

# ZAAWANSOWANE METODY EKSPŁORACJI

## UCZENIE SIĘ INDUKCYJNE

Prof. dr hab. inż. Grzegorz Dudek  
Wydział Matematyki i Informatyki  
Uniwersytet Łódzki

Wiedza pozyskana przez ucznia ma charakter odwzorowania informacji wejściowej za zbiór wartości wyjściowych.

**Informacja wejściowa** – opisy obiektów pewnej dziedziny

**Dziedzina** – zbiór obiektów  $X$ , których dotyczy wiedza zdobywana przez ucznia (przedmioty, osoby, wydarzenia, sytuacje, procesy, ...)

**Przykład** – obiekt, element dziedziny  $x \in X$

**Atrybut** (cecha, zmienna) – przykłady opisywane są za pomocą atrybutów  $x_i$ :  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$

**Wartości wyjściowe** – klasa lub wartość funkcji



## Typy atrybutów:

- **nominalne** – o skończonym zbiorze nieuporządkowanych wartości dyskretnych,  
np. kolor, kształt, marka; relacje  $>$ ,  $<$  są nieokreślone
- **porządkowe** – o skończonym zbiorze uporządkowanych wartości dyskretnych,  
np. rozmiar (S, M, L, XL), wzrost (niski, średni, wysoki); relacje  $>$ ,  $<$  są określone
- **ciągłe** – o wartościach ze zbioru liczb rzeczywistych,  
np. temperatura, prędkość, masa

*Przykład: punkty na płaszczyźnie.*

**Dziedzina** –  $X = \mathbb{R}^2$

**Atrybuty** – dwie współrzędne kartezjańskie:  $x_1, x_2 \in \mathbb{R}$

**Przykład** – punkt na płaszczyźnie:  $\mathbf{x} = [1.23, 2.56]^T$

**Klasa** – punkty z pierwszej ćwiartki, punkty z drugiej ćwiartki, ...

*Przykład: n-elementowe łańcuchy binarne*

**Dziedzina** –  $X = \{0, 1\}^n$

**Atrybuty** – wartości binarne:  $x_1, x_2, \dots, x_n \in \{0, 1\}$

**Przykład** – łańcuch binarny:  $\mathbf{x} = [01001010\dots]^T$

**Klasa** – łańcuchy z jedną jedynką, łańcuchy z dwiema jedynkami, ...

*Przykład: figury geometryczne*

**Dziedzina** – kolorowe figury geometryczne

**Atrybuty** – rozmiar, kolor, kształt

**Przykład** –  $\mathbf{x}$  = (duży, niebieski, trójkąt)

**Klasa** – czworoboki, pięcioboki, trójkąty, ...

*Przykład: pogoda*

**Dziedzina** – stany pogody

**Atrybuty** – aura (słoneczna, pochmurna, deszczowa), temperatura (zimna, umiark., ciepła), wilgotność (normalna, duża), wiatr (słaby, silny)

**Przykład** –  $\mathbf{x}$  = (słoneczna, zimna, duża, słaby)

**Klasa** – pogoda ładna, pogoda brzydka

Przykłady trenujące do uczenia się pojęć (przykłady etykietowane): **opis obiektu + etykieta klasy**

$\langle \mathbf{x}, d \rangle$

np.  $\langle [1.23, 2.56], 1 \rangle$ ;  $\langle [1001000], 2 \rangle$ ;  $\langle (\text{duży, niebieski, trójkąt}), \text{trójkąty} \rangle$ ;  $\langle (\text{słoneczna, zimna, duża, silny}), \text{pogoda brzydka} \rangle$

**Zadanie ucznia** (uczenie z nadzorem) – znalezienie hipotezy (wiedzy), która jest spójna (zgodna) z pojęciem docelowym (klasą) dla przykładów trenujących i która klasyfikuje również inne przykłady z jak najmniejszym błędem.

