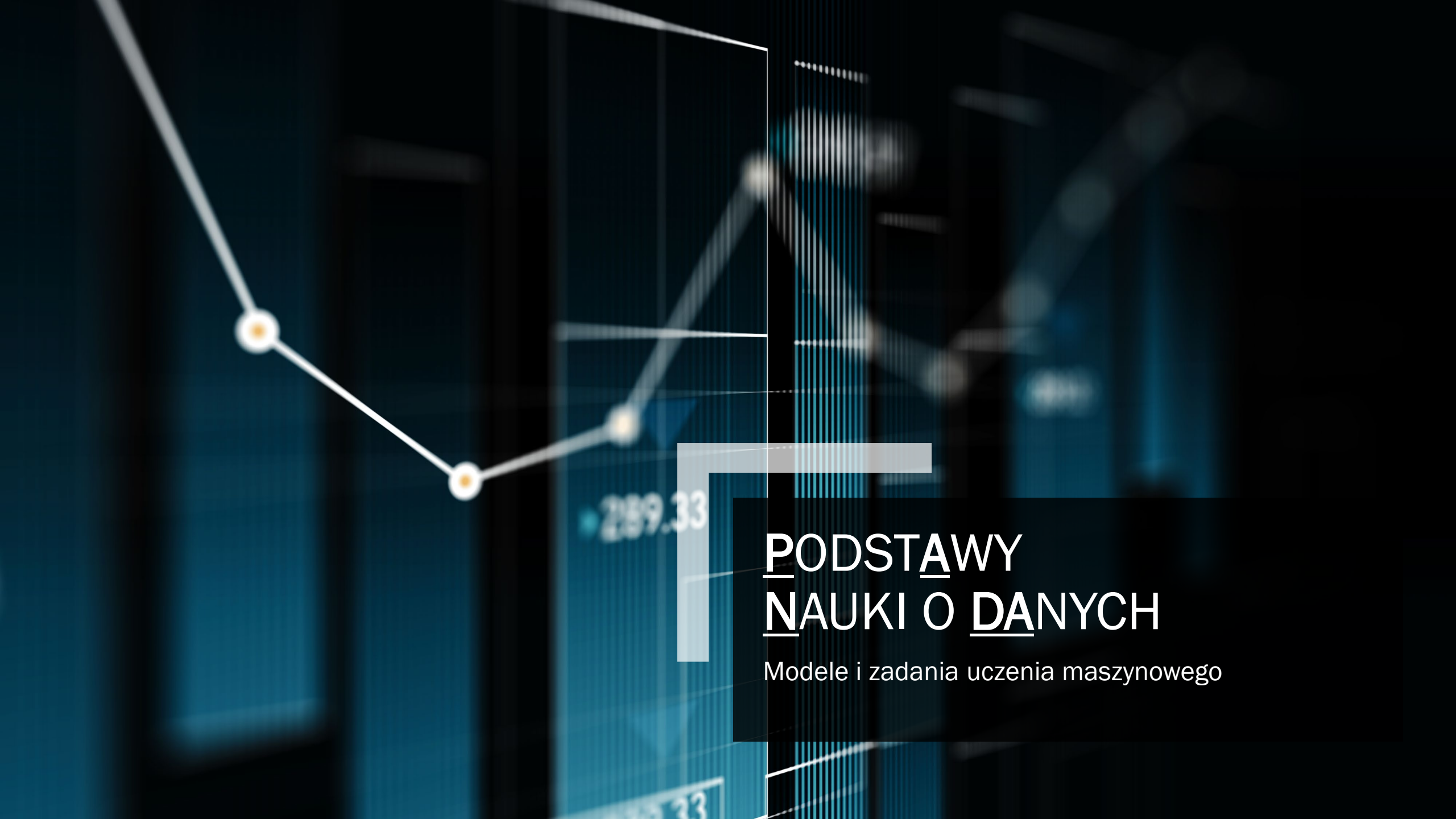


PODSTAWY NAUKI O DANYCH

Modele i zadania uczenia maszynowego

Dane

„uczenie"

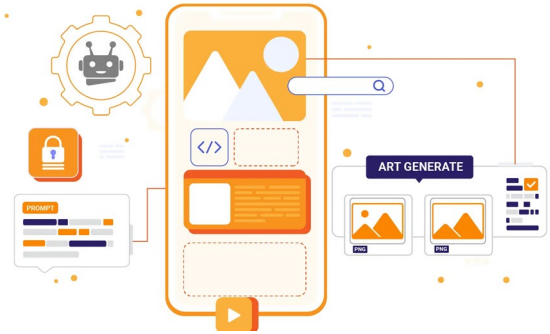
Modele danych

wnioskowanie

Wiedza

wdrażanie

Rzeczywisty proces



Aplikacje

Czym jest uczenie maszynowe

- ang. *machine learning* *learning* ≠ uczenie

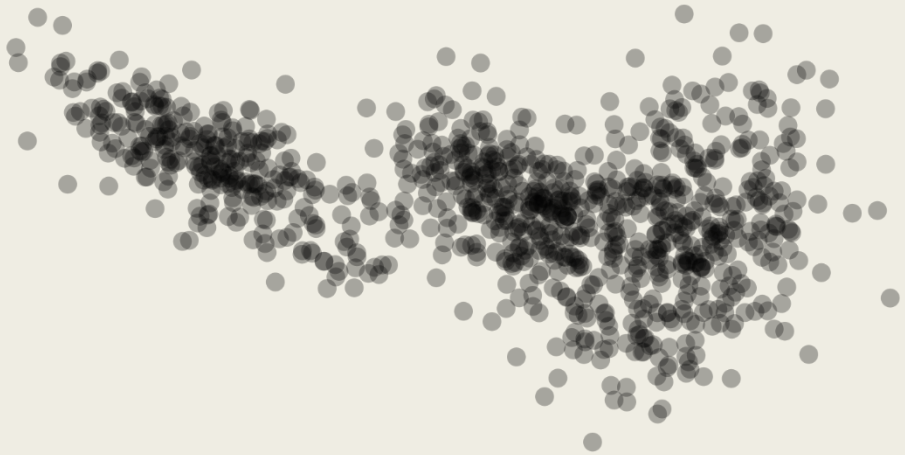
I've learned that you are a confident person

- modelowanie statystycznych regularności w obserwacjach / pomiarach
- modelowanie danych ≠ modelowanie procesu
- dopasowanie modelu do danych (zadanie estymacji)
- **Def.** Proces automatycznego poszukiwania lepszej **reprezentacji danych**

  w ramach zdefiniowanej przestrzeni możliwości

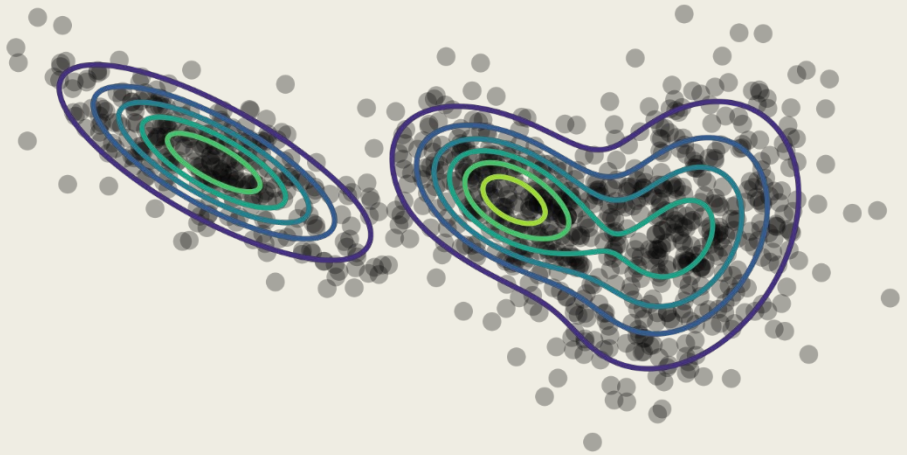
  na podstawie sygnału **informacji zwrotnej**

PRZYKŁADY ZADAŃ UCZENIA MASZYNOWEGO



Mieszanina rozkładów Gaussa

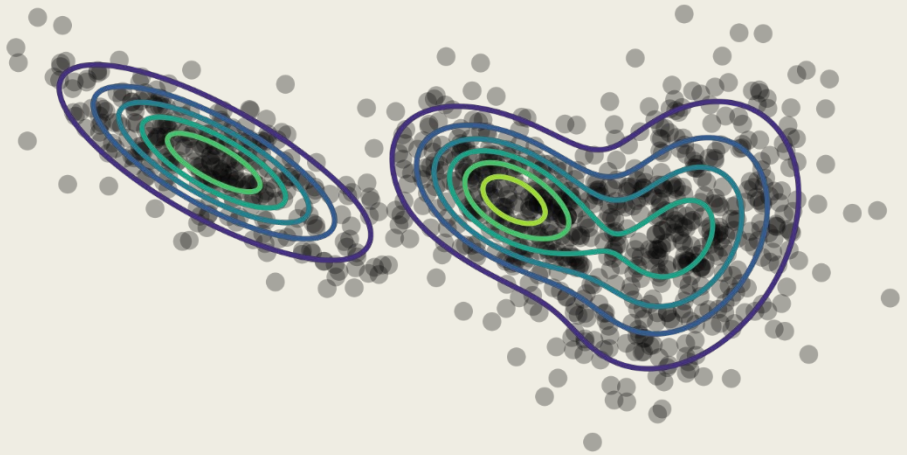
$$p(\mathbf{x}|\theta) = \sum_{m=1}^{M} \pi_m \, \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \qquad 0 \leq \pi_m \leq 1, \quad \sum_{m=1}^{M} \pi_m = 1$$

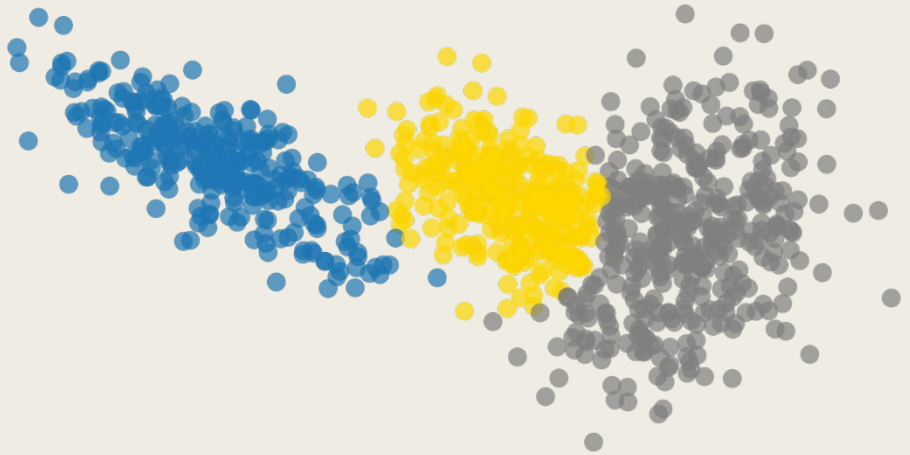


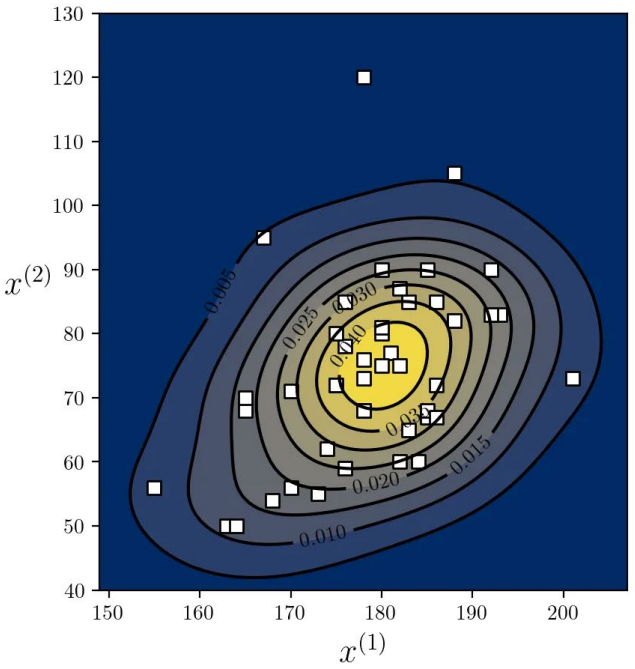
Mieszanina rozkładów Gaussa

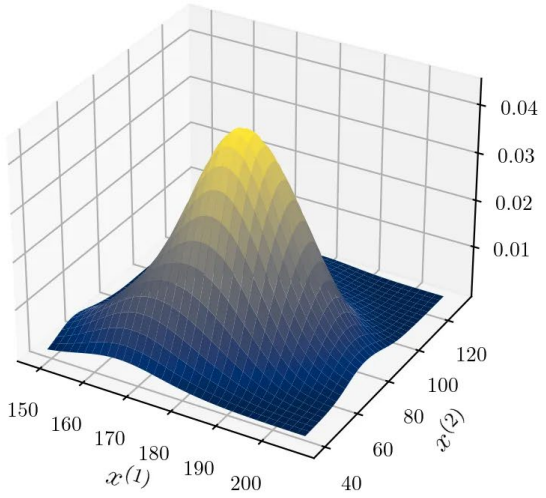
$$p(\mathbf{x}|\theta) = \sum_{m=1}^{M} \pi_m \, \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$$

