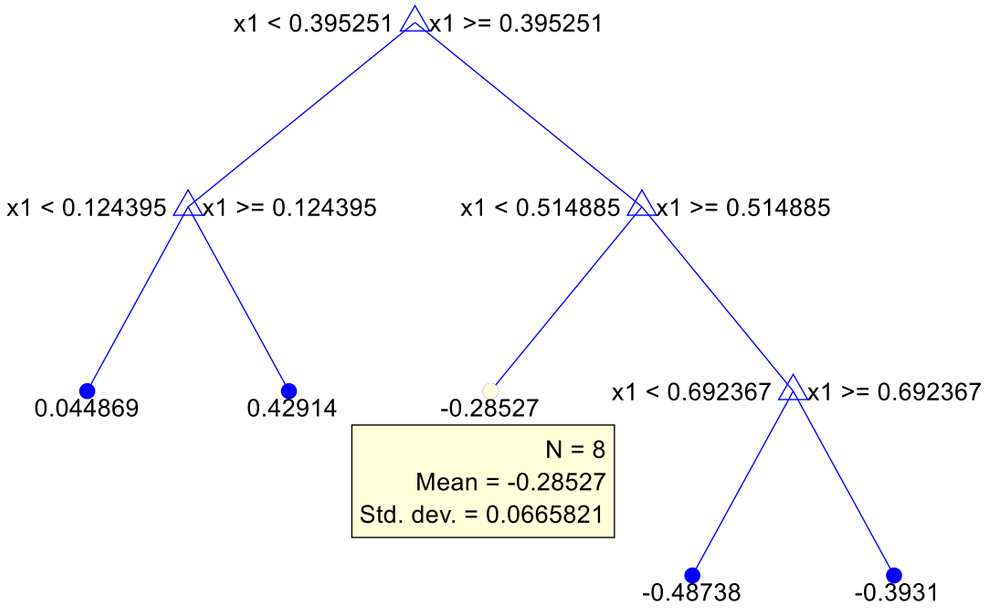
- aktualny zbiór przykładów zawiera przykłady o niewielkim zróżnicowaniu ($\varepsilon < \varepsilon_{max}$) lub
- nie istnieje taki podział przykładów w węźle, który prowadzi do zmniejszenia zróżnicowania lub
- liczba przykładów w węźle jest nie większa niż N_{max}

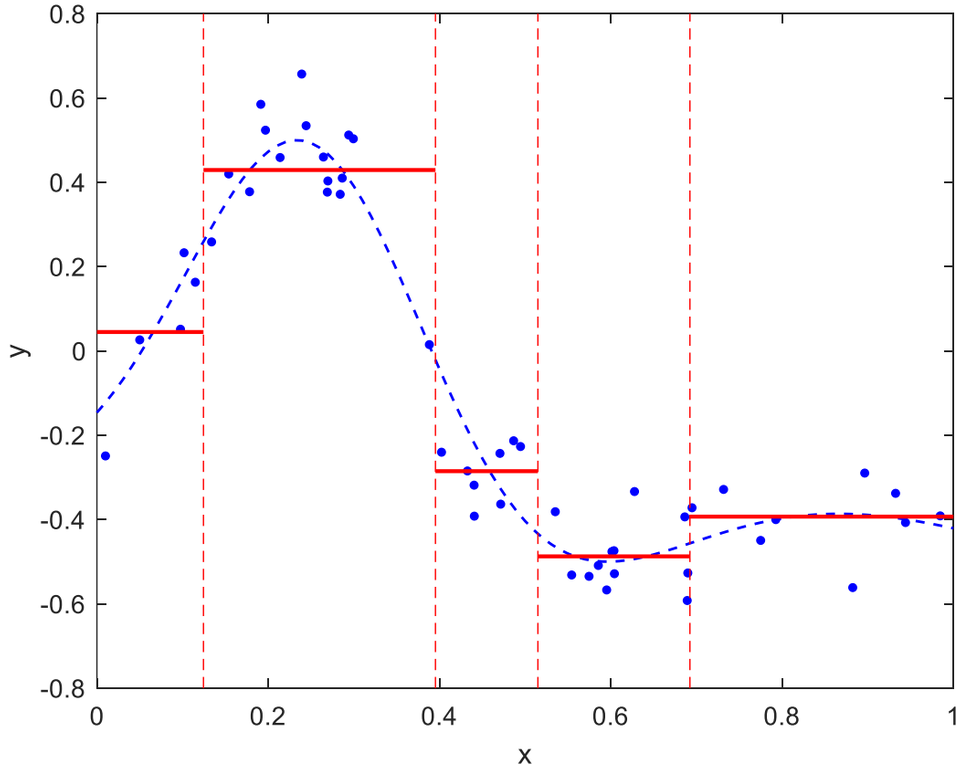
Do etykiety liścia wpisujemy średnią wartość y-ków przykładów uczących, które dotarły do tego liścia.

Dodatkowo, dla celów informacyjnych, możemy wpisać liczbę przykładów, ich zróżnicowanie (wyrażone np. jako odchylenie standardowe) i granice przedziału x reprezentowanego przez liść.

Kryterium stopu i ustalenie etykiety liścia

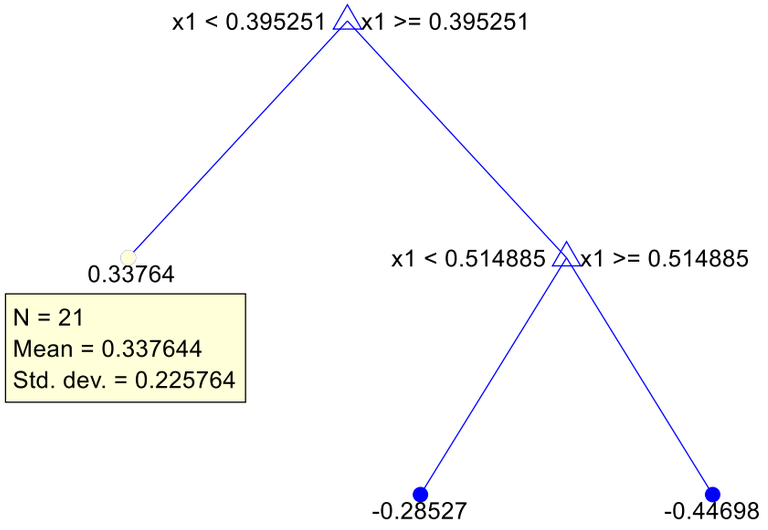
W naszym przykładzie kryterium stopu miało postać: liczba przykładów w węźle jest nie większa niż $N_{max} = 18$.

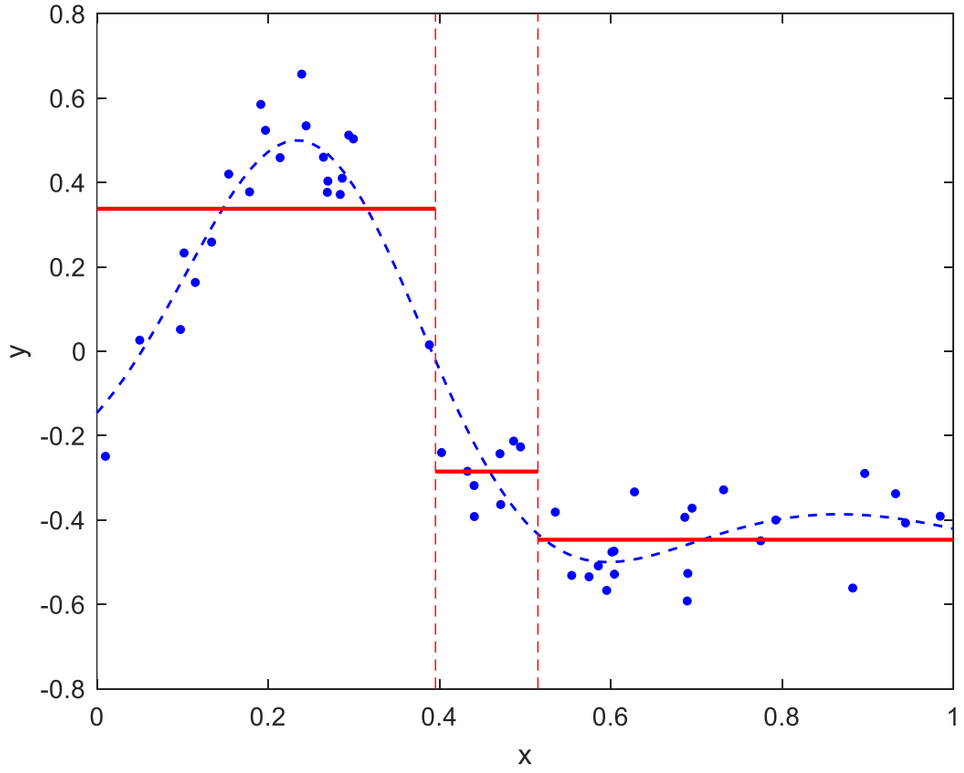




Kryterium stopu i ustalenie etykiety liścia

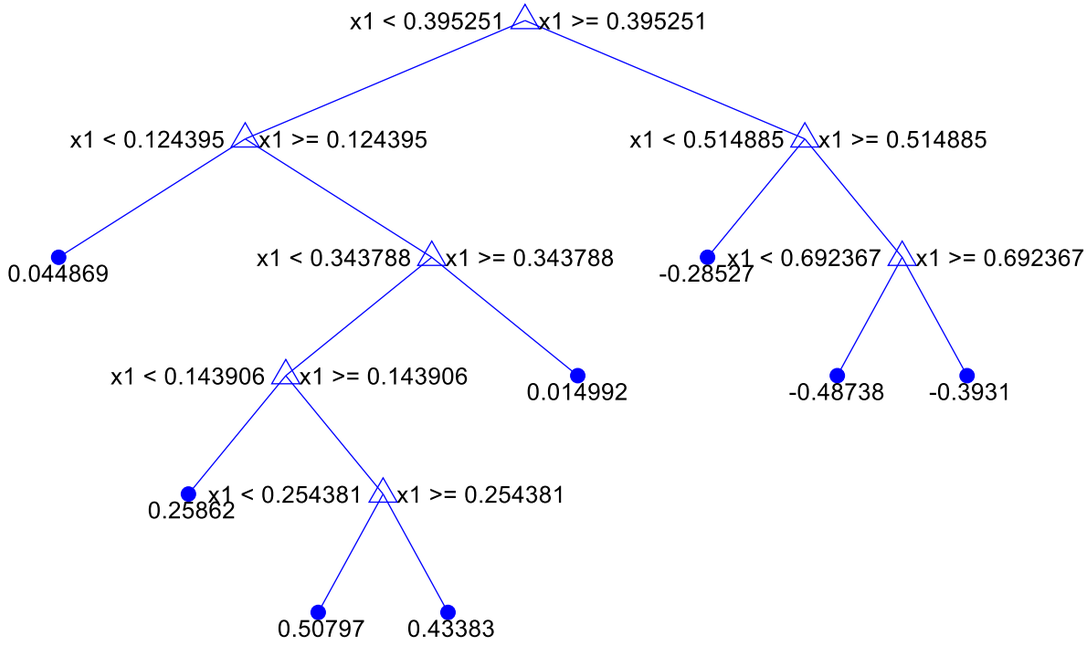
Dla $N_{max} = 25$.