**Hipoteza** jest funkcją przypisującą przykładom ich klasy –  $h : X \rightarrow C$ , gdzie  $X$  jest dziedziną (zbiorem obiektów), a  $C$  jest zbiorem ich klas.

Dla zadania "punkty na płaszczyźnie" hipoteza może mieć postać:

Jeśli  $\text{znak}(x_1) = "+"$  i  $\text{znak}(x_2) = "+"$  to  $\hat{d} = 1$  (punkty z pierwszej ćwiartki)

Jeśli  $\text{znak}(x_1) = "-"$  i  $\text{znak}(x_2) = "+"$  to  $\hat{d} = 2$  (punkty z drugiej ćwiartki)

Jeśli  $\text{znak}(x_1) = "-"$  i  $\text{znak}(x_2) = "-"$  to  $\hat{d} = 3$  (punkty z trzeciej ćwiartki)

Jeśli  $\text{znak}(x_1) = "+"$  i  $\text{znak}(x_2) = "-"$  to  $\hat{d} = 4$  (punkty z pierwszej ćwiartki)

# UCZENIE SIĘ POJĘĆ

Dla zadania " $n$ -elementowe łańcuchy binarne" hipoteza może mieć postać:

$$\hat{d} = x_1 + x_2 + \dots + x_n$$

Hipoteza uzyskana w wyniku uczenia się pojęć może być stosowana do klasyfikacji innych przykładów z dziedziny. Uczeń na wejściu otrzymuje przykład  $\mathbf{x}$ , a na wyjściu podaje jego klasę  $\hat{d}$ .

Prezentacja przykładu na wejściu nazywa się **zapytaniem**, a jego klasyfikacja przez ucznia – **odповідzią na zapytanie**.

Zbiór trenujący – zbiór przykładów etykietowanych  $T = \{\langle \mathbf{x}_1, d_1 \rangle, \langle \mathbf{x}_2, d_2 \rangle, \dots, \langle \mathbf{x}_N, d_N \rangle\}$

L.p.	$\mathbf{x}$		Klasa $d$
	$x_1$	$x_2$	
1	2.65	1.26	1
2	-34.89	3.56	2
...	...	...	...
$N$	5.87	-7.94	4

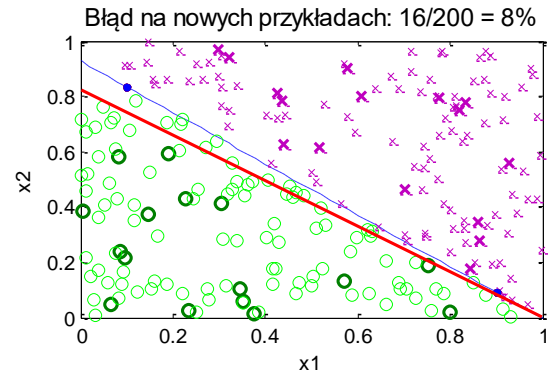
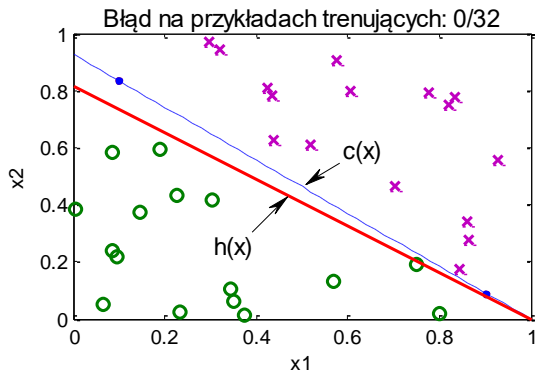
L.p.	$\mathbf{x}$				Klasa $d$
	$x_1$	$x_2$	$x_3$	$x_4$	
1	słoneczna	zimna	duża	silny	brzydka
2	słoneczna	umiark.	normalna	słaby	ładna
...	...	...	...	...	...
$N$	deszczowa	zimna	duża	silny	brzydka