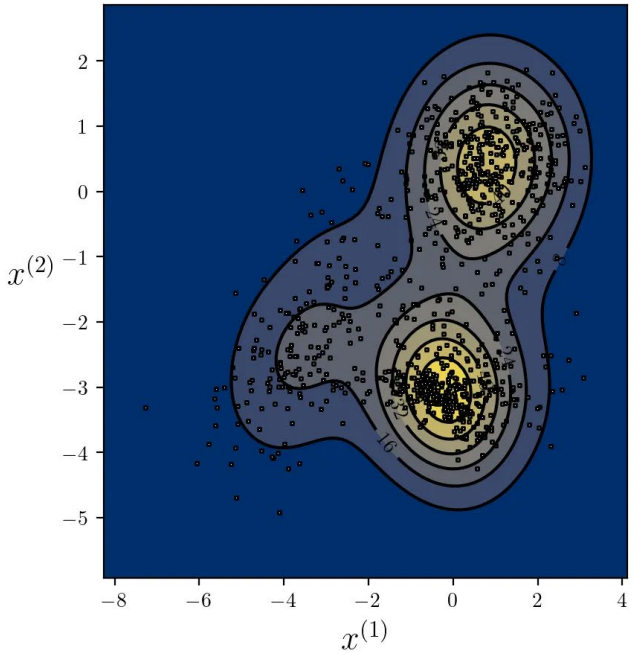
$$0 \leq \pi_m \leq 1, \quad \sum_{m=1}^{M} \pi_m = 1$$

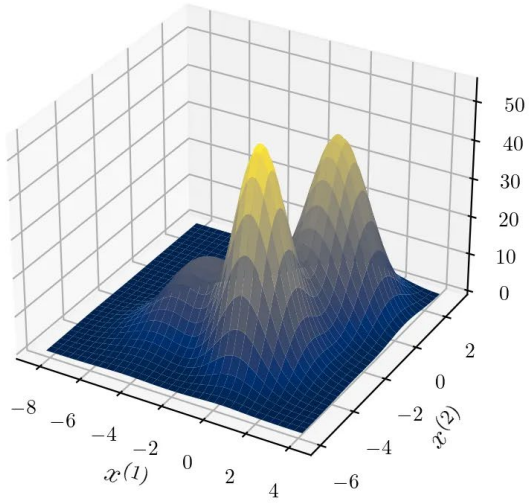





$x^{(2)}$
$x^{(1)}$
0.035
0.030
0.025
0.020
0.015
0.010


$x^{(1)}$
$x^{(2)}$


$x^{(2)}$
$x^{(1)}$


$x^{(1)}$
$x^{(2)}$


Medium

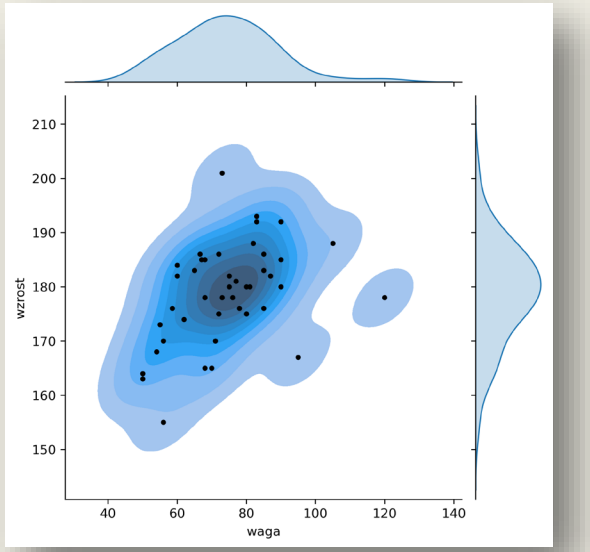
Przypominajka statystyczna

$p(X;\theta)$ – rozkład zmiennej losowej X, którego parametrem jest θ

zwróć uwagę na tę różnicę

$p(X,\theta)$ – **rozkład łączny** dwóch zmiennych losowych X oraz θ

ang. *joint probability distribution*



$$p(X,\theta) = p(\theta,X)$$

$p(X|\theta)$ – **rozkład warunkowy** zmiennej losowej X pod warunkiem θ

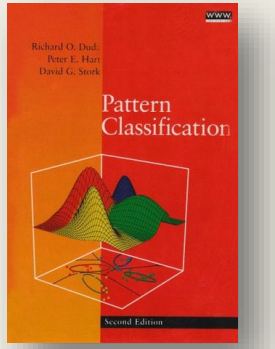
ang. *conditional PDF*

$p(\theta|X)$ – **rozkład warunkowy** zmiennej losowej θ pod warunkiem X

Probability Distribution Function

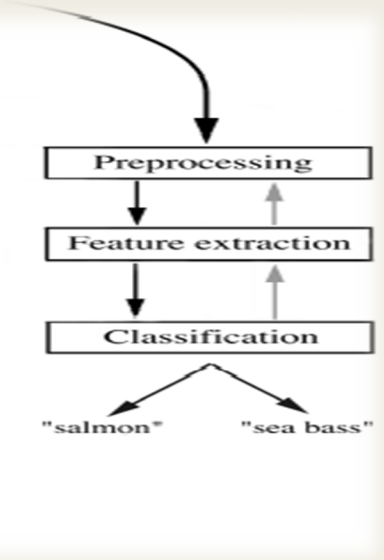
Probability Density Function > *PDF*

Klasyfikacja

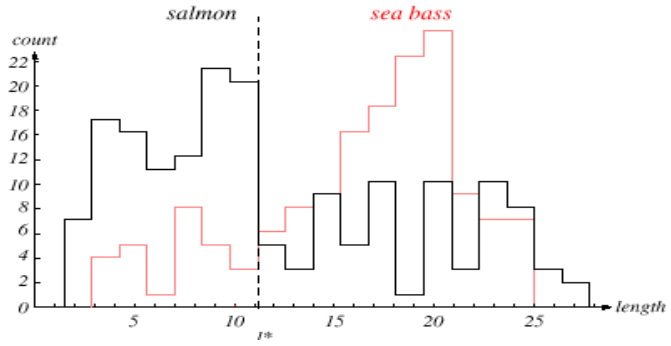


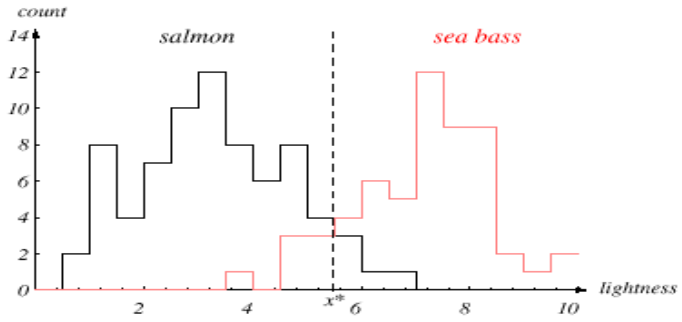
Cechy:

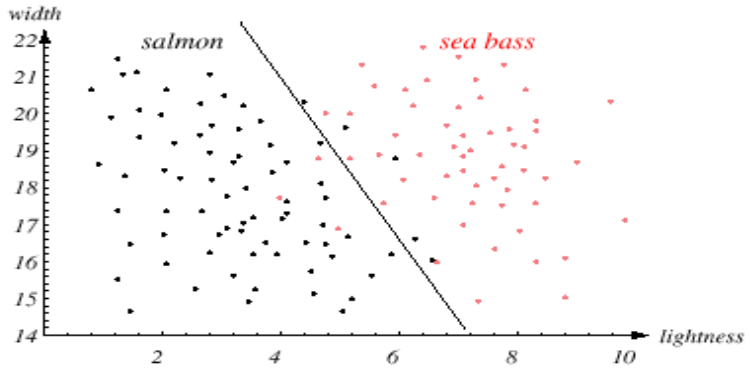
- ✓ *długość*
- ✓ *jasność*
- ✓ *szerokość*
- ✓ *liczba i kształt płetw*
- ✓ *położenie otworu gębowego*