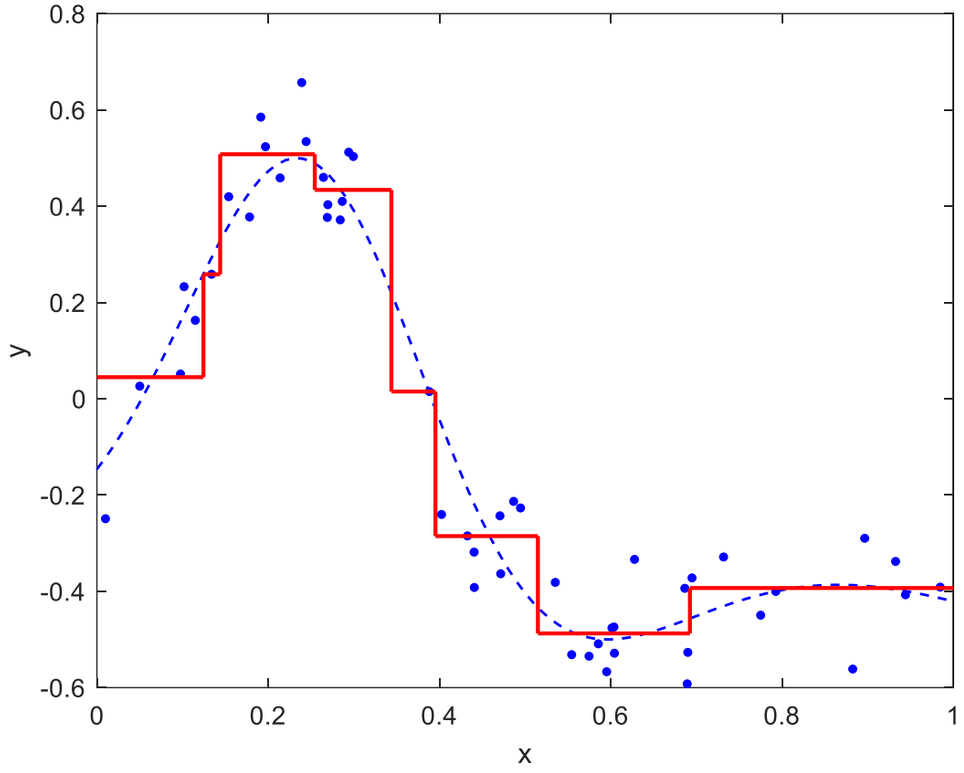




Kryterium stopu i ustalenie etykiety liścia

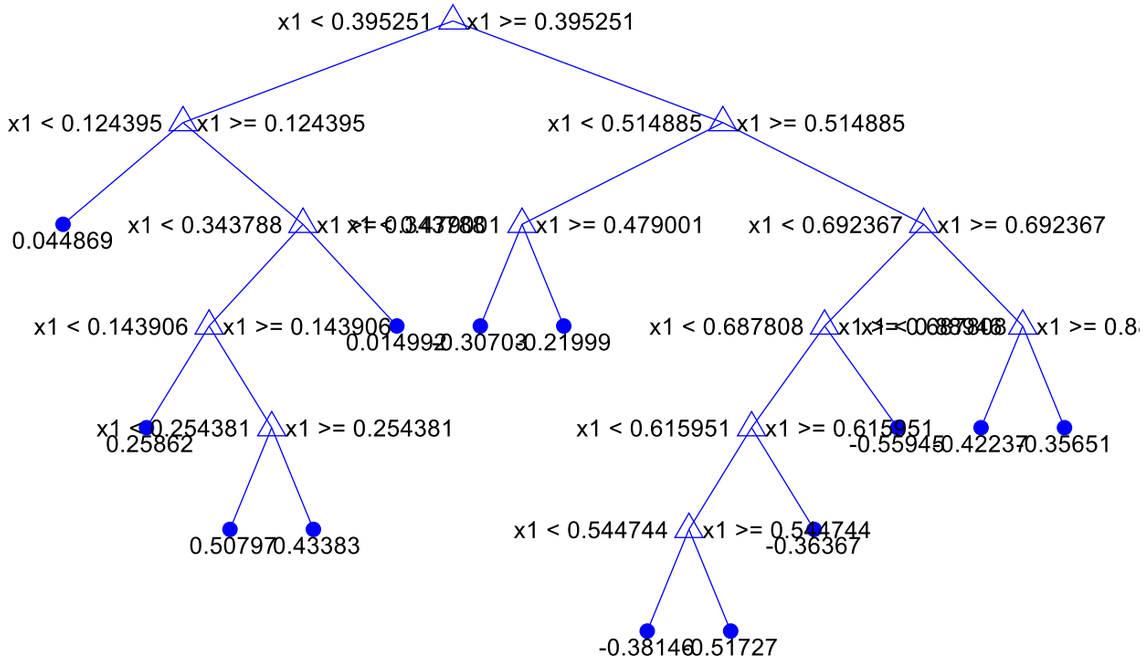
Dla $N_{max} = 13$.

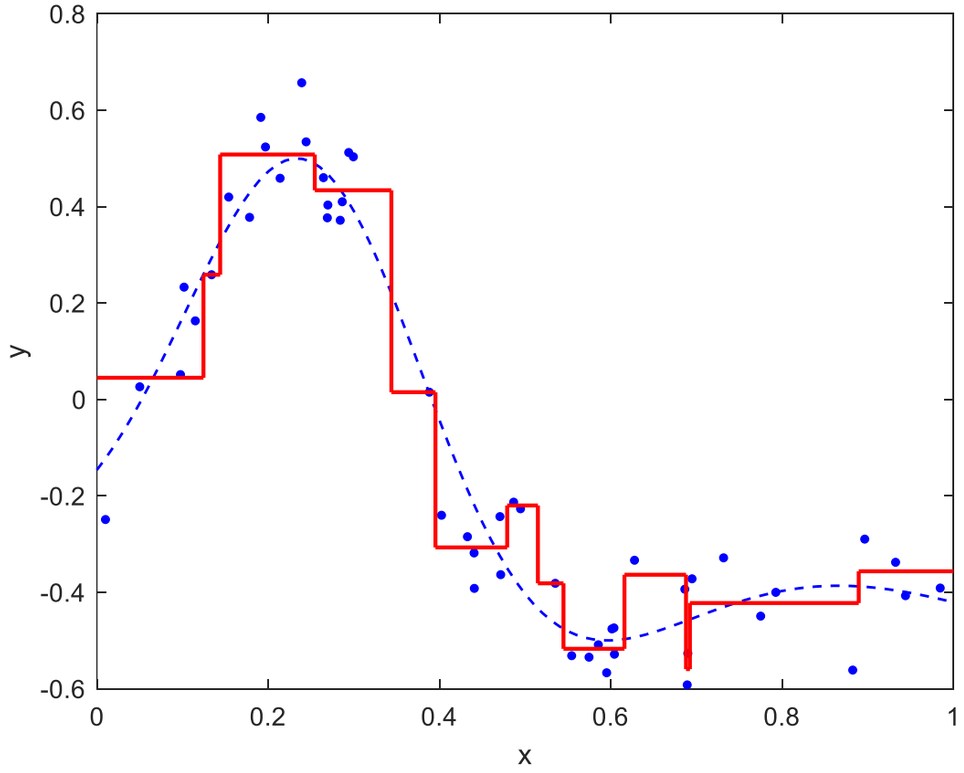




Kryterium stopu i ustalenie etykiety liścia

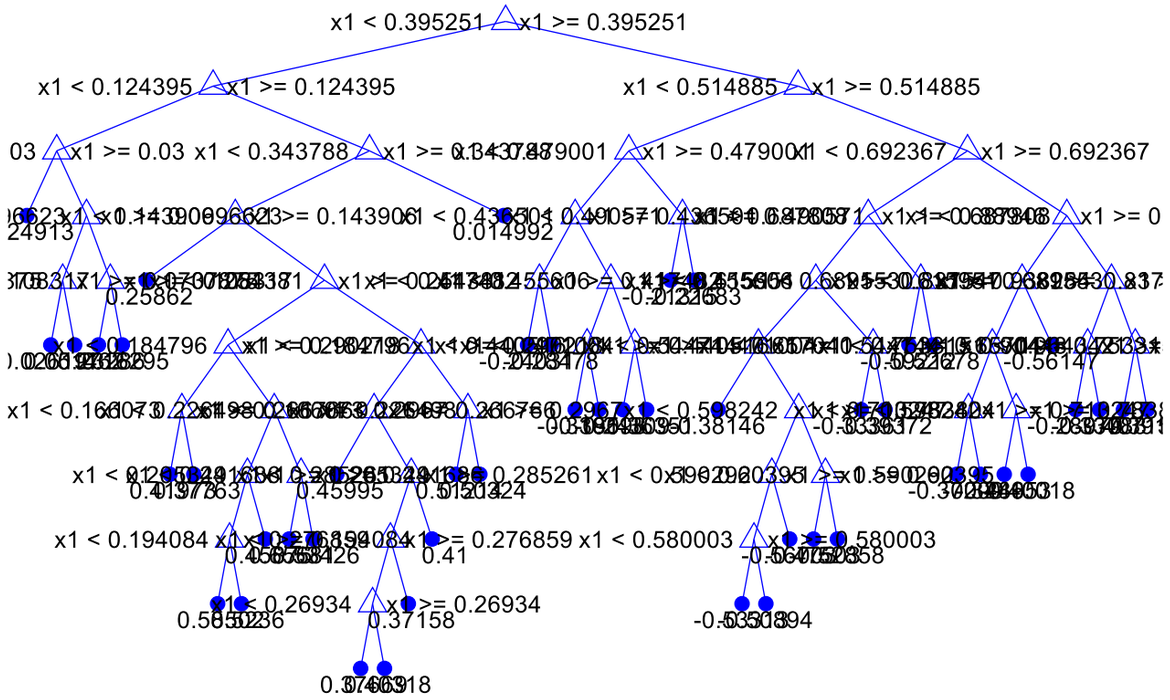
Dla $N_{max} = 8$.