# BŁĄD W UCZENIU SIĘ POJĘĆ

**Błąd klasyfikacji** – stosunek liczby niepoprawnie sklasyfikowanych przykładów do liczby wszystkich przykładów:

$$E(h) = \frac{|\{\mathbf{x} \in P \mid h(\mathbf{x}) \neq d\}|}{|P|}$$

gdzie:  $|\cdot|$  – moc zbioru,  $P \in X$  – zbiór przykładów,  $h(\mathbf{x})$  – klasa przypisana przez ucznia.



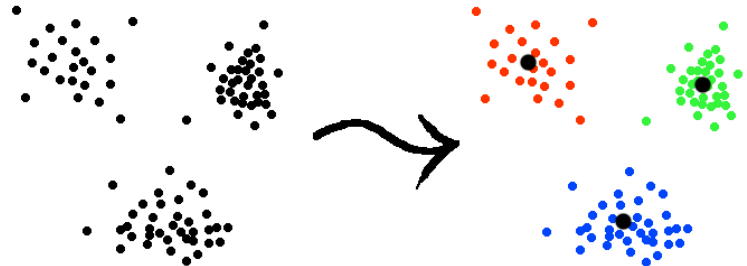
**Uczenie bez nadzoru** – klasy przykładów są nieznane (przykłady nieetykietowane).

Uczeń grupuje samodzielnie zaobserwowane przykłady w grupy zgodnie z pewnymi kryteriami podobieństwa.

**Hipoteza** – funkcja odwzorowująca przykłady na zbiór grup –  $h : X \rightarrow C_h$ , gdzie  $C_h$  jest zbiorem grup utworzonym przez ucznia.

Hipoteza określa:

- jak są tworzone grupy
- jak do grup przypisane są przykłady



Odpowiedzią na zapytanie jest wskazanie przez ucznia grupy przykładu prezentowanego na wejściu

**Zbiór trenujący** – zbiór przykładów nieetykietowanych  $T = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$



# UCZENIE SIĘ APROKSYMACJI FUNKCJI

Uczenie się funkcji odwzorowującej przykłady na zbiór liczb rzeczywistych.

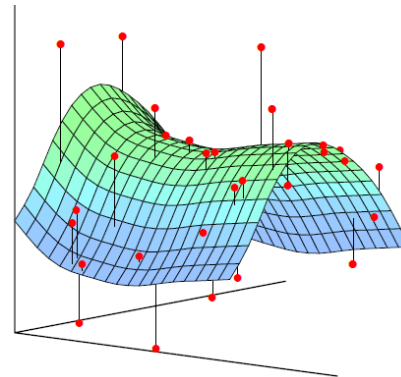
Przykłady trenujące – **argument funkcji + wartość funkcji**:  $\langle \mathbf{x}, y \rangle$ , gdzie  $y = f(\mathbf{x}) + \varepsilon$ ,  $\varepsilon$  – błąd

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

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

**Zbiór trenujący** – zbiór przykładów etykietowanych  $T = \{\langle \mathbf{x}_1, y_1 \rangle, \langle \mathbf{x}_2, y_2 \rangle, \dots, \langle \mathbf{x}_N, y_N \rangle\}$

L.p.	$\mathbf{x}$		$y$
	$x_1$	$x_2$	
1	5.5	6.0	275.375
2	-4.5	-7.0	-261.625
...	...	...	...
$N$	0.5	8.0	257.125



# UCZENIE SIĘ APROKSYMACJI FUNKCJI

## Błędy aproksymacji

Średni błąd względny:

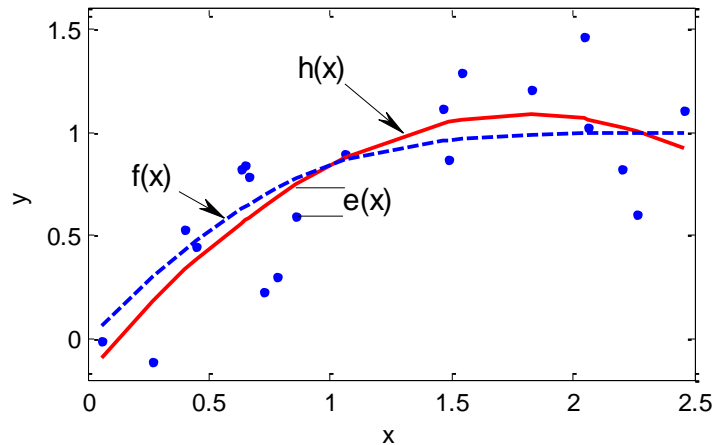
$$E(h) = \frac{1}{|P|} \sum_{\mathbf{x} \in P} \frac{|f(\mathbf{x}) - h(\mathbf{x})|}{f(\mathbf{x})} = \text{mean} \left( \frac{|e(\mathbf{x})|}{f(\mathbf{x})} \right)$$

Błąd średniokwadratowy (MSE):

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

gdzie:  $|\cdot|$  – moc zbioru,  $P \in X$  – zbiór przykładów,  $y$  – wartość pożądaney odpowiedzi,  $h(\mathbf{x})$  – odpowiedź ucznia.

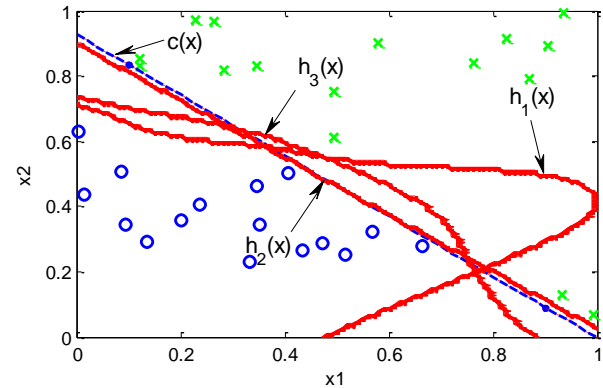
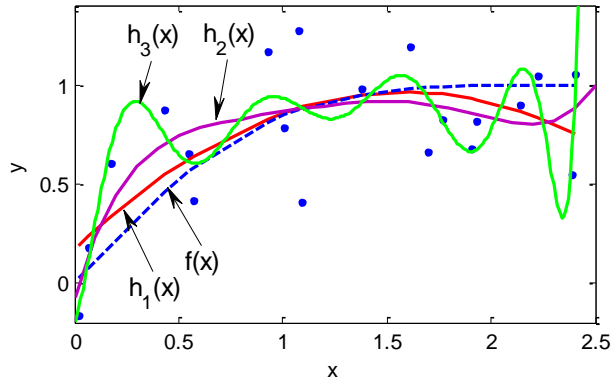
Ponieważ zwykle nie znamy prawdziwej wartości funkcji docelowej  $f(\mathbf{x})$ , w jej miejsce w powyższych wzorach podstawiamy  $y$  ( $y = f(\mathbf{x}) + \varepsilon$ )



# PROBLEM NADMIERNEGO DOPASOWANIA

**Generalizacja** (uogólnianie) – zdolność SUS do poprawnych odpowiedzi na przykłady spoza zbioru trenującego. Aby osiągnąć najlepszą generalizację złożoność hipotezy powinna odpowiadać złożoności pojęcia/funkcji docelowej.

Hipoteza jest nadmiernie dopasowana do zbioru trenującego (**przeuczona**), gdy istnieje inna hipoteza o większym błędzie na zbiorze trenującym, ale o mniejszym błędzie generalizacji (na nowych przykładach).



# PROBLEM NADMIERNEGO DOPASOWANIA

---

Zagrożenie przeuczenia – SUS dopasowuje hipotezę do przykładów uczących, często zaszumionych. Charakterystyczne cechy pojęcia/funkcji docelowej pozostają niewykryte.