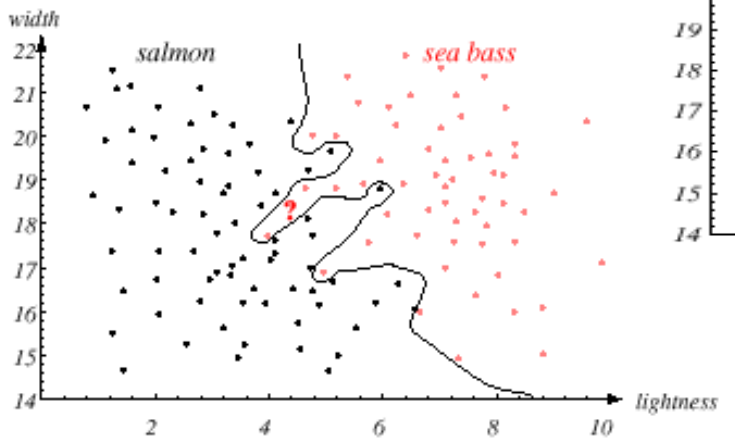


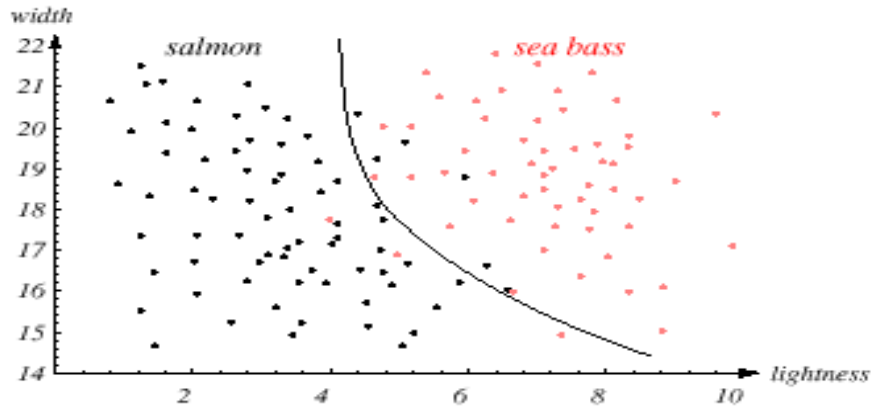
Klasyfikacja

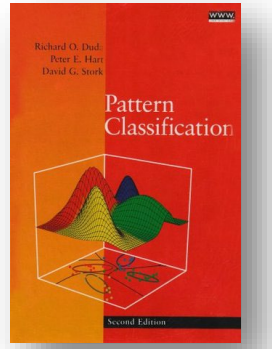




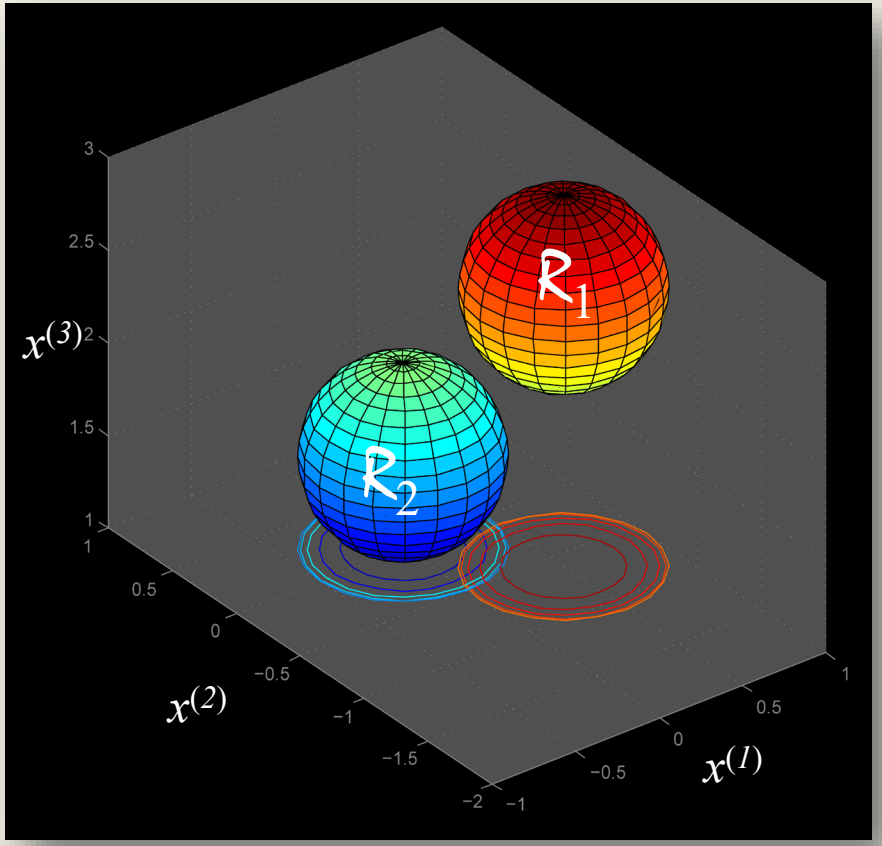


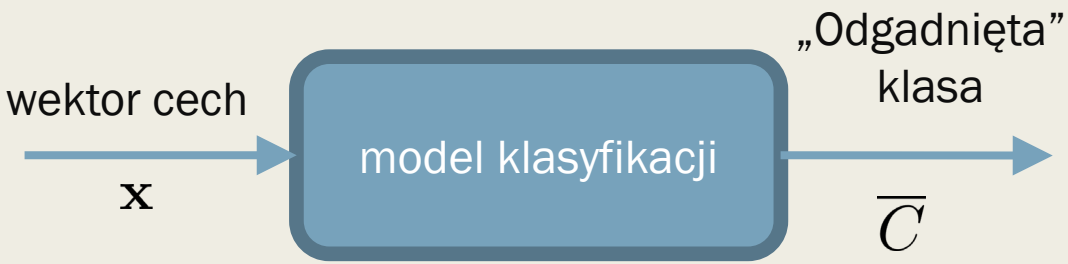




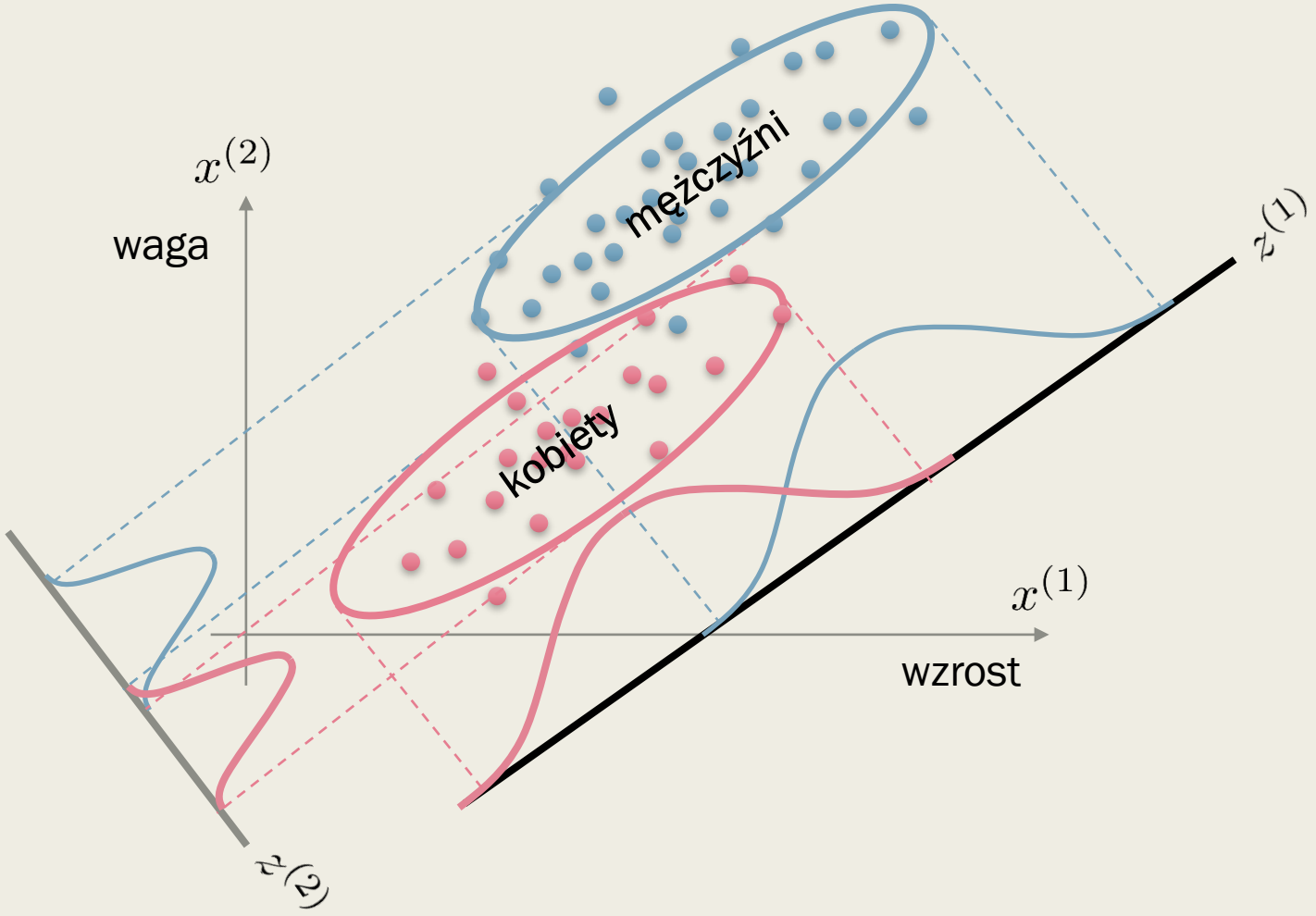


Klasyfikacja





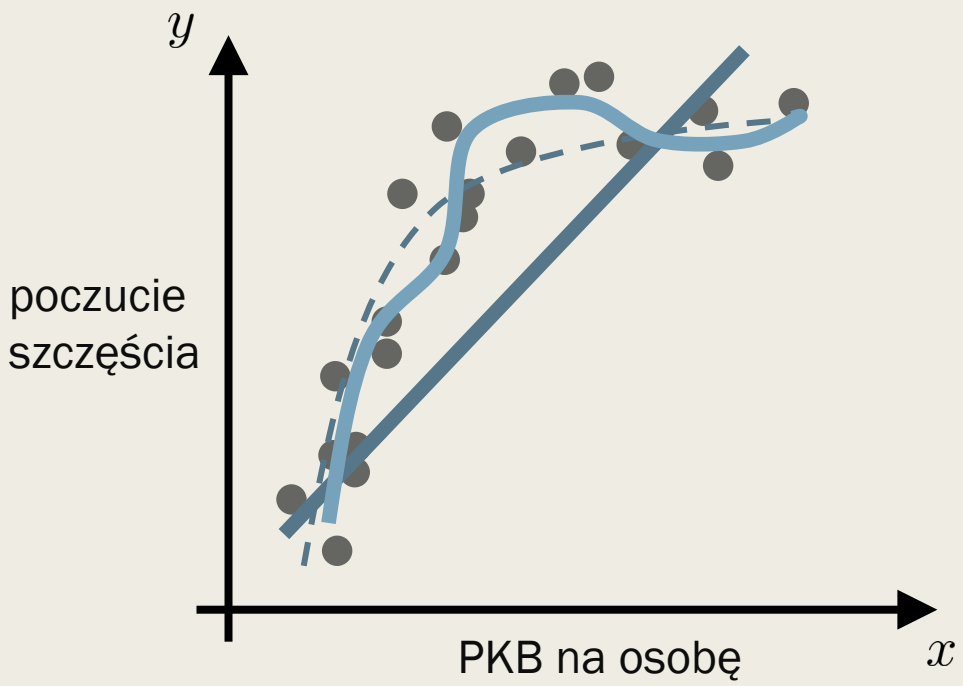
Transformacje

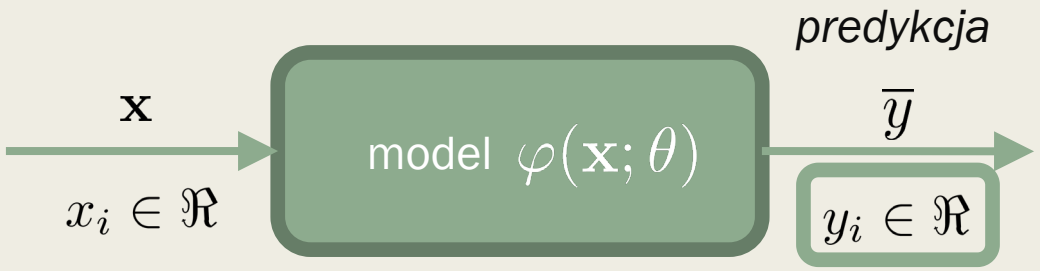


Przykłady modeli

- *k*-najbliższych sąsiadów
- klasyfikator liniowy
- regresja logistyczna
- maszyny wektorów wspierających (ang. *Support Vector Machines*)
- drzewa decyzyjne (lasy)
- sieci neuronowe
- ...

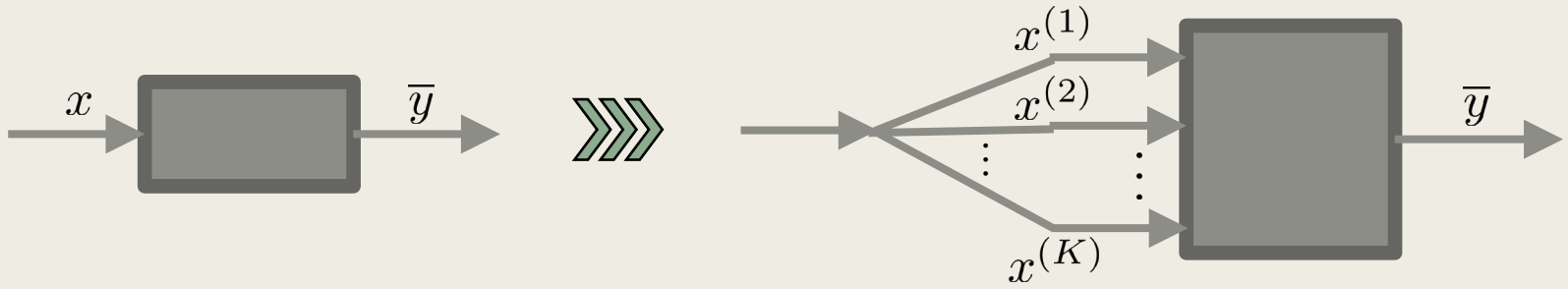
Regresja





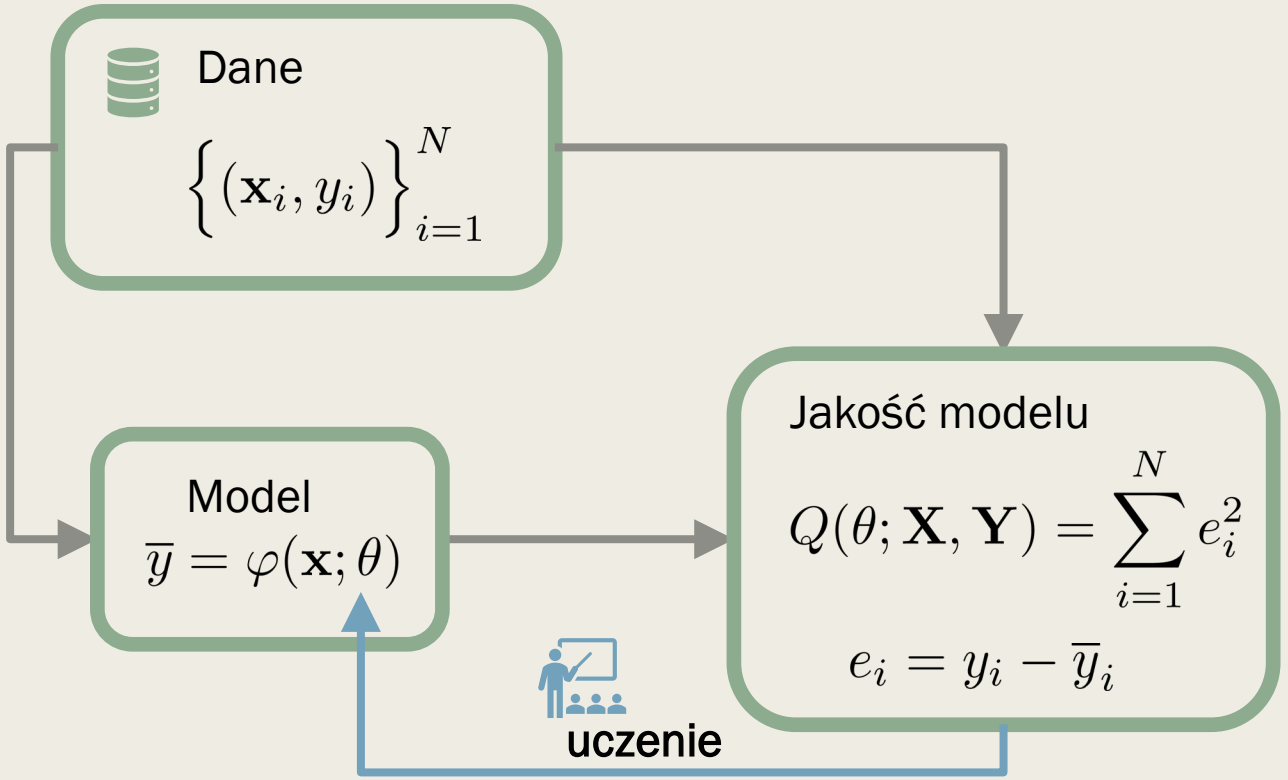
→ *Model odniesienia / naiwny*

Uogólnione modele liniowe



Regresja

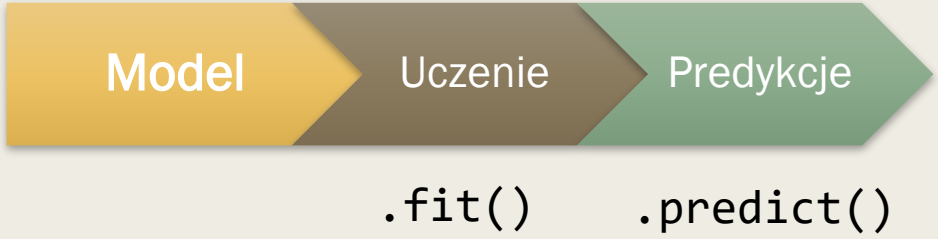
Dopasowanie modelu do danych = **uczenie**