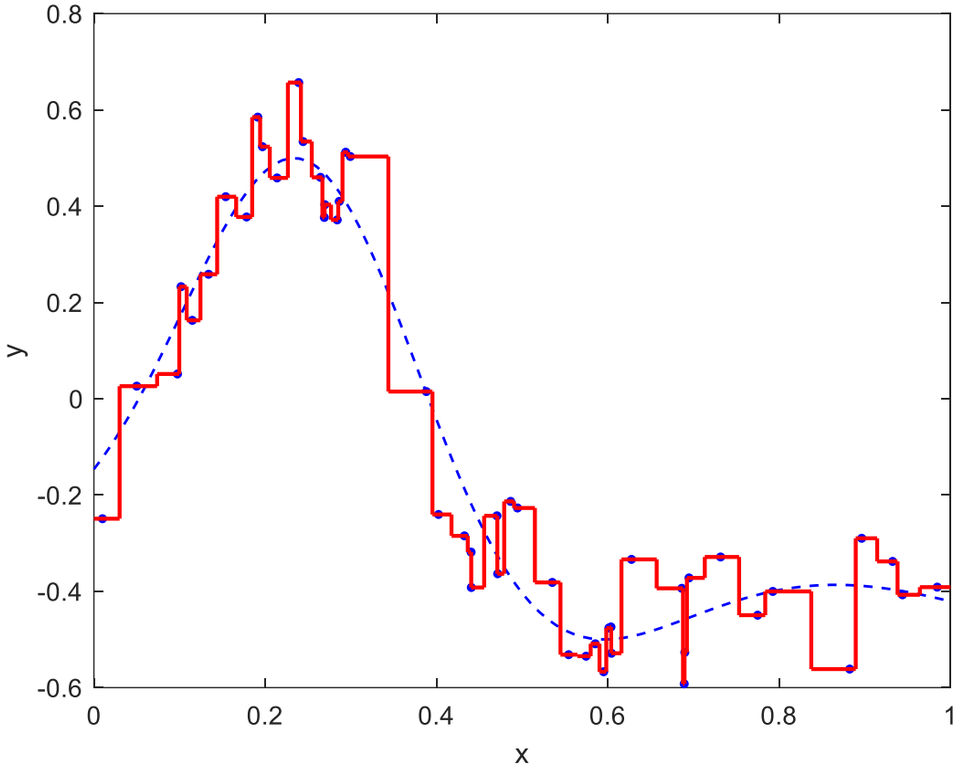




Kryterium stopu i ustalenie etykiety liścia

Dla $N_{max} = 1$.

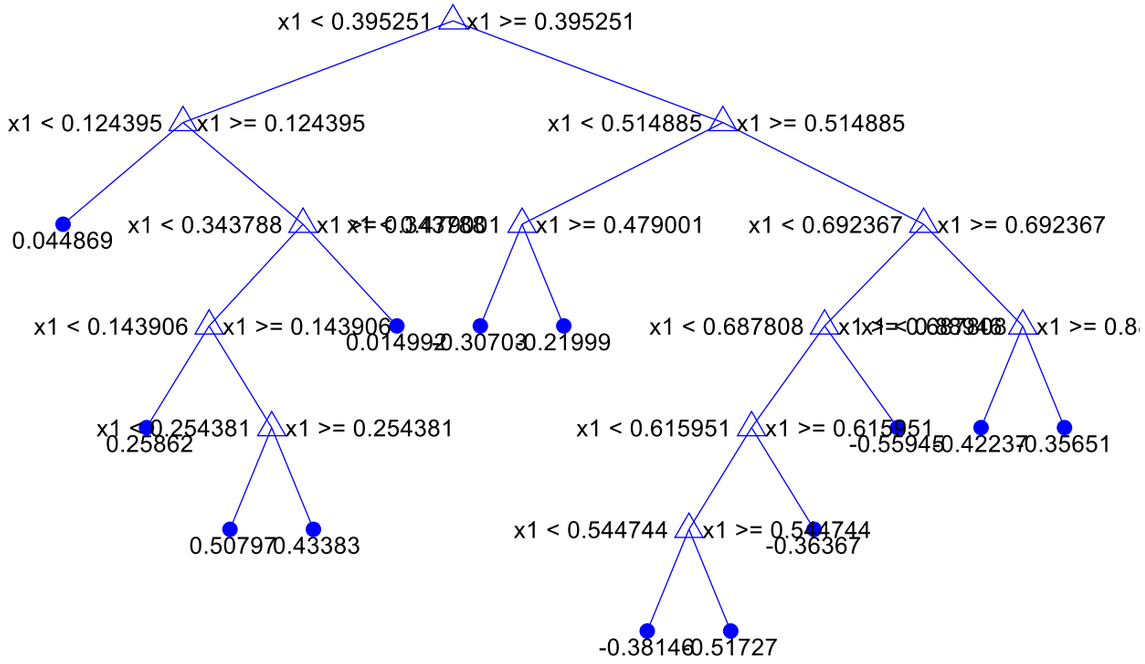


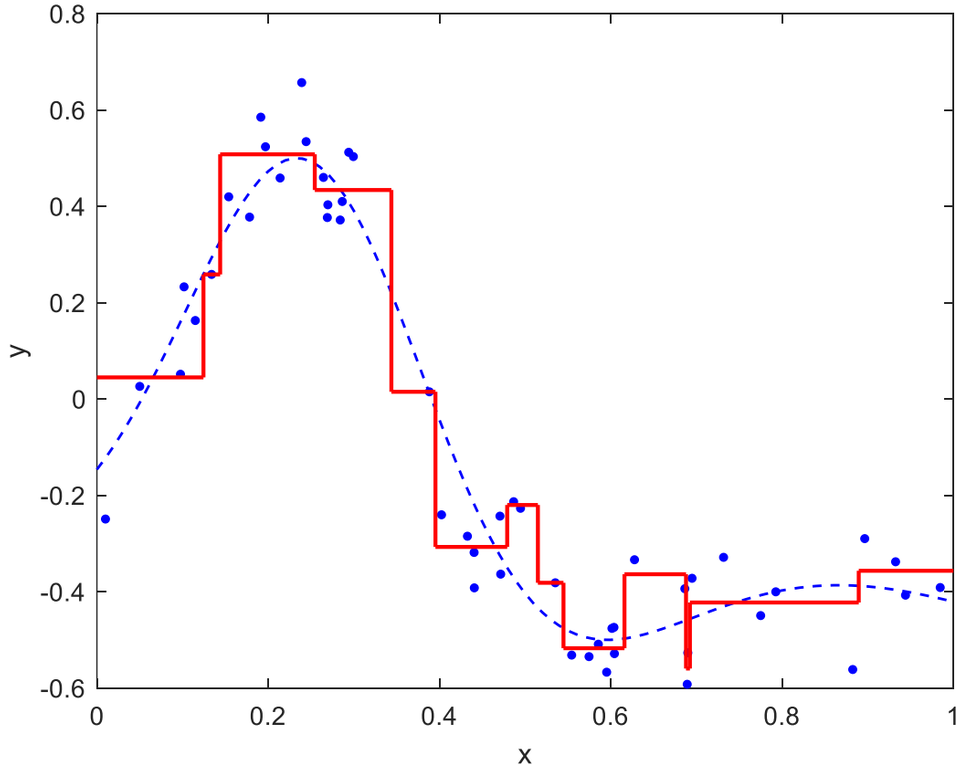


Przycinanie drzewa

Przycinanie drzewa – zastępowanie wybranych węzłów (a tym samym całych poddrzew) liśćmi.

Cel – redukcja przeuczenia (nadmienego dopasowania)