Przeuczeniu sprzyja:

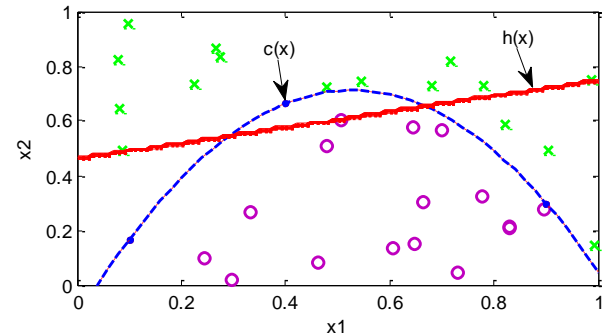
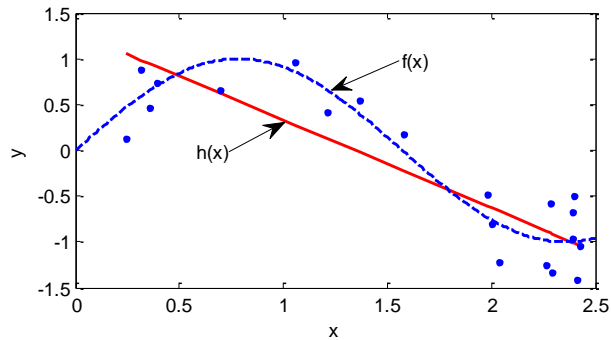
- nadmiernie złożony (elastyczny) model (bogate przestrzenie hipotez)
- deficyt danych

Zapobieganie nadmiernemu dopasowaniu:

- kontrola dopasowania hipotezy na przykładach walidacyjnych
- uwzględnienie złożoności hipotez w kryterium wyboru najlepszej hipotezy (regularyzacja)
- adaptacyjny dobór złożoności hipotezy (struktury modelu)
- komitety modeli
- dodanie składnika losowego do danych uczących
- wzbogacenie zbioru trenującego o nowe przykłady

# PROBLEM NIEDOUCZENIA

Złożoność hipotezy jest mniejsza niż złożoność funkcji docelowej.



Zapobieganie niedouczeniu:

- rozbudowanie modelu (wzbogacenie przestrzeni hipotez)
- adaptacyjny dobór złożoności hipotezy (struktury modelu)

## SELEKCJA NAJLEPSZEJ HIPOTEZY

---

Zbiór przykładów dzielimy na trzy rozłączne podzbiory:

- zbiór trenujący,
- zbiór walidacyjny i
- zbiór testowy.

Przykłady do tych zbiorów wybieramy losowo, np. 50% przykładów do zbioru trenującego, 25% – do zbioru walidacyjnego, 25% – do zbioru testowego.