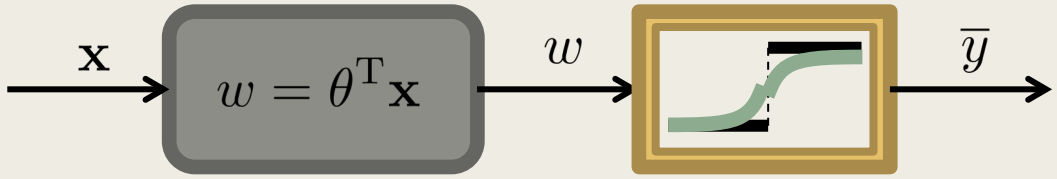
$$\widehat{\theta}_{\mathrm{LS}} = \arg\min_{\theta} Q(\theta; \mathbf{X}, \mathbf{Y})$$

Przykłady modeli

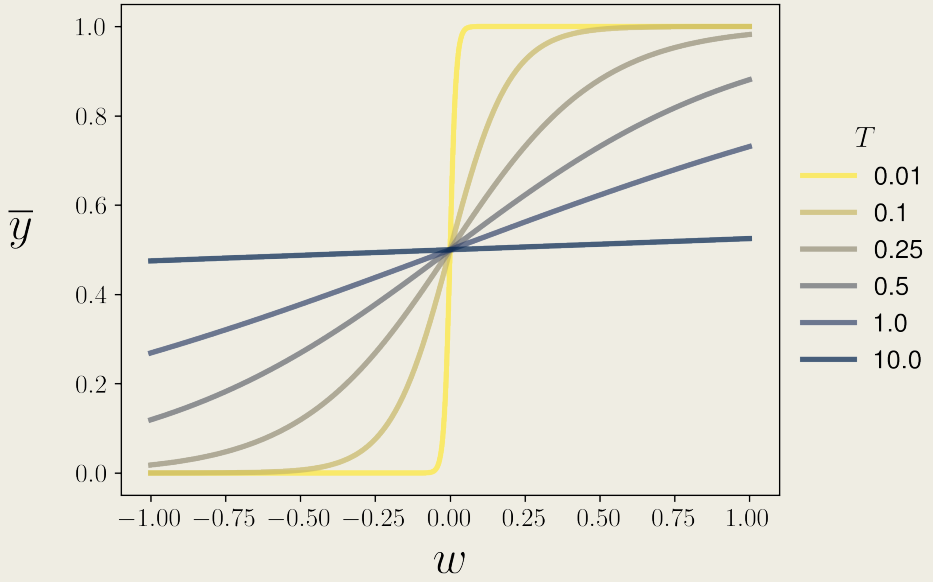
- regresja liniowa
- uogólnione modele liniowe (ang. *Generalized linear models*)
- drzewa decyzyjne (lasy)
- regresja bayesowska
- maszyny wektorów wspierających (ang. *Support Vector Machines*)
- sieci neuronowe
- ...



Model regresji jako klasyfikator

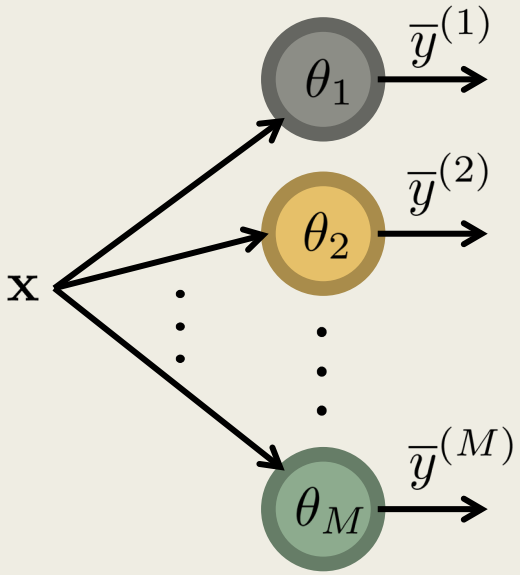


$$\overline{y} = \frac{1}{1 + e^{-w}}$$



Model regresji jako klasyfikator

Wiele klas



0.12

0.08

0.93

$$\overline{y}^{(m)} = \frac{1}{1 + e^{-\theta_m^{\mathrm{T}}\mathbf{x}}} \qquad (m = 1, 2, \ldots, M)$$

vs

Softmax $$\overline{y}^{(m)} = \frac{e^{\theta_m^{\mathrm{T}}\mathbf{x}}}{\sum_{l=1}^{M} e^{\theta_l^{\mathrm{T}}\mathbf{x}}} \qquad (m = 1, 2, \ldots, M)$$

Ujęcie probabilistyczne klasyfikacji

wektor cech

$$\mathbf{x} = \begin{bmatrix} x^{(1)} \\ \vdots \\ x^{(K)} \end{bmatrix}$$

etykiety klas

$$\{C_1, C_2, \cdots, C_M\}$$

$p(\mathbf{x}, C)$

$p(\mathbf{x})$ – rozkład cech

$p(C)$ – rozkład klas

$p(\mathbf{x}|C)$ – rozkład cech w klasach

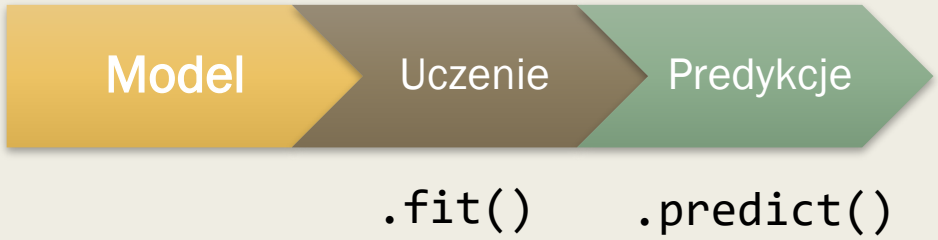
$$\begin{matrix} p\left(x^{(1)}|C\right) \\ \vdots \\ p\left(x^{(K)}|C\right) \end{matrix}$$

$$p(C|\mathbf{x}) = \frac{p(\mathbf{x}|C) \cdot p(C)}{p(\mathbf{x})}$$

→ *Modele generatywne*

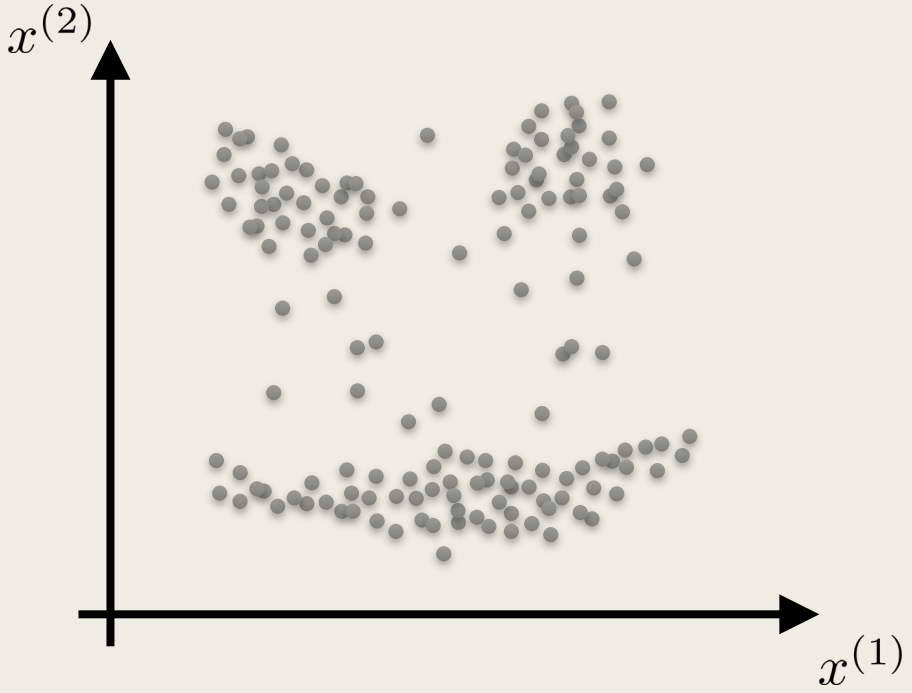
Przykłady modeli

- optymalny klasyfikator Bayesa
- naiwny klasyfikator Bayesa (ang. *Naive Bayes Classifier*)
- ukryte modele Markowa (ang. *Hidden Markov Models*)
- ...



Grupowanie

$$\left\{\mathbf{x}_i\right\}_{i=1}^{N}$$



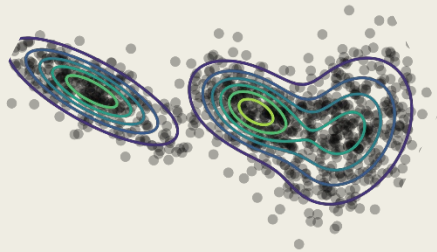
Przykłady modeli / algorytmów

- *k*-średnich (ang. *k-means*) / centroidów
- mieszanina rozkładów Gaussa
  - *maksymalizacja oczekiwania (ang. **E**xpectation–**M**aximization)*
- DBSCAN
- widmowa analiza skupień
- grupowanie twarde i miękkie
  - *rozmyte c-średnich (ang. fuzzy c-means*)
- metody aglomeracyjne

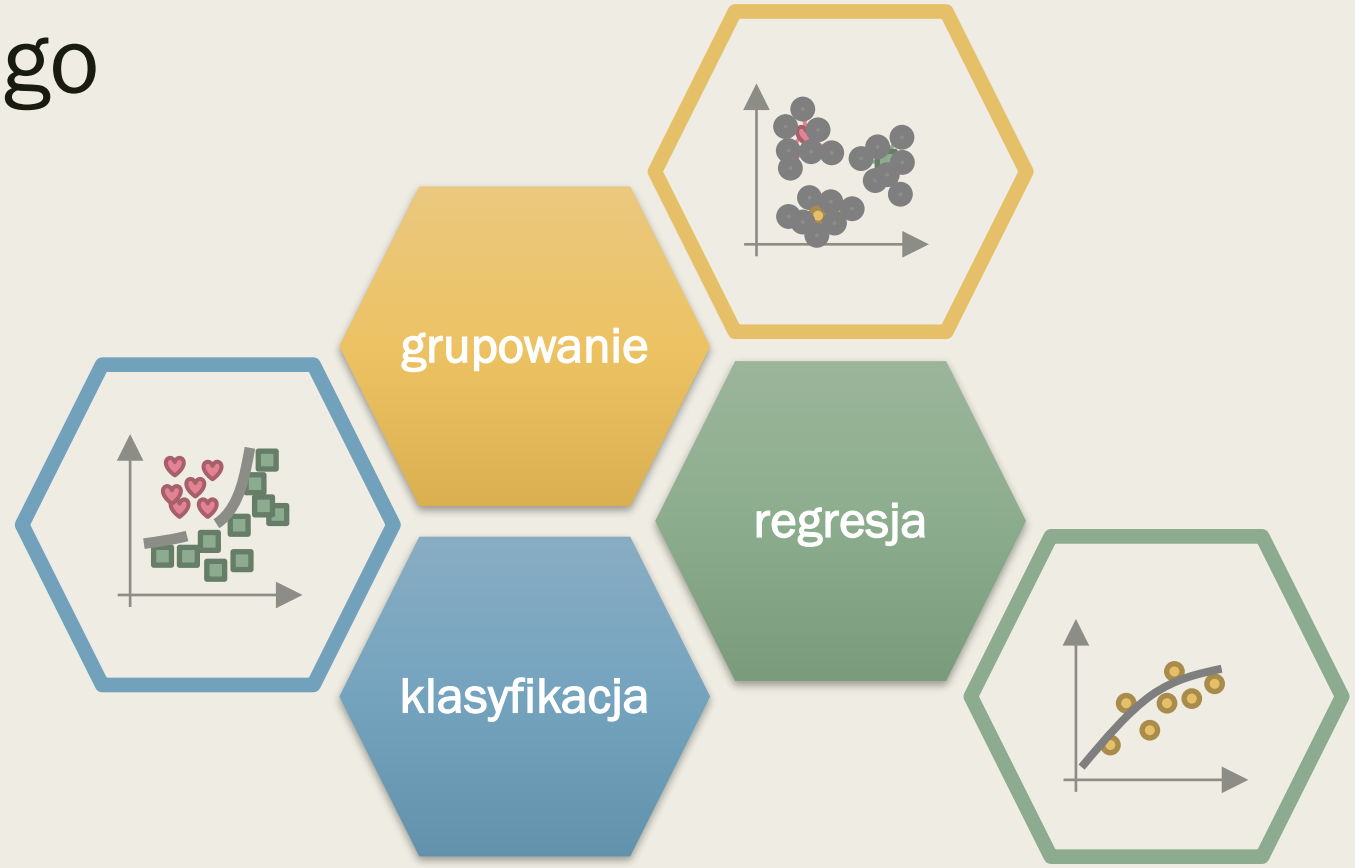
```
sklearn.mixture.GaussianMixture
                .BayesianGaussianMixture

sklearn.cluster.Kmeans
                .DBSCAN
            .AgglomerativeClustering
                .SpectralClustering
scipy.cluster.hierarchy.linkage
                       .dendrogram
```