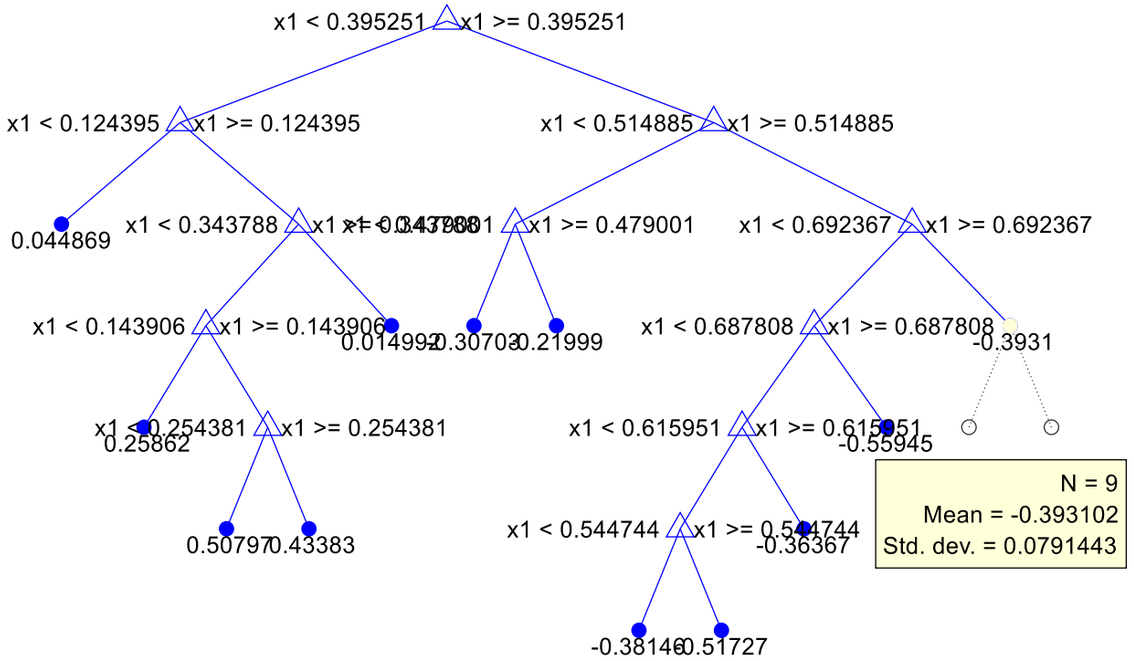


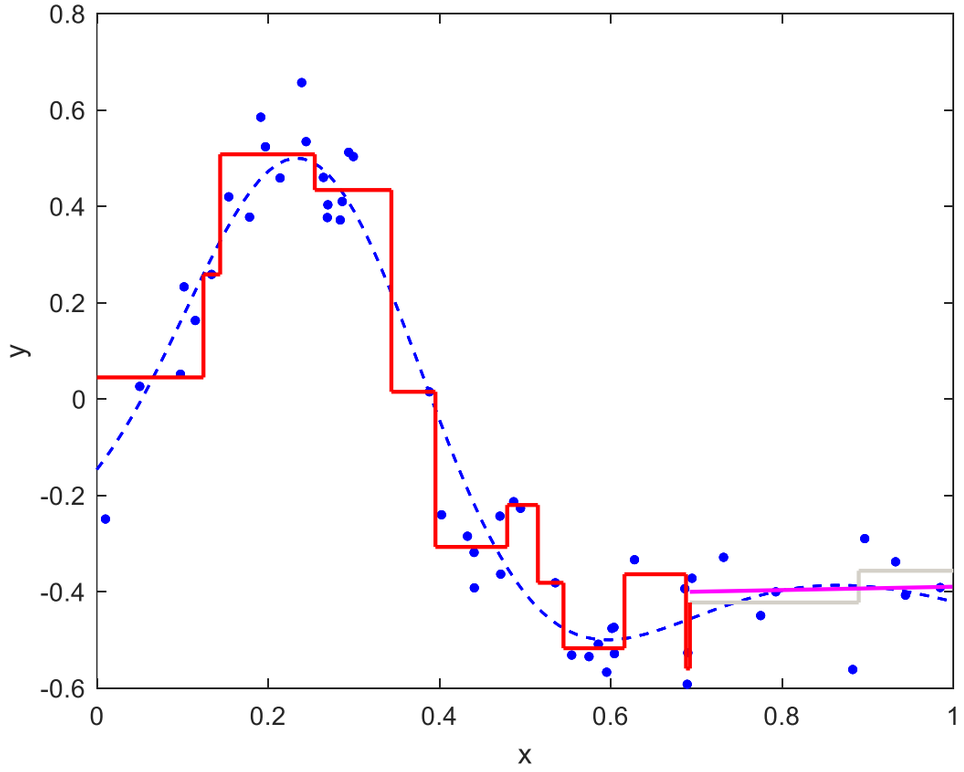


Przycinanie drzewa

Przycinanie drzewa – zastępowanie wybranych węzłów (a tym samym całych poddrzew) liśćmi.

Cel – redukcja przeuczenia (nadmienego dopasowania)

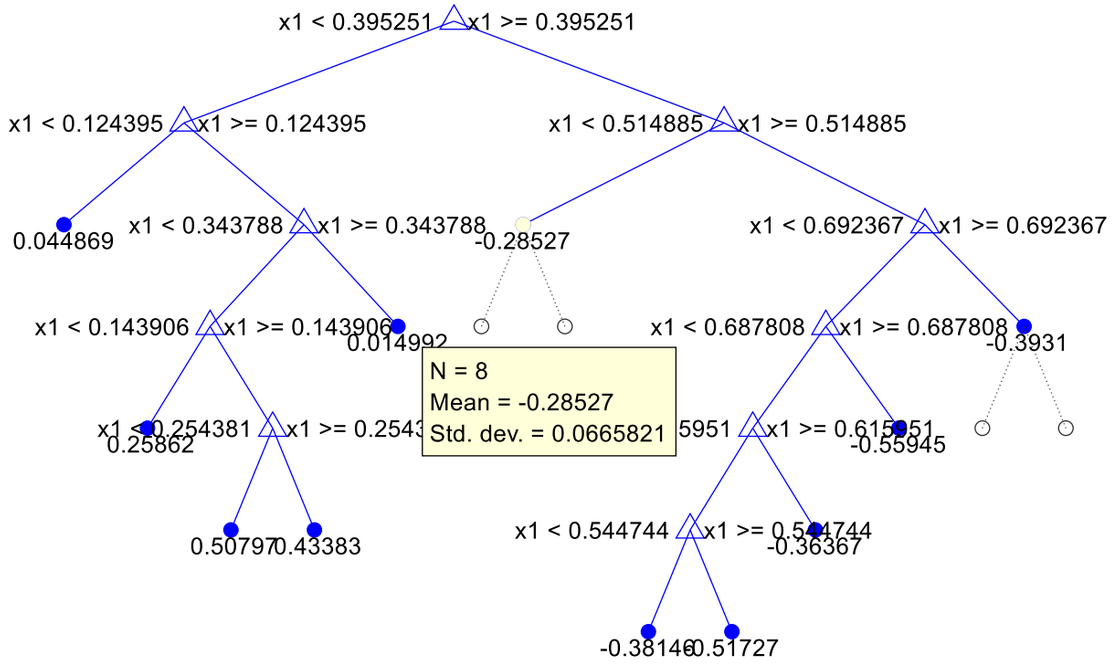


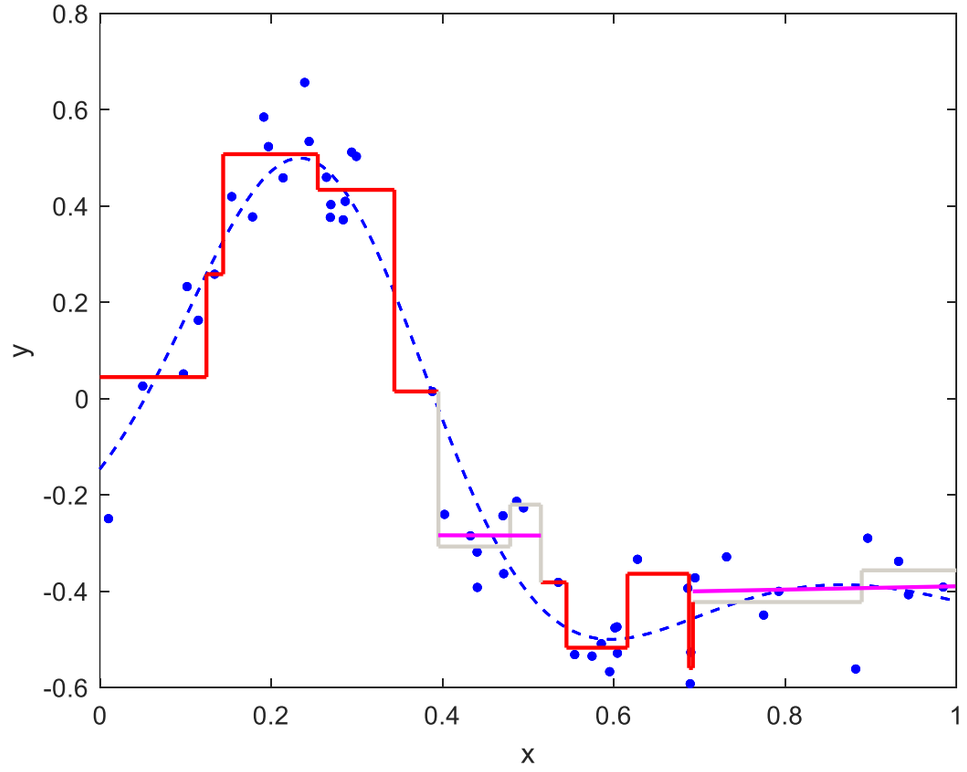


Przycinanie drzewa

Przycinanie drzewa – zastępowanie wybranych węzłów (a tym samym całych poddrzew) liśćmi.

Cel – redukcja przeuczenia (nadmienego dopasowania)