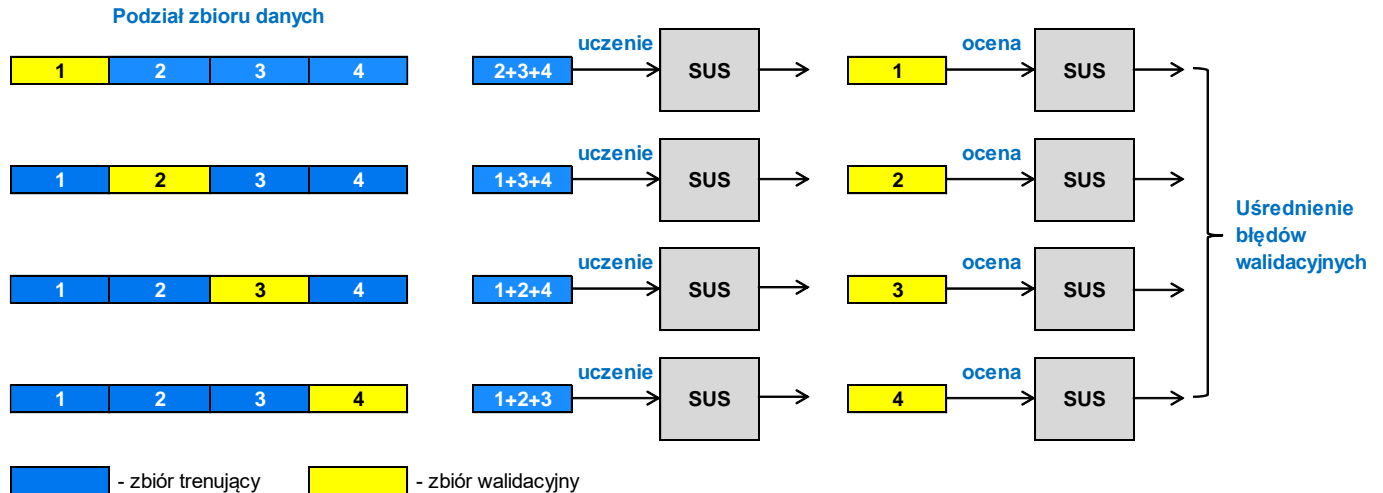
Użycie zbiorów:

- SUS uczy się na zbiorze trenującym.
- Błąd generalizacji hipotezy w trakcie uczenia szacujemy na zbiorze walidacyjnym, różnym od zbioru trenującego.
- Ostateczny błąd generalizacji SUS mierzymy na zbiorze testowym.

Hipoteza o najmniejszym błędzie testowym jest najlepsza.

# ESTYMACJA BŁĘDU GENERALIZACJI

**Kroswalidacja** – procedura uczenia i oceny SUS, w której zbiór przykładów dzieli się losowo na  $m$  równolicznych, rozłącznych podzbiorów. Następnie kolejno każdy z tych podzbiorów bierze się jako zbiór walidacyjny, a pozostałe razem jako zbiór trenujący, którym uczy się SUS. Błąd generalizacji szacuje się uśredniając błędy walidacyjne obliczane po każdej z  $m$  sesji uczących.



Postać modelu:

$$g(\mathbf{x} | \theta)$$

gdzie:  $\mathbf{x}$  – wejścia,  $\theta$  – parametry.

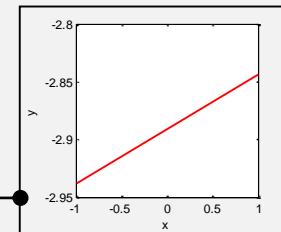
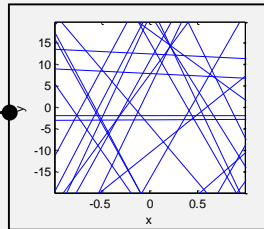
$g(\cdot)$  reprezentuje klasę hipotez  $H$ . Konkretnie wartości parametrów  $\theta$  reprezentują konkretną hipotezę  $h \in H$ .

*Przykład: model liniowy*

Klasa hipotez:  $g(\mathbf{x} | \theta) = ax + b$

$$\theta = [a, b]$$

Konkretna hipoteza:  $g(\mathbf{x} | [0.0472, -2.8912]) = 0.0472 x - 2.8912$



Postać modelu ustala projektant na podstawie przewidywanej postaci funkcji docelowej i własnych doświadczeń (obciążenie indukcyjne).



## Kryterium oceny modelu:

- błąd generalizacji
- złożoność (liczba parametrów)

## Procedura optymalizacji (uczenia) modelu:

Znaleźć wartości parametrów minimalizujące kryterium oceny

$$\theta^* = \arg \min_{\theta} E(\theta | X)$$

Metody optymalizacji:

- Analityczne (w zadaniach regresji)
- Gradientowe
- Stochastyczne (algorytmy ewolucyjne i rojowe)
- ...