Podstawowe zadania
uczenia maszynowego



Estymacja rozkładu



$\left\{(\mathbf{x}_i, y_i)\right\}_{i=1}^{N}$

uczenie nadzorowane
(ang. *supervised learning*)

- **etykieta**
  „prawdziwa odpowiedź”
- **regresja**
  wyjście ciągłe
- **klasyfikacja**
  wyjście kategorialne

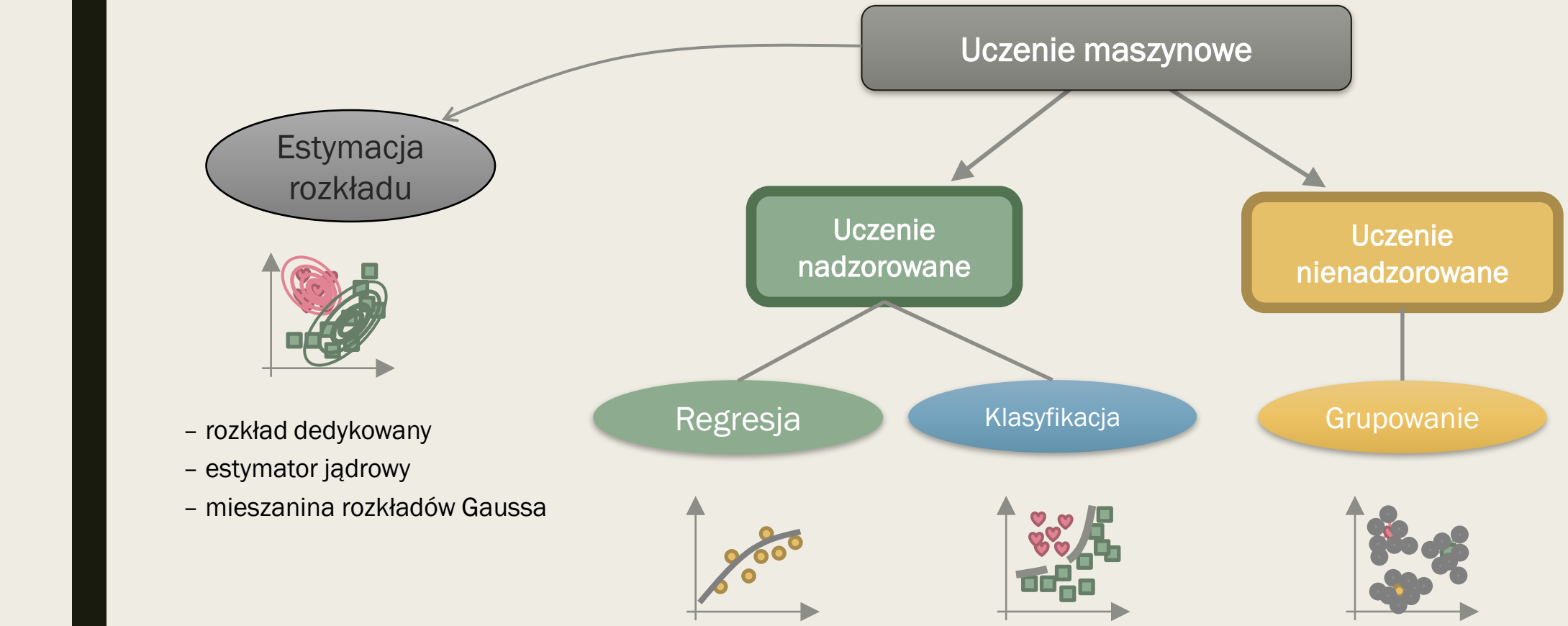
$\left\{\mathbf{x}_i\right\}_{i=1}^{N}$

uczenie nienadzorowane
(ang. *unsupervised learning*)

- brak etykiet
- **grupowanie**

uczenie półnadzorowane
(ang. *semisupervised learning*)

uczenie ze wzmocnieniem
(ang. *reinforcement learning*)



– rozkład dedykowany
– estymator jądrowy
– mieszanina rozkładów Gaussa

– regresja liniowa
– uogólnione modele liniowe
– SVR
– drzewa decyzyjne
– sieci neuronowe

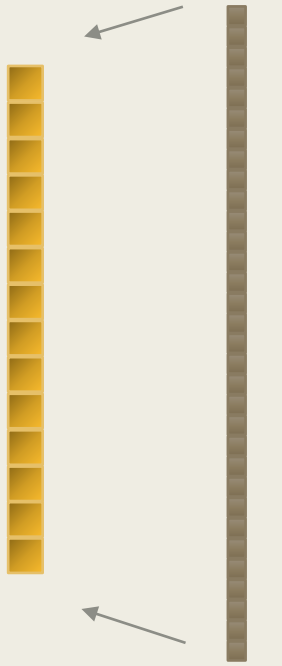
– *k*-najbliższych sąsiadów
– naiwny Bayes
– regresja logistyczna
– SVM
– drzewa decyzyjne
– lasy losowe
– sieci neuronowe

– *k*-średnich
– rozmyte c-średnich
– mieszanina rozkładów Gaussa

Uczenie głębokie

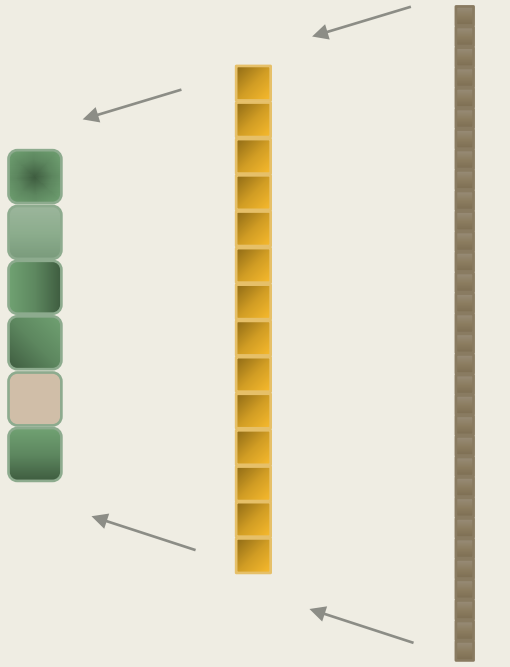
Uczenie głębokie

$$\varphi_0\left(\mathbf{W_0} \cdot \mathbf{x}\right)$$



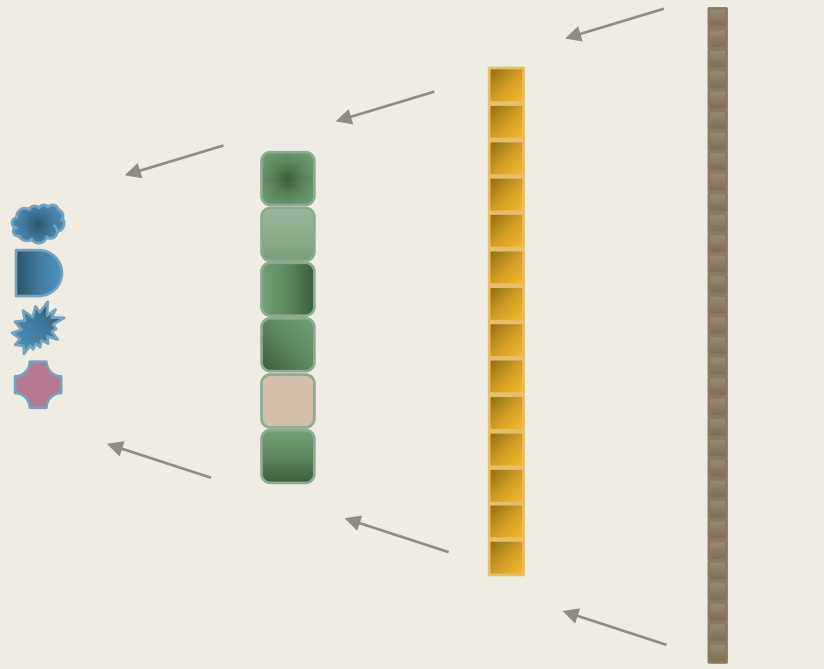
Uczenie głębokie

$$\varphi_1\left(\mathbf{W_1} \cdot \varphi_0\left(\mathbf{W_0} \cdot \mathbf{x}\right)\right)$$



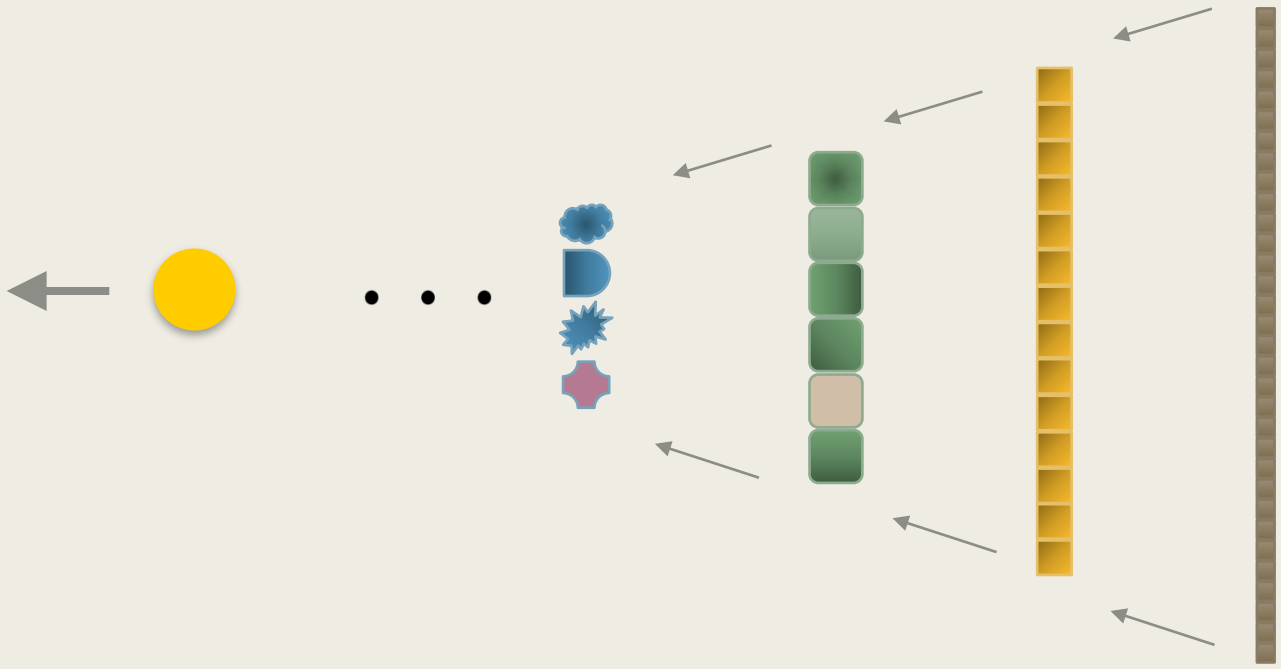
Uczenie głębokie

$$\varphi_2\Big(\mathbf{W_2}\cdot\varphi_1\Big(\mathbf{W_1}\cdot\varphi_0\left(\mathbf{W_0}\cdot\mathbf{x}\right)\Big)\Big)$$



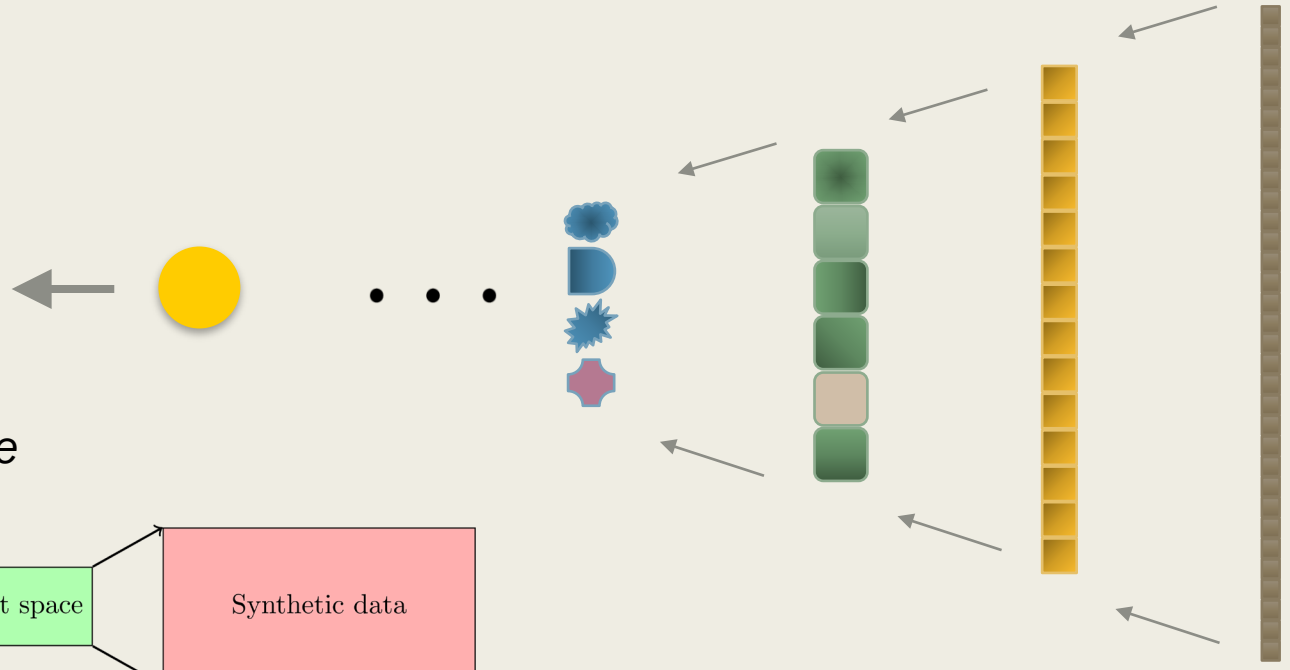
Uczenie głębokie

$$y = \varphi_K\left(\mathbf{W_K} \cdot \cdots \cdot \varphi_2\left(\mathbf{W_2} \cdot \varphi_1\left(\mathbf{W_1} \cdot \varphi_0\left(\mathbf{W_0} \cdot \mathbf{x}\right)\right)\right)\cdots\right)$$