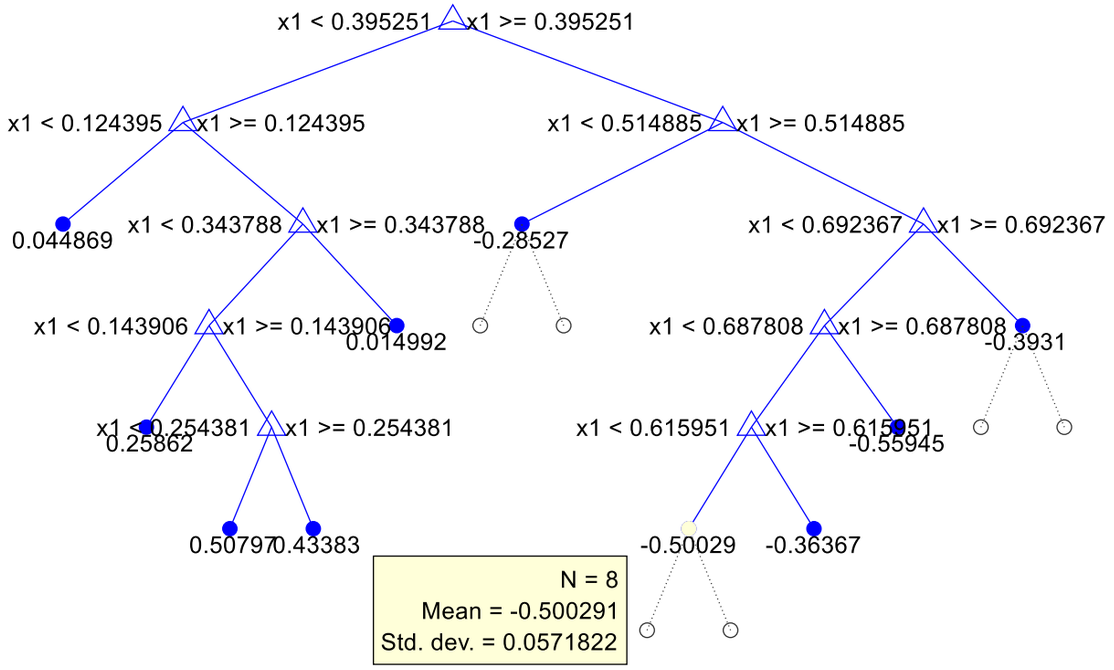


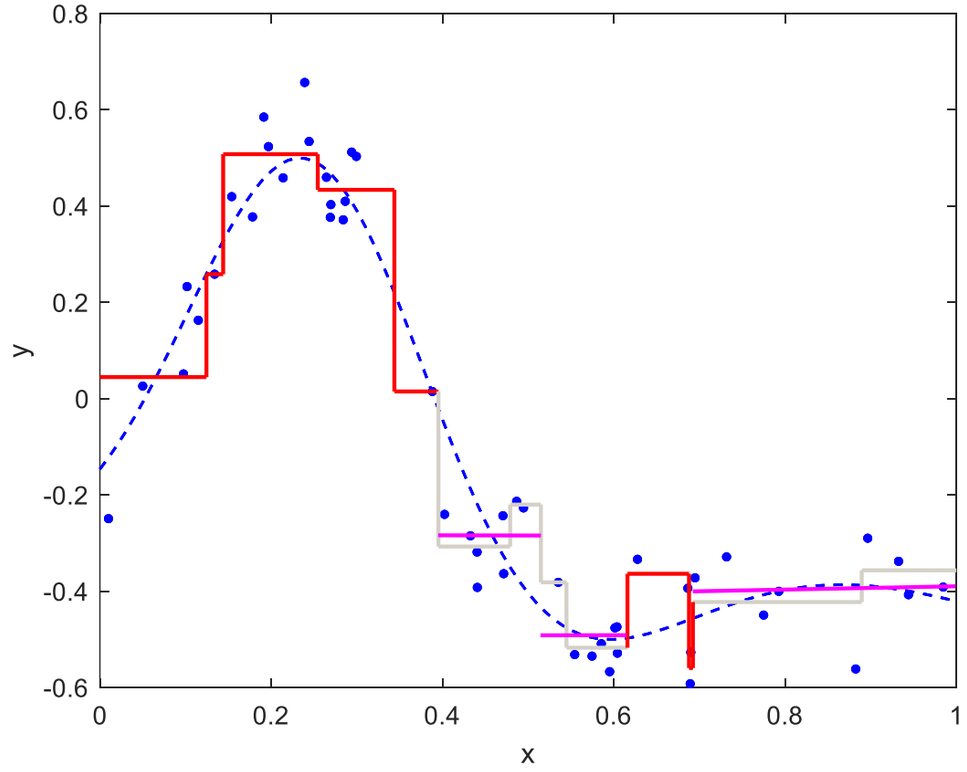


Przycinanie drzewa

Przycinanie drzewa – zastępowanie wybranych węzłów (a tym samym całych poddrzew) liśćmi.

Cel – redukcja przeuczenia (nadmienego dopasowania)

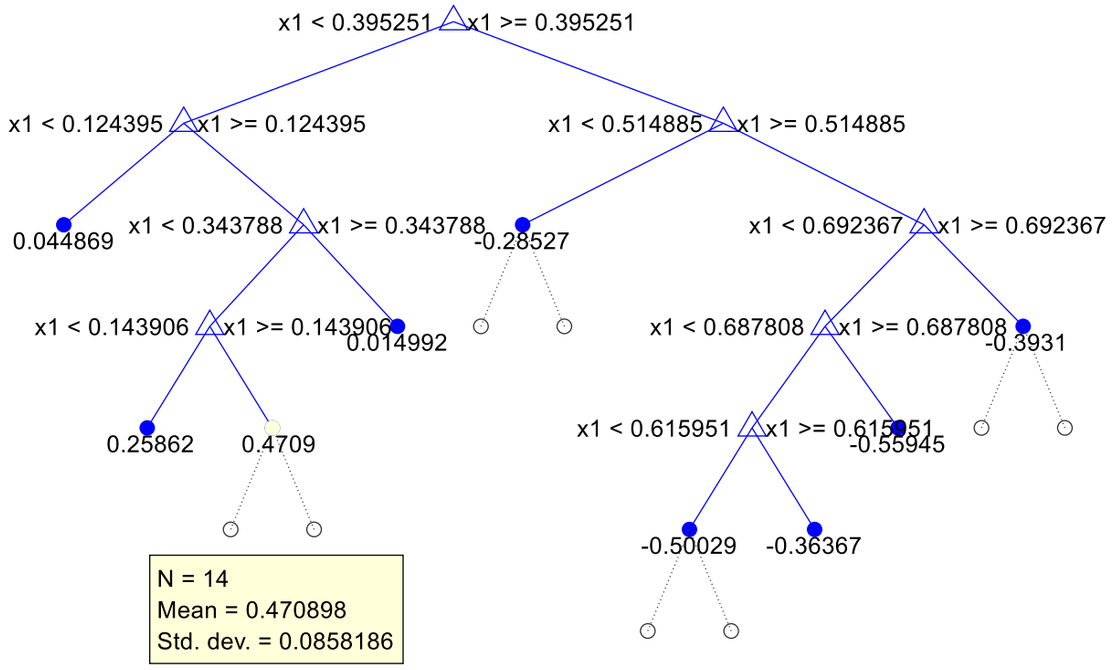


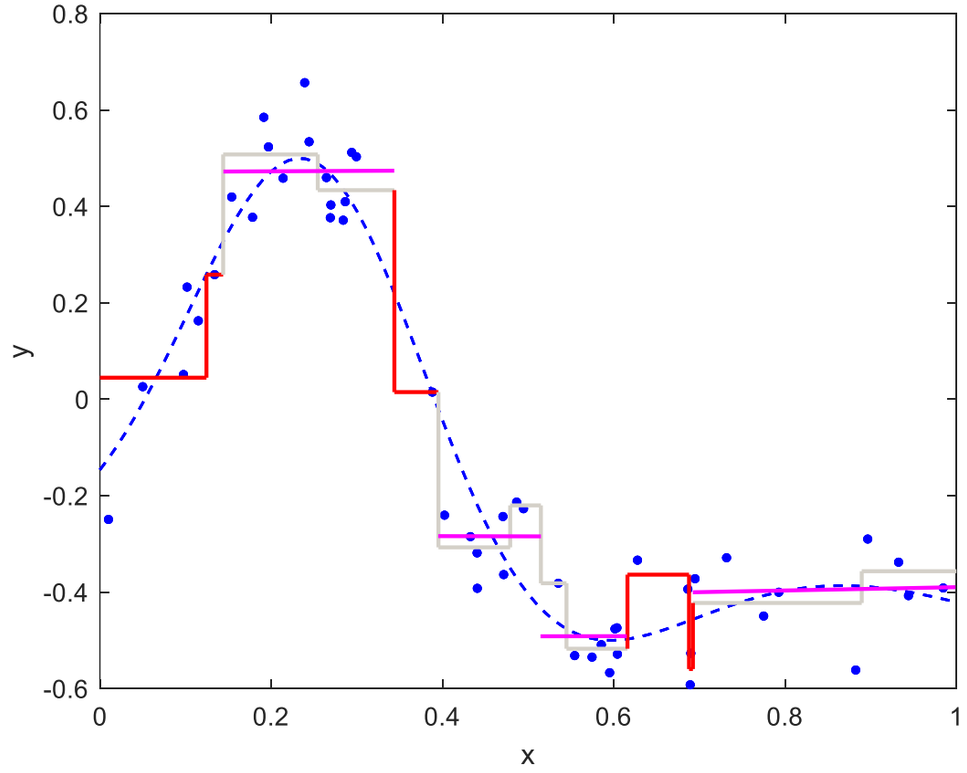


Przycinanie drzewa

Przycinanie drzewa – zastępowanie wybranych węzłów (a tym samym całych poddrzew) liśćmi.

Cel – redukcja przeuczenia (nadmienego dopasowania)

