



Uczenie maszynowe

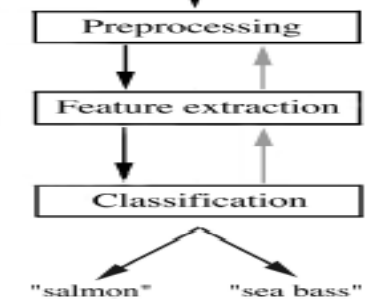