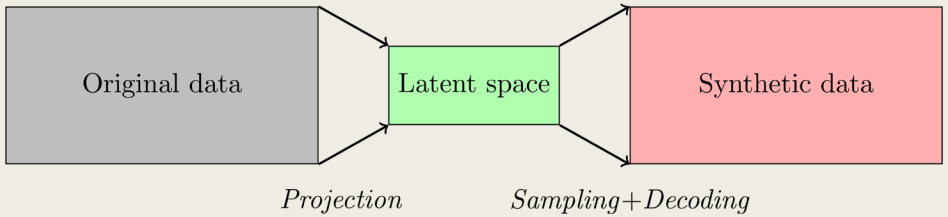


Uczenie głębokie

$$y = \varphi_K\left(\mathbf{W_K} \cdot \cdots \cdot \varphi_2\left(\mathbf{W_2} \cdot \varphi_1\left(\mathbf{W_1} \cdot \varphi_0\left(\mathbf{W_0} \cdot \mathbf{x}\right)\right)\right)\cdots\right)$$

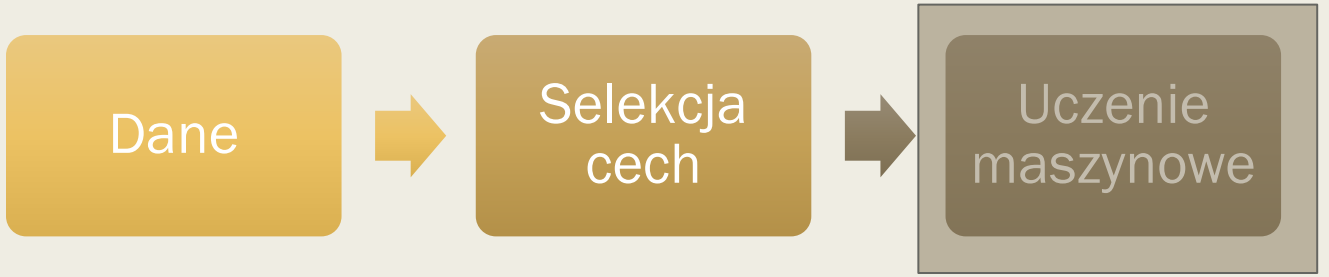


→ *Modele generatywne*



Uczenie głębokie

Uczenie „**płytkie**”

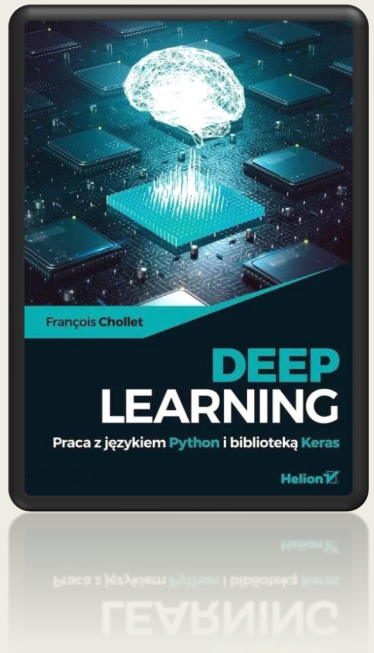


Model „płytki”

Uczenie **głębokie**

Model wielowarstwowy, hierarchiczny

Uczenie głębokie



- Uczenie **reprezentacji warstwowych i hierarchicznych**
- Model samodzielnie uczy się cech
- Uczenie transferowe
- Przestrzeń ukryta
- Zdominowane przez sieci neuronowe
- Mistyczna otoczka tego, że technologia ta przypomina działanie mózgu

Parametryczne

Ustalona liczba parametrów

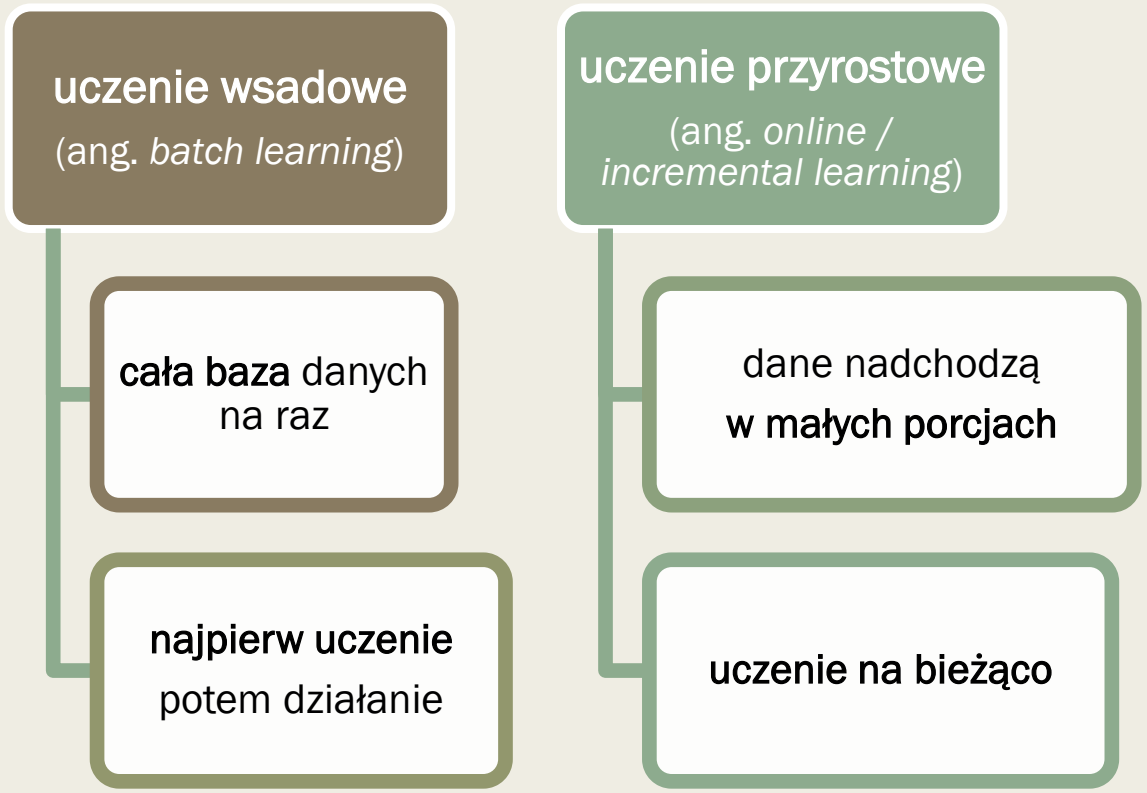
- klasyfikator liniowy
- regresja liniowa i logistyczna
- mieszanina rozkładów Gaussa
- liniowy SVM
- naiwny klasyfikator Bayesa
- ...

Nieparametryczne

Postać modelu i liczba parametrów zależą od danych

- *k*-NN
- drzewa decyzyjne
- estymatory jądrowe (KDE)
- sieci neuronowe
- ...

Sposoby korzystania z bazy danych

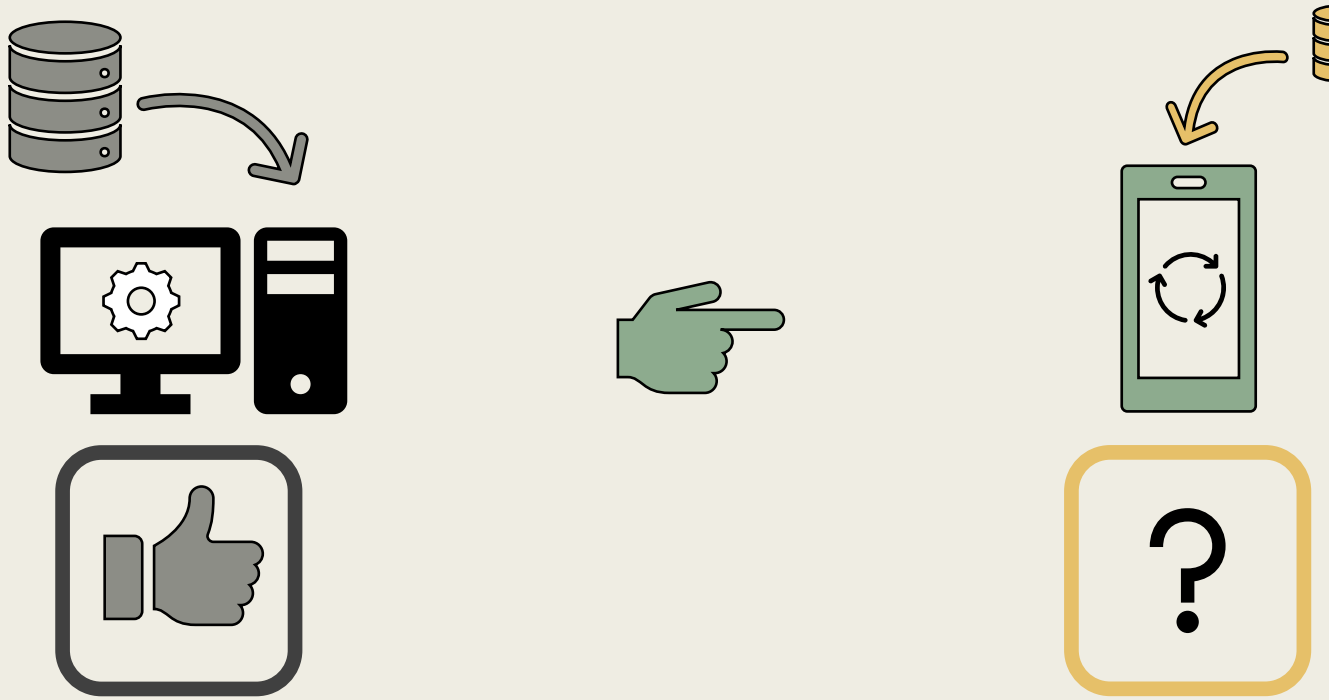



Dane
Reguły
Programowanie
(system ekspertowy)
Odpowiedzi
A
B
C


Dane
Odpowiedzi
Uczenie
maszynowe
Reguły
A
B
C

DOBRY MODEL

Jaki model jest dobry?

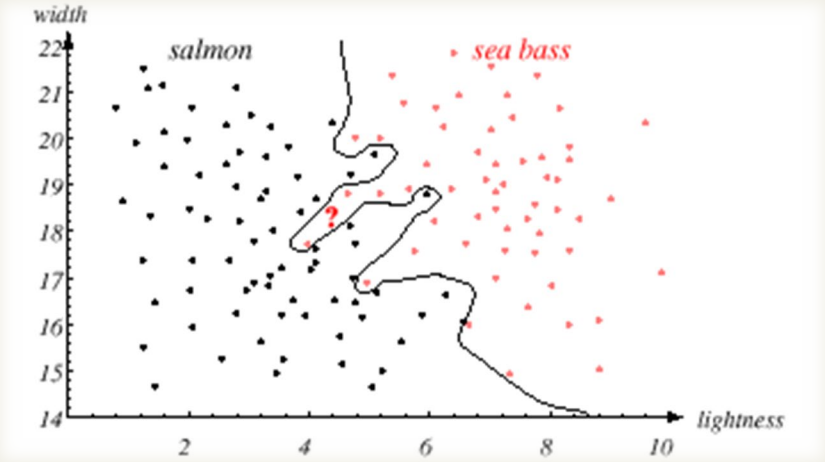


Jakość na danych treningowych ⟹ Jakość na **nowych danych**

Złożoność modelu

kompromis obciążenie – wariancja
(ang. *bias-variance dillema*)

duża wariancja

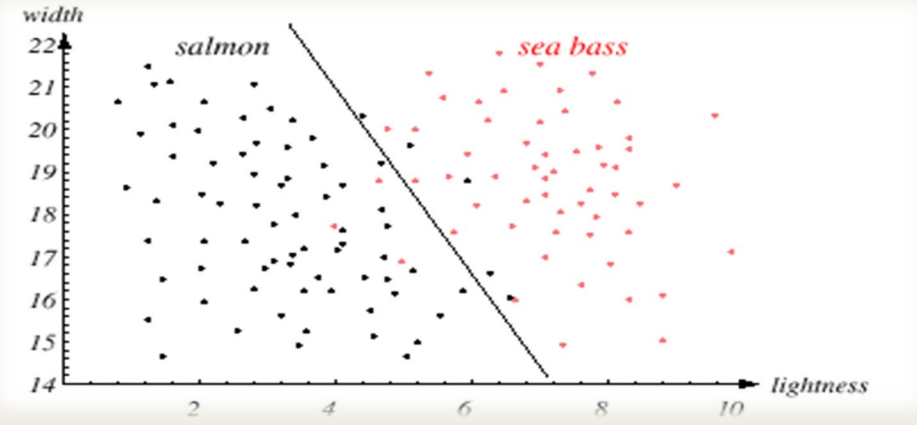


nadmiernie dopasowany do tych danych

dane

wiedza

duże obciążenie



mocno trzymający się założeń

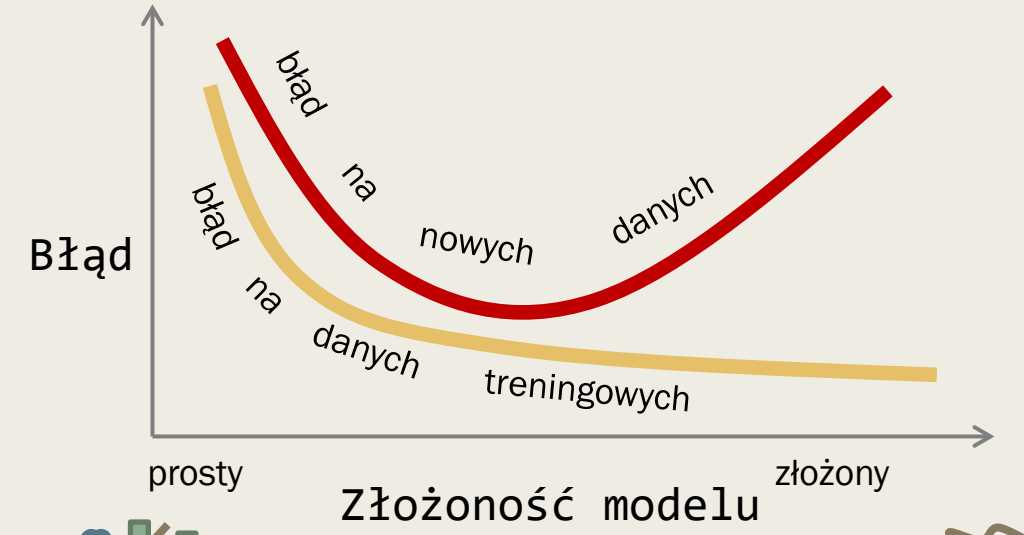
Im bardziej **model elastyczny**, tym lepiej **dopasuje się do aktualnych danych**

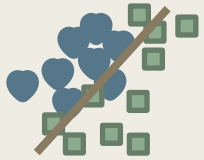
Złożoność modelu

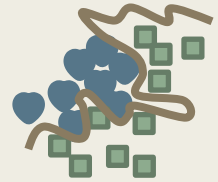
- mały błąd na danych uczących nie gwarantuje małego błędu na nowych obserwacjach
- przeuczenie / **nadmierne dopasowanie** / przetrenowanie (ang. *overfitting*)
  - *model mapuje jak słownik*
- niedouczenie / niedotrenowanie

  (ang. *underfitting*)
- błąd uogólniania

  (ang. *generalization error*)

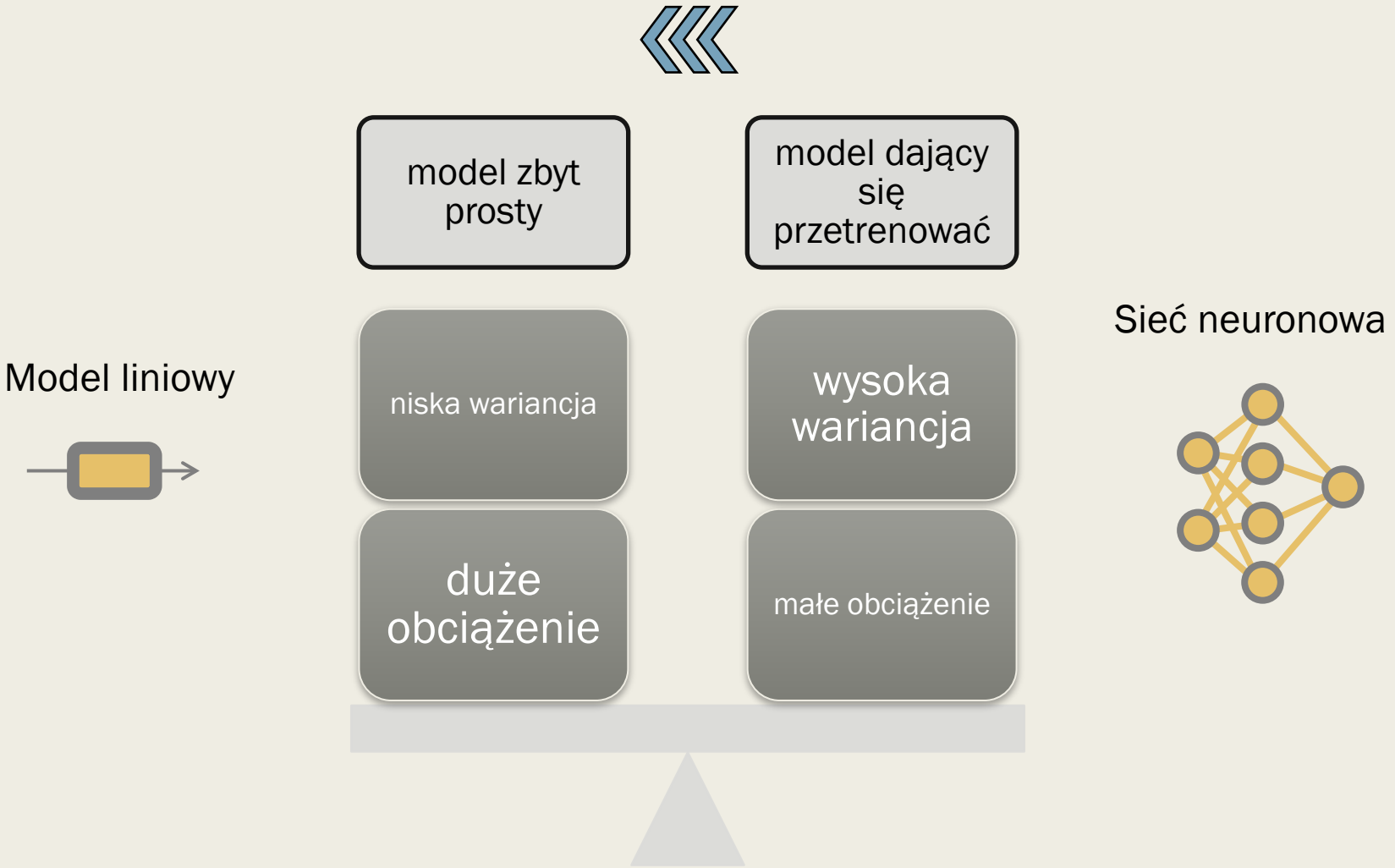






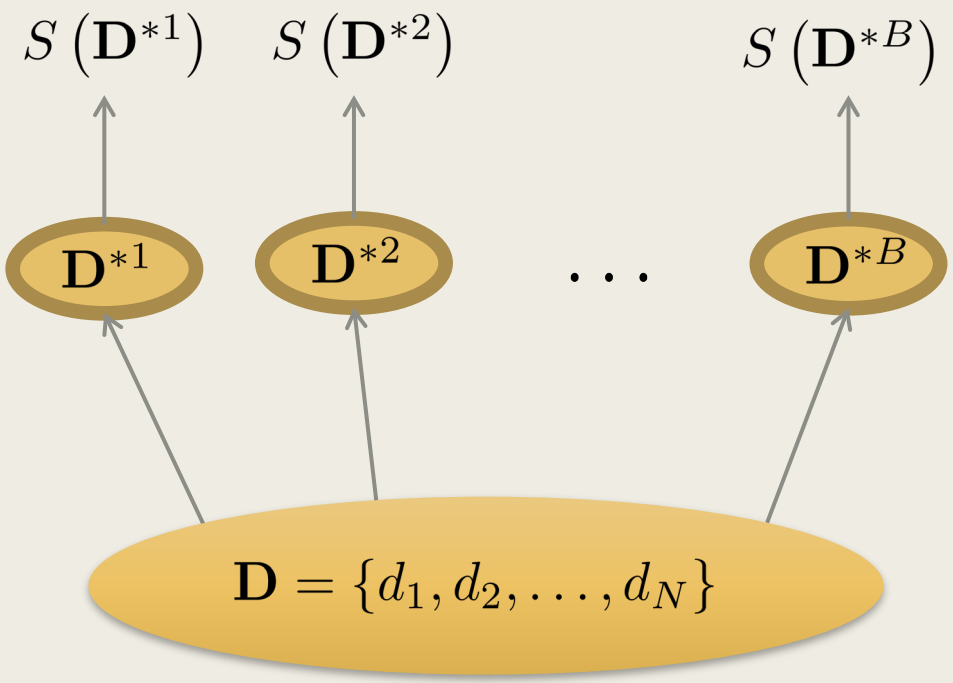
Regularyzacja

Ocena modelu = błąd dopasowania + $\lambda \cdot$ kara za złożoność



OCENA JAKOŚCI MODELU

Metoda bootstrap



replikacje *bootstrapowe*

N – elementowe próby *bootstrapowe* losowane ze zwracaniem

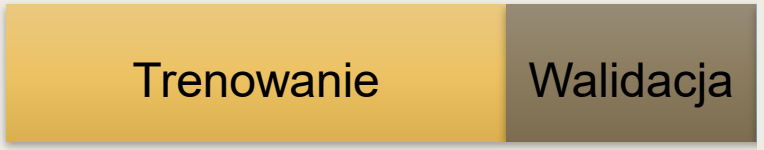
Estymacja Monte-Carlo dokładności wyznaczenia

$$\overline{S}^* = \frac{1}{B}\sum_{b=1}^{B} S(D^{*b})$$

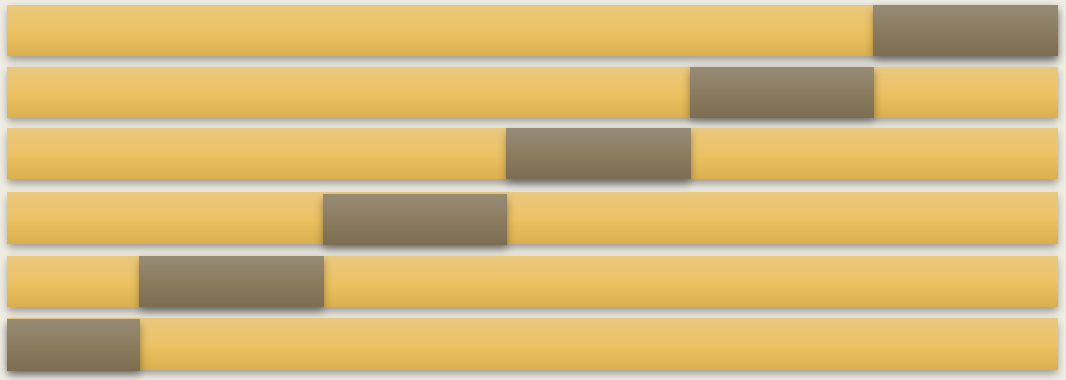
$$\widehat{\mathrm{Var}}\left[S(D)\right] = \frac{1}{B-1}\sum_{b=1}^{B}\left(S\left(D^{*b}\right) - \overline{S}^*\right)^2$$

Walidacja krzyżowa
(ang. C*ross-Validation – CV*)

- Losowy podział
- Wielokrotny podział

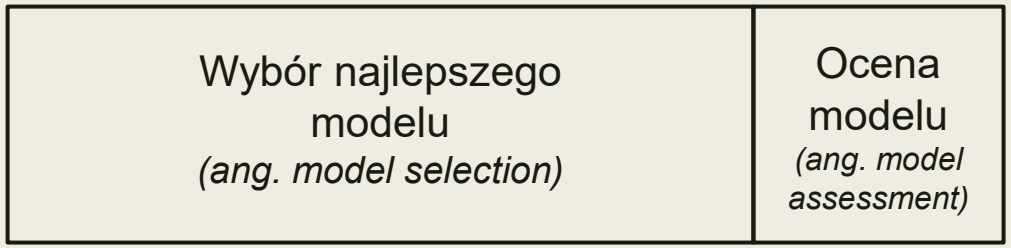


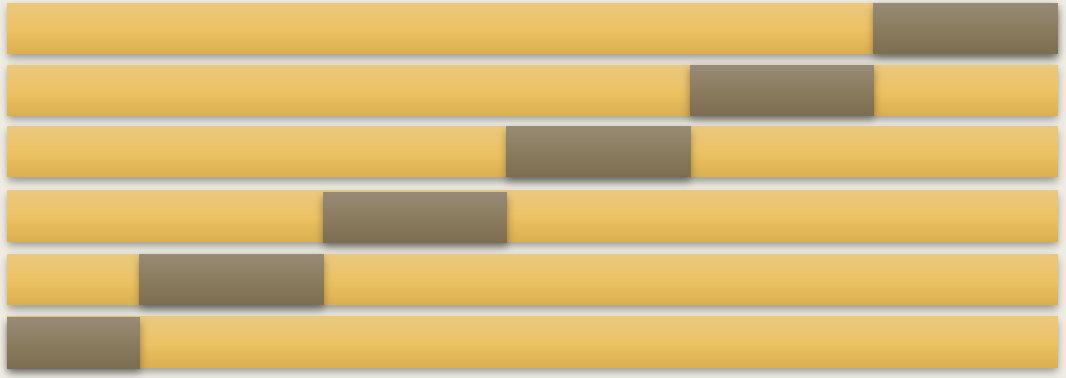
Wybór najlepszego modelu
(ang. model selection)



Walidacja krzyżowa
(ang. *Cross-Validation – CV*)

- Losowy podział
- Wielokrotny podział
- Oszacowanie błędu na zbiorze testującym





Potok wydobywania wiedzy z danych

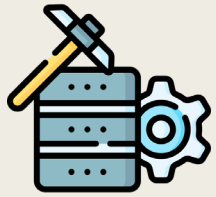
Gromadzenie danych

Potok wydobywania wiedzy z danych

Gromadzenie danych → Sformułowanie problemu →

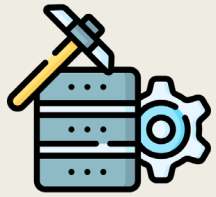
Potok wydobywania wiedzy z danych

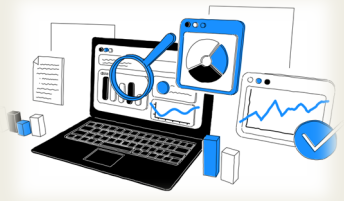
Gromadzenie danych → Sformułowanie problemu → Przygotowanie danych



Potok wydobywania wiedzy z danych

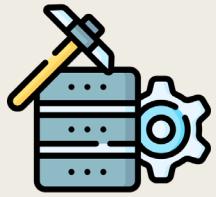
Gromadzenie danych → Sformułowanie problemu → Przygotowanie danych → Eksploracja danych →

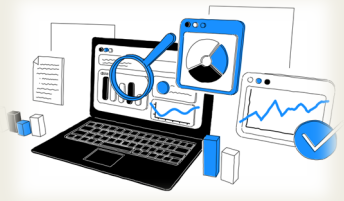


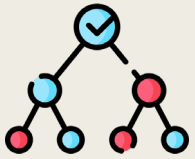


Potok wydobywania wiedzy z danych

Gromadzenie danych → Sformułowanie problemu → Przygotowanie danych → Eksploracja danych → Uczenie modeli →



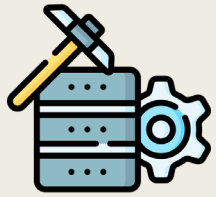
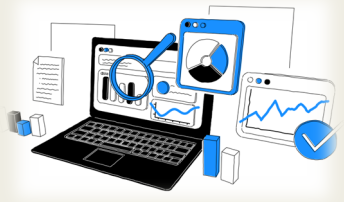
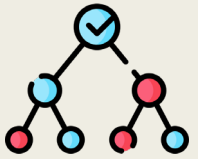




Potok wydobywania wiedzy z danych

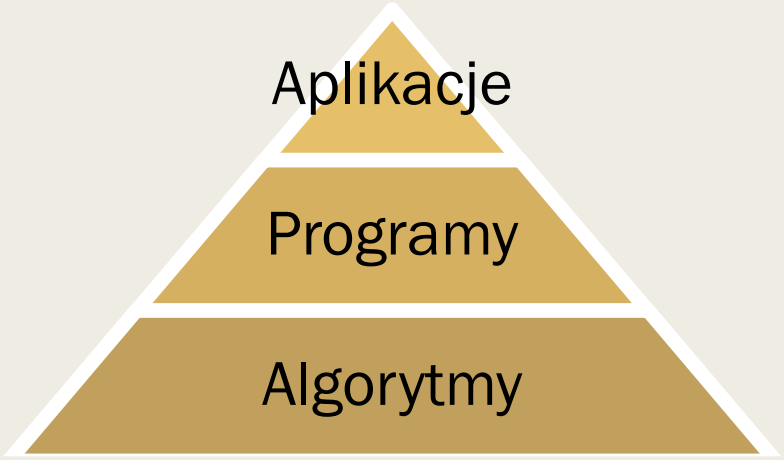
Gromadzenie danych → Sformułowanie problemu → Przygotowanie danych → Eksploracja danych → Uczenie modeli → Opracowanie narracji



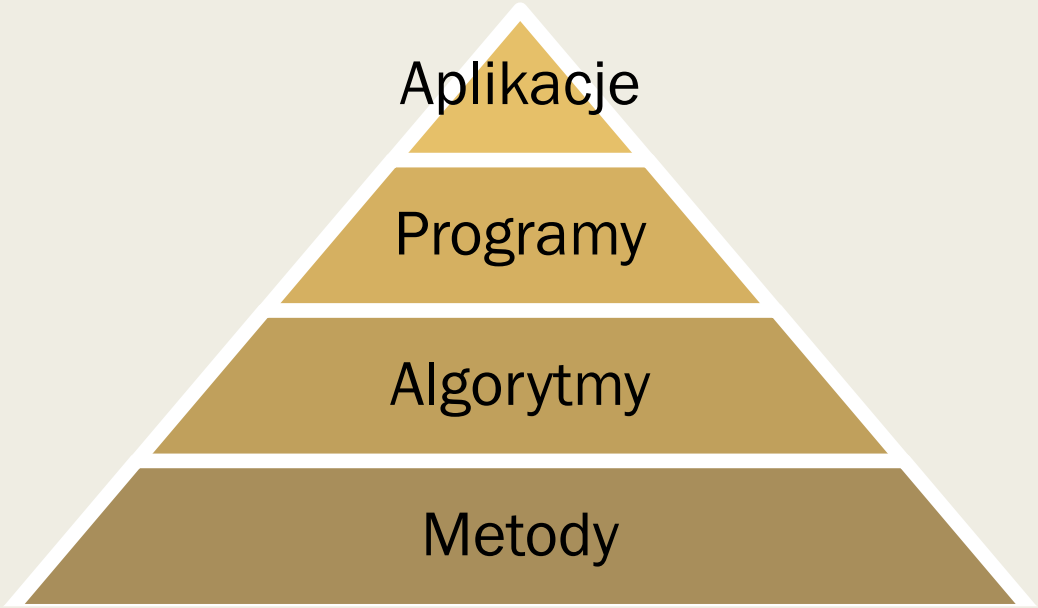





Skąd się biorą aplikacje ze sztuczną inteligencją na pokładzie?

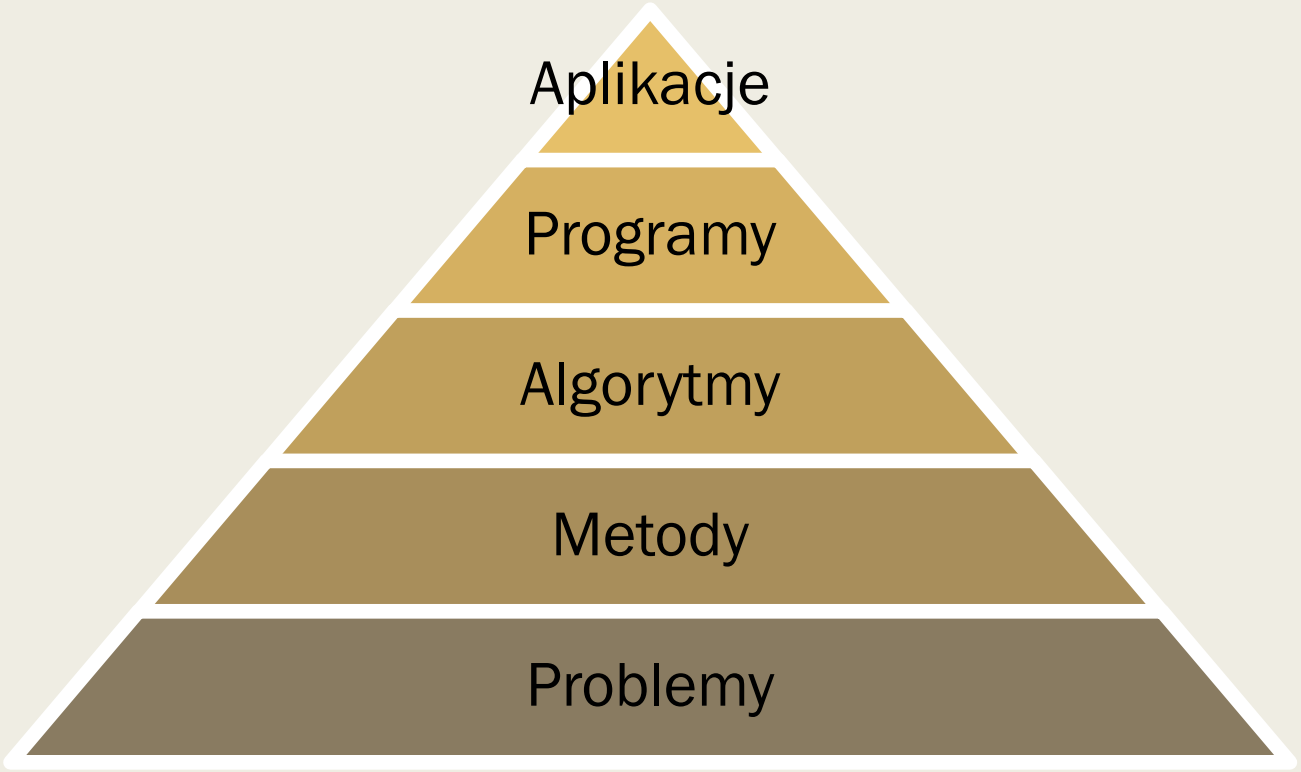
Skąd się biorą aplikacje ze sztuczną inteligencją na pokładzie?



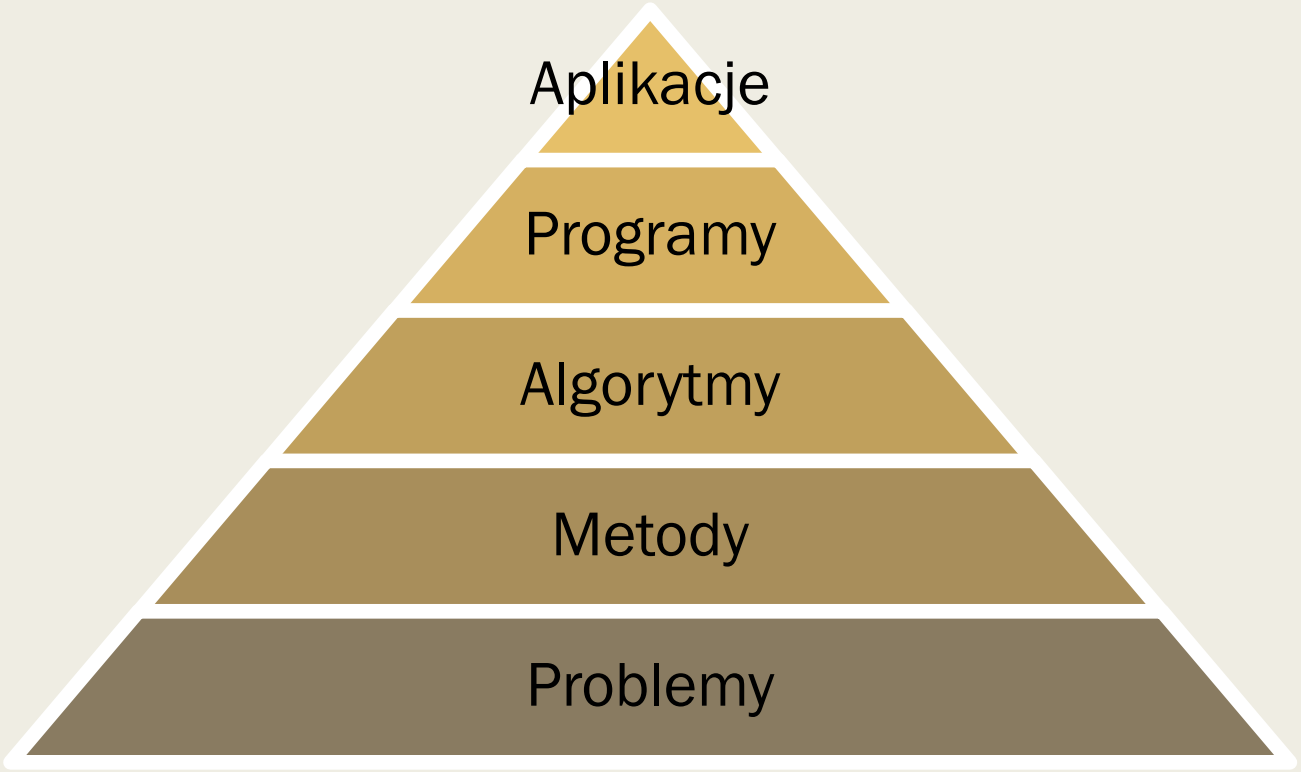
Skąd się biorą aplikacje ze sztuczną inteligencją na pokładzie?



Skąd się biorą aplikacje ze sztuczną inteligencją na pokładzie?



Skąd się biorą aplikacje ze sztuczną inteligencją na pokładzie?



analiza, algebra, statystyka, optymalizacja

koniec