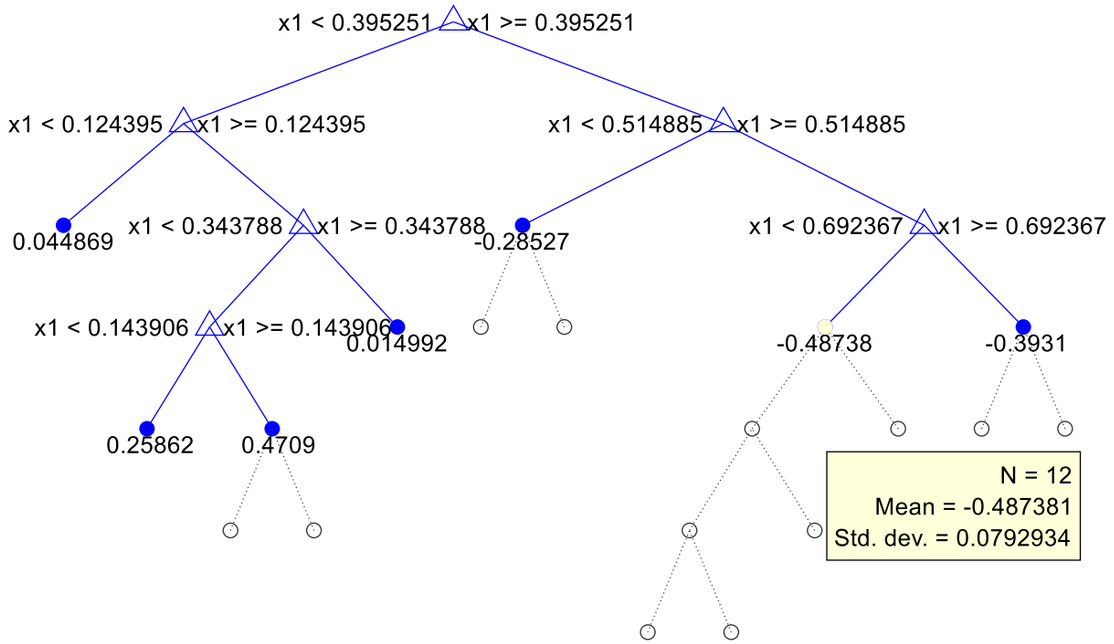


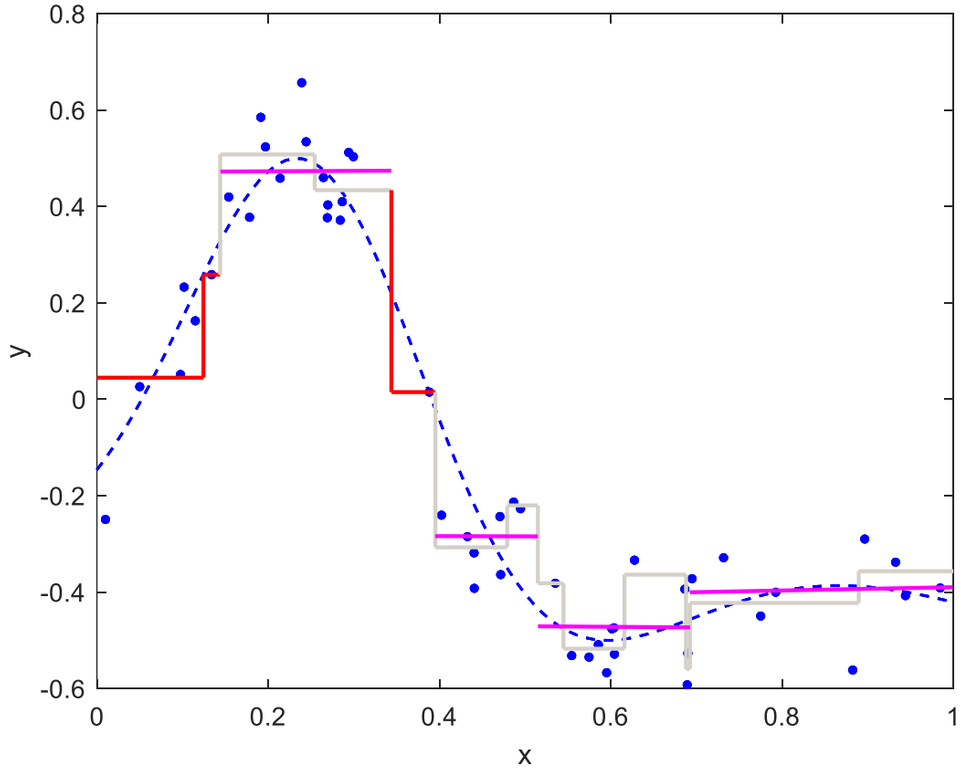


Przycinanie drzewa

Przycinanie drzewa – zastępowanie wybranych węzłów (a tym samym całych poddrzew) liśćmi.

Cel – redukcja przeuczenia (nadmienego dopasowania)





Przycinanie drzewa

- Przycinanie drzewa redukuje efekt przeuczenia (mały błąd uczący, duży testowy).
- Alternatywą przycinania jest zapobieganie nadmiernemu wzrostowi w trakcie konstrukcji drzewa poprzez tworzenie liści w miejsce węzłów.
- Zazwyczaj przycinanie realizowane jest w sposób wstępujący, tzn. w pierwszej kolejności rozważa się przycięcie węzłów położonych najniżej w drzewie. Podstawowe znaczenie ma kryterium przycinania, które decyduje, czy węzeł będzie zastąpiony liściem.
- Najczęściej kryterium przycinania jest błąd na oddzielnym zbiorze przykładów (niezależnym od zbioru trenującego). Przycięcie następuje, jeśli na tym oddzielnym zbiorze liść uzyskuje błąd nie większy, niż błąd poddrzewa.

Algorytm konstrukcji drzewa regresyjnego

Wejście: zbiór przykładów uczących Φ, parametry modelu (np. N_{max})

Wyjście: drzewo regresyjne

Procedura:

1. Utwórz liść lub węzeł dla przykładów w zbiorze Φ:
   a. Jeśli spełniony jest warunek zatrzymania (np. liczba przykładów w węźle jest nie większa niż N_{max}), to utwórz liść,

      w przeciwnym przypadku wykonaj b, c i d:

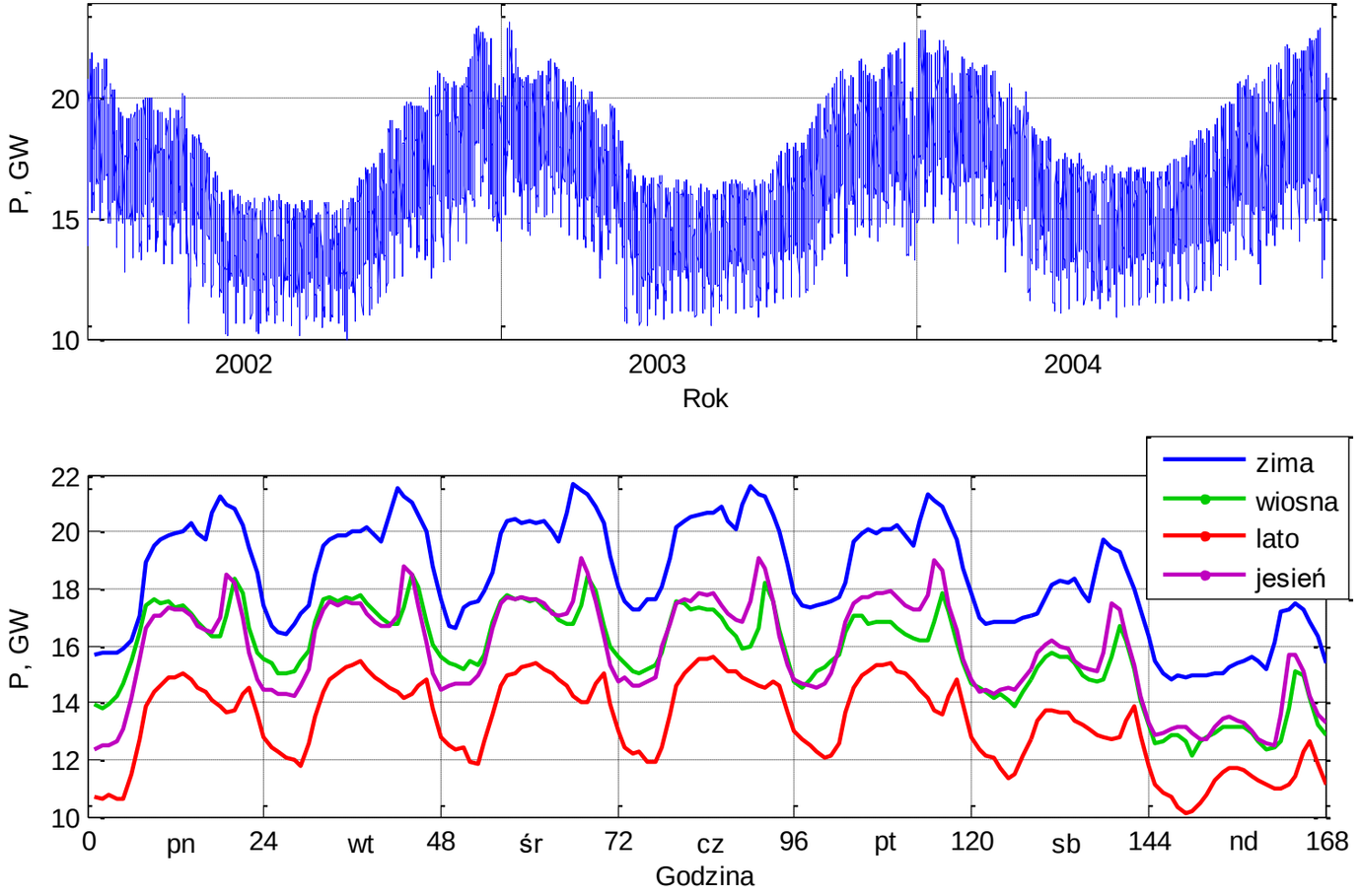
   b. Utwórz węzeł z testem T. Do testu wybierz atrybut x_i i punkt podziału p_i, dla których uzyskano największą redukcję zróżnicowania ΔI w zbiorze Φ.

   c. Podziel zbiór Φ na dwa podzbiory, tj. przykładów zaliczających test T (Φ_1) i niezaliczających (Φ_0).

   d. Powtórz p. 1 dla Φ_1 i Φ_0.

Przykład zastosowania drzewa regresyjnego

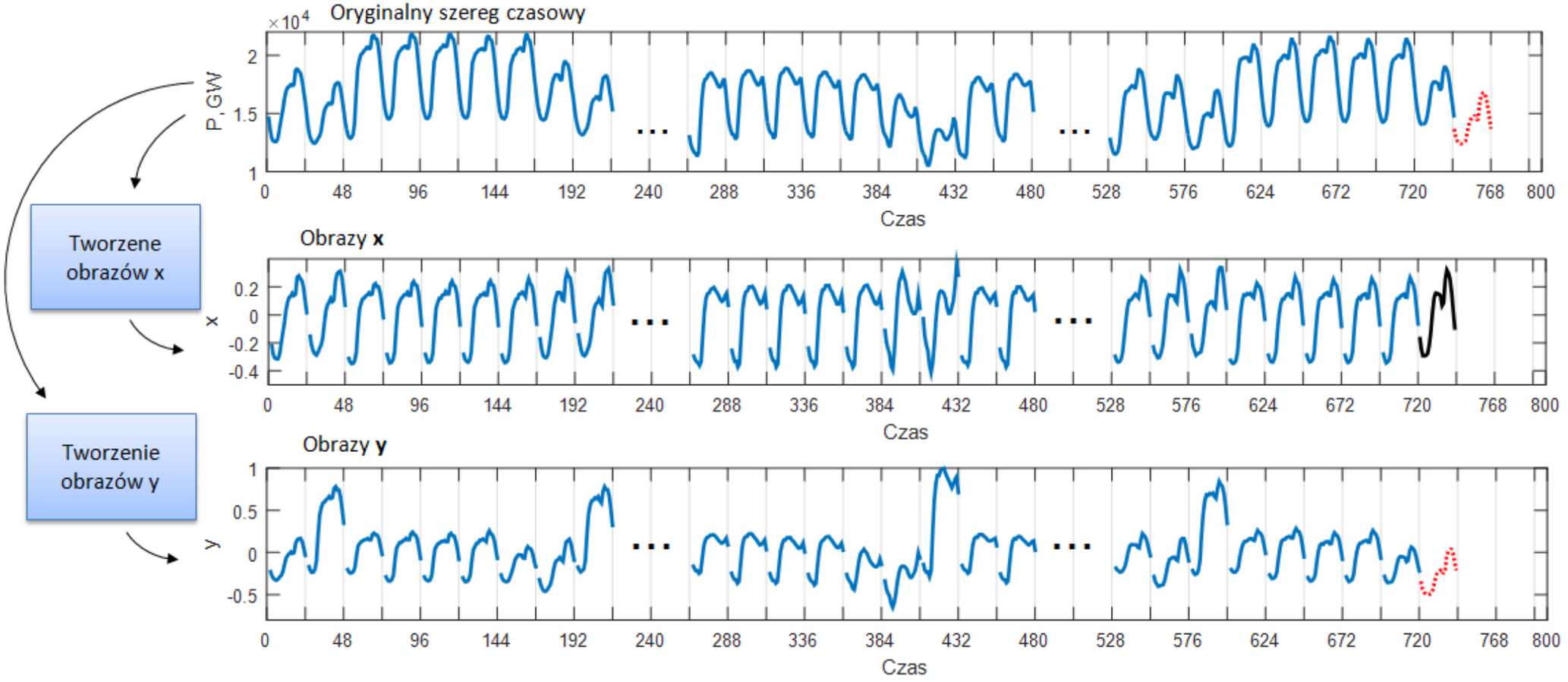
Problem: Predykcja szeregu czasowego z wieloma cyklami wahań sezonowych w horyzoncie τ na podstawie przebiegu historycznego.



Obciążenie system elektroenergetycznego z cyklami rocznymi, tygodniowymi i dobowymi

Przykład zastosowania drzewa regresyjnego

Reprezentacja szeregów czasowych za pomocą obrazów/wzorców sekwencji dobowych

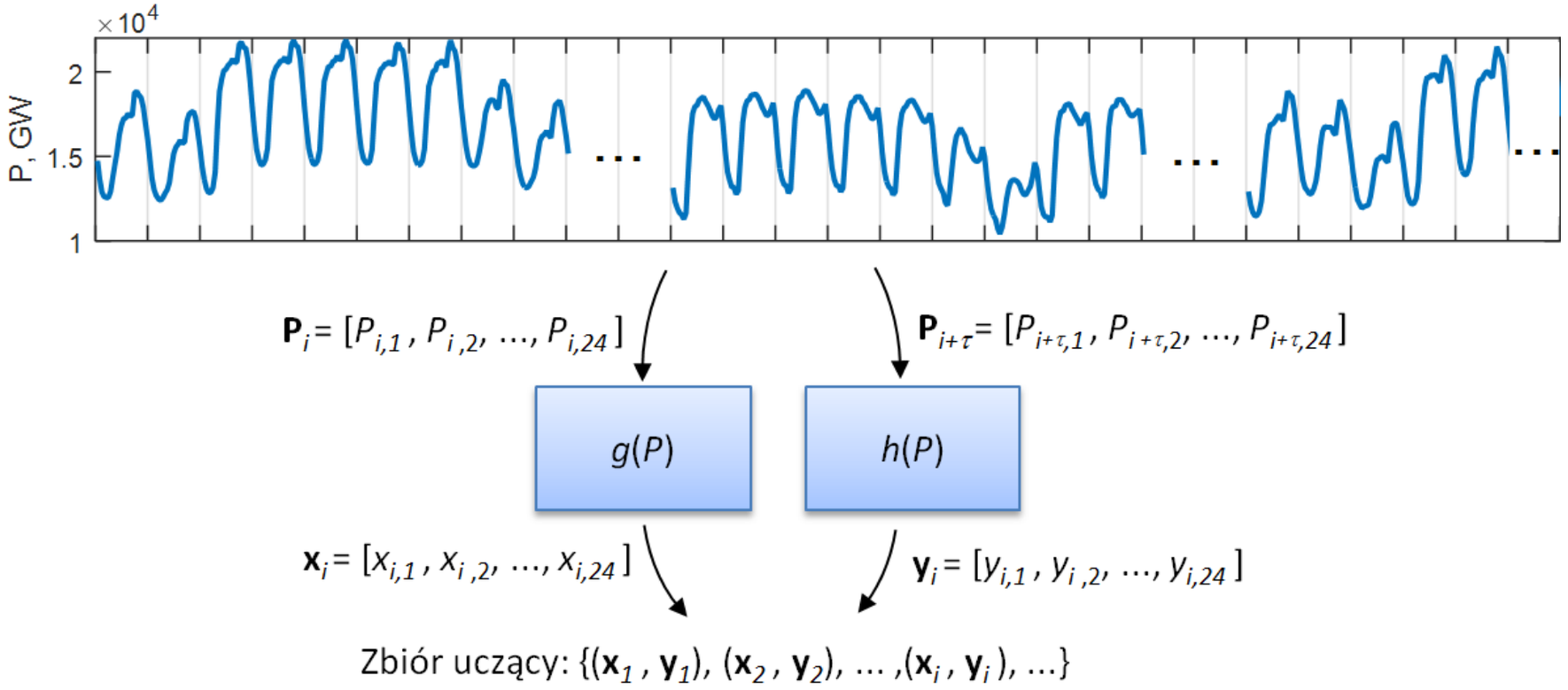


⇒ Dudek G.: *Systemy uczące się oparte na podobieństwie obrazów do prognozowania szeregów czasowych obciążeń elektroeneregtycznych*. EXIT, Warszawa 2012

⇒ Dudek G.: *Pattern Similarity-based Methods for Short-term Load Forecasting – Part 1: Principles*. Applied Soft Computing, vol. 37, pp. 277-287, 2015

Przykład zastosowania drzewa regresyjnego

Reprezentacja szeregów czasowych za pomocą obrazów/wzorców sekwencji dobowych

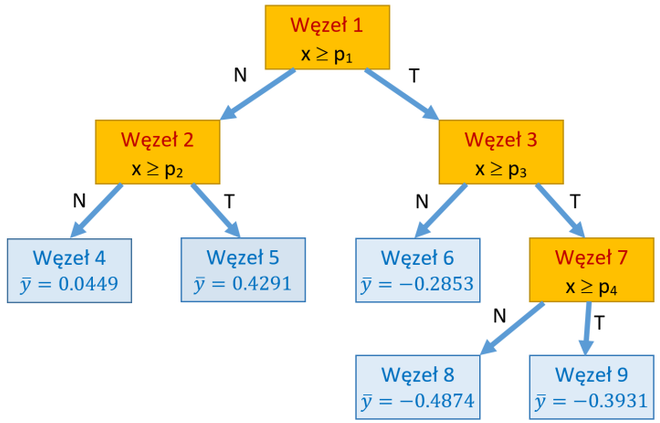


Przykład zastosowania drzewa regresyjnego

Model prognostyczny:

$$f : X \rightarrow Y$$

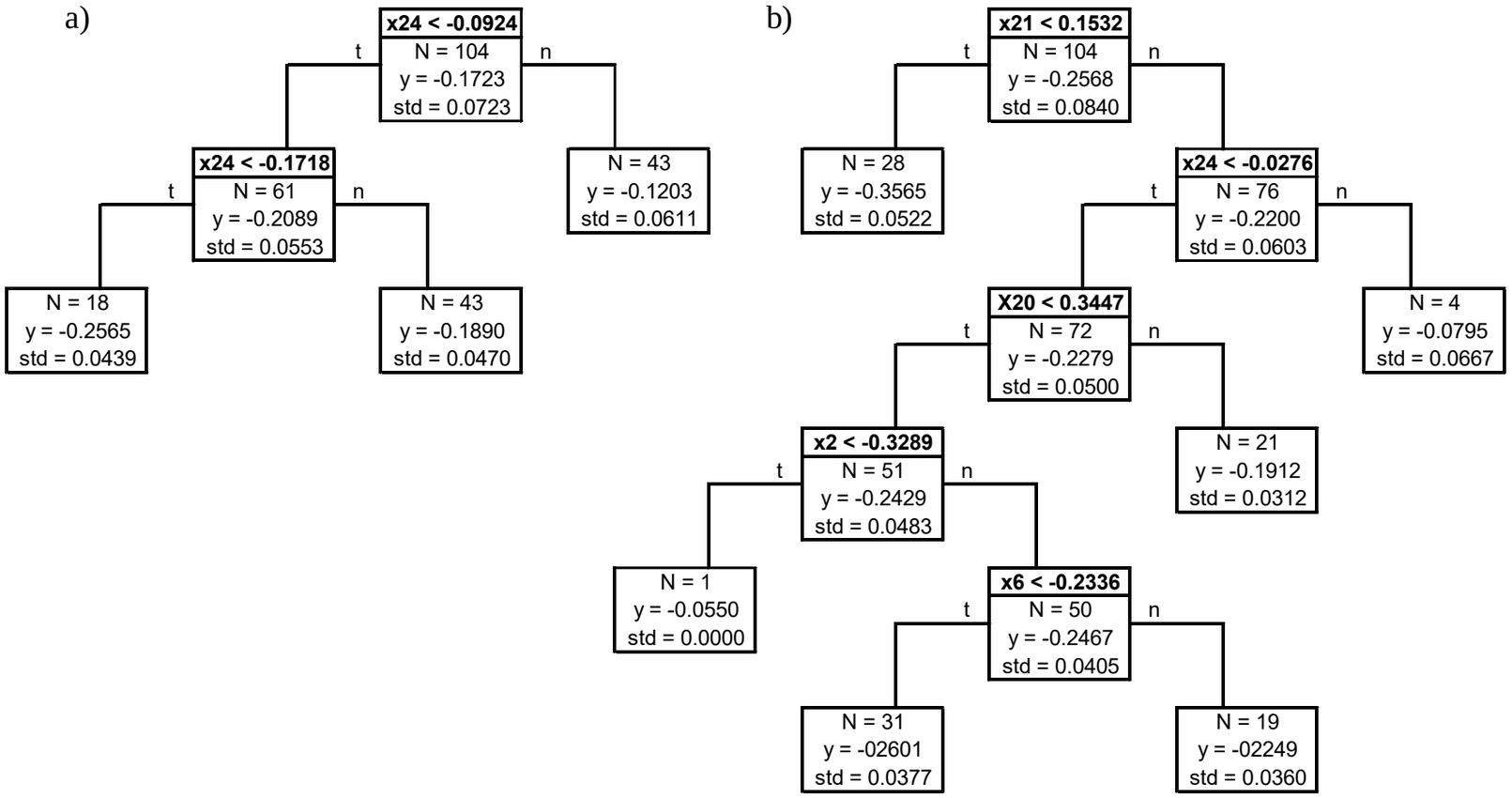
$\mathbf{x} \rightarrow$



$\rightarrow y_i$

Przykład zastosowania drzewa regresyjnego

Wyniki:

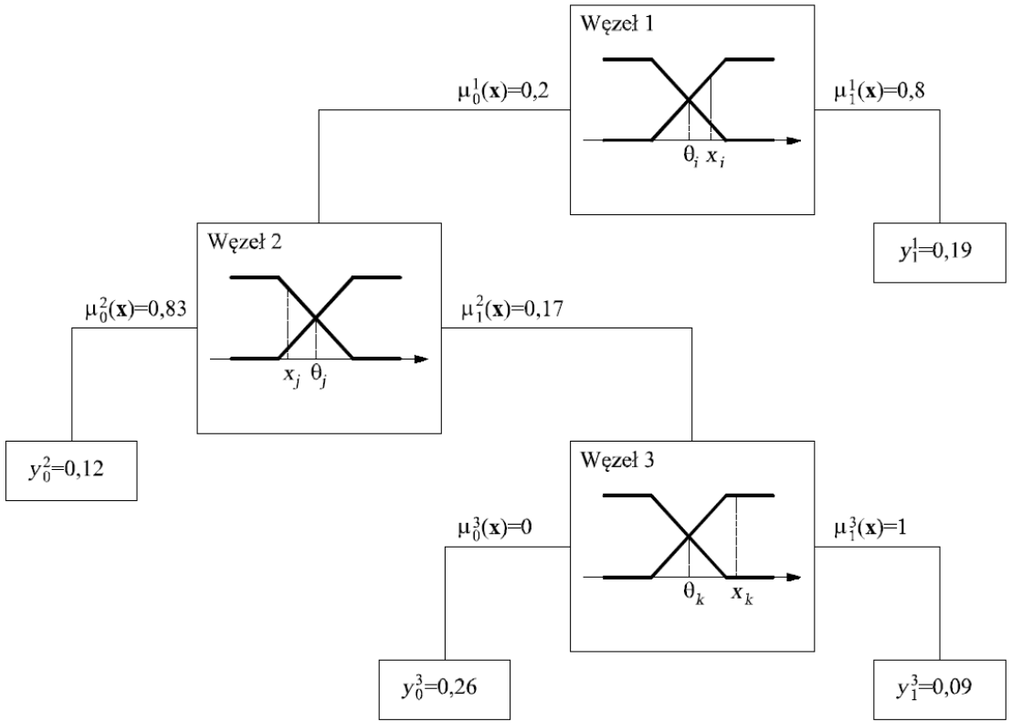


Drzewa regresyjne dla prognoz obciążeń o godz. 1 (a) i 6 (b) w środy.

⇒ Dudek G.: *Drzewo regresyjne jako narzędzie prognostyczne*. Mat. Konf. Prognozowanie w Elektroenergetyce 2004, str. 99-105.

Przykład zastosowania drzewa regresyjnego

Inne przykłady drzew regresyjnych do prognozowania szeregów czasowych – drzewo regresyjne z rozmytymi węzłami

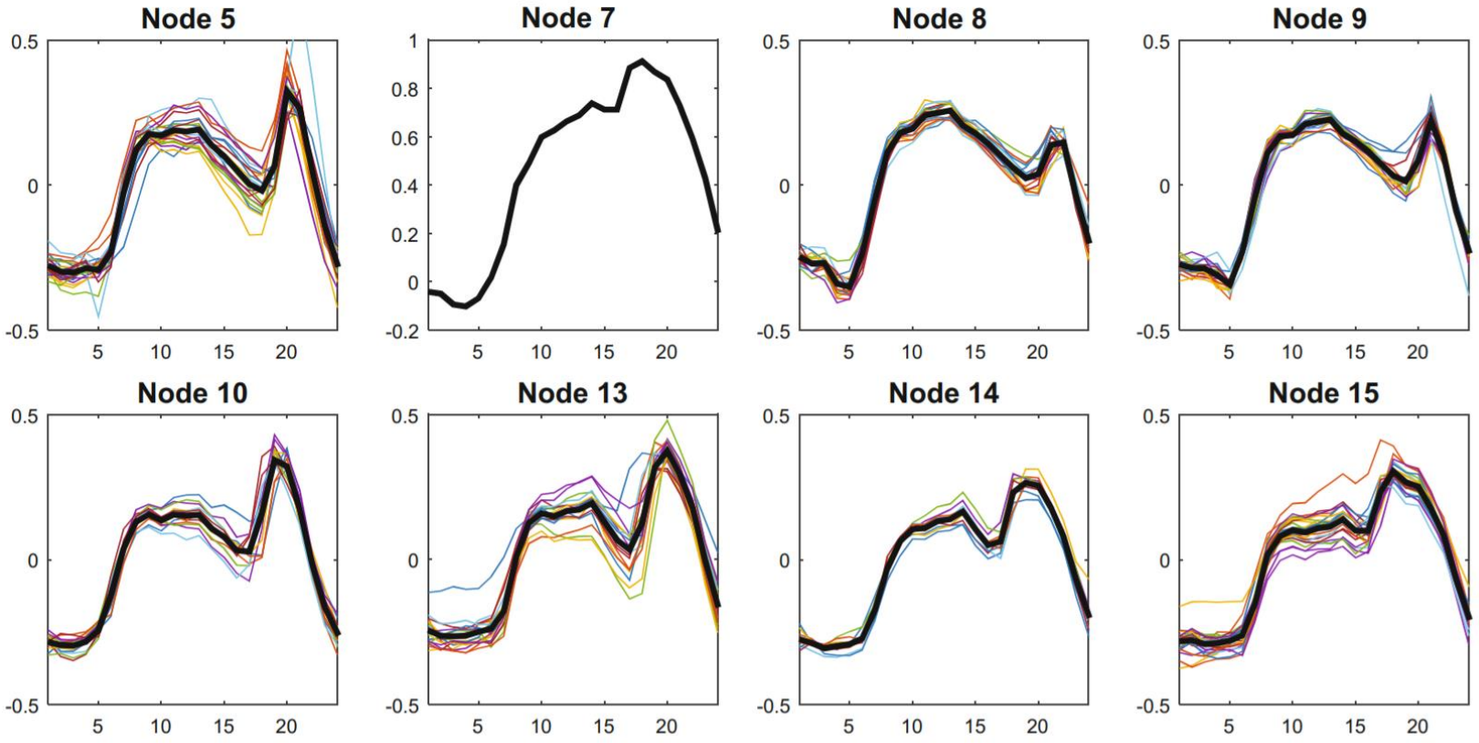


$$y(\mathbf{x}) = y_1^1\mu_1^1(\mathbf{x}) + y_0^2\mu_0^2(\mathbf{x})\mu_0^1(\mathbf{x}) + y_0^3\mu_0^3(\mathbf{x})\mu_1^2(\mathbf{x})\mu_0^1(\mathbf{x})$$
$$+ y_1^3\mu_1^3(\mathbf{x})\mu_1^2(\mathbf{x})\mu_0^1(\mathbf{x}) = 0{,}19\cdot 0{,}8 + 0{,}12\cdot 0{,}83\cdot 0{,}2$$
$$+ 0{,}26\cdot 0\cdot 0{,}17\cdot 0{,}2 + 0{,}09\cdot 1\cdot 0{,}17\cdot 0{,}2 = 0{,}175$$

⇒ Dudek G.: *Prognozowanie krótkoterminowe obciążeń systemów elektroenergetycznych z wykorzystaniem rozmytych drzew regresyjnych*. Przegląd Elektrotechniczny, r. 90, nr 4, s. 108-111, 2014.

Przykład zastosowania drzewa regresyjnego

Inne przykłady drzew regresyjnych do prognozowania szeregów czasowych – wielowymiarowe drzewo regresyjne



⇒ Dudek G.: *Multivariate Regression Tree for Pattern-based Forecasting Time Series with Multiple Seasonal Cycles*. ISAT 2017. Advances in Intelligent Systems and Computing 655, pp. 85-94, 2018.

Przykład zastosowania drzewa regresyjnego

Inne przykłady drzew regresyjnych do prognozowania szeregów czasowych – losowy las regresyjny

1. For $k = 1$ to K:
   1.1. Draw a bootstrap sample L of size N from the training data.
   1.2. Grow a random-forest tree T_k to the bootstrapped data, by recursively repeating the following steps for each node of the tree, until the minimum node size m is reached.
      1.2.1. Select F variables at random from the n variables.
      1.2.2. Pick the best variable/split-point among the F.
      1.2.3. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_k\}_{k=1,2,\ldots,K}$.

To make a prediction at a new point $\mathbf{x}$:

$$f(\mathbf{x}) = \frac{1}{K}\sum_{k=1}^{K} T_k(\mathbf{x}) \qquad (1)$$

Fig. 1. Algorithm of RF for regression.

Table 1. Results of forecasting.

Model	January		July		Mean	
	$MAPE_{tst}$	IQR	$MAPE_{tst}$	IQR	$MAPE_{tst}$	IQR
RF	1.42	1.39	0.92	0.98	1.16	1.17
CART	1.70	1.58	1.16	1.17	1.42	1.39
Fuzzy CART	1.62	1.47	1.13	1.12	1.37	1.35
ARIMA	2.64	2.34	1.21	1.24	1.91	1.67
ES	2.35	1.88	1.19	1.30	1.76	1.56
ANN	1.32	1.30	0.97	1.01	1.14	1.15
Naïve	6.37	5.36	1.29	1.20	3.78	3.82

⇒ Dudek G.: *Short-Term Load Forecasting using Random Forests*. Intelligent Systems'2014, Advances in Intelligent Systems and Computing 323, pp. 821-828, 2015.

Podsumowanie

- Model interpretowalny – reprezentacja drzewiasta lub zbiór reguł decyzyjnych "jeśli – to"
- Czytelny algorytm konstrukcji modelu
- Działanie na atrybutach nominalnych, porządkowych i ciągłych
- Nie wymaga specjalnego przygotowania danych
- Efektywne pamięciowo
- Wbudowany mechanizm selekcji atrybutów
- Podział przestrzeni cech na hiperprostopadłościany
- Lokalna aproksymacja funkcji stałą wewnątrz hiperprostopadłościanu (aproksymacja dyskretna)
- Zależnie od funkcji docelowej drzewo decyzyjne może pełnić rolę klasyfikatora lub modelu regresyjnego
- Zachłanny algorytm optymalizacyjny nie zawsze prowadzi do optymalnych rozwiązań
- Model niestabilny – niewielkie zmiany w danych uczących mogą powodować duże zmiany modelu
- Na bazie drzew decyzyjnych budowane są komitety: lasy losowe, *boosted decision trees* (XGBoost, Light GMB, CatBoost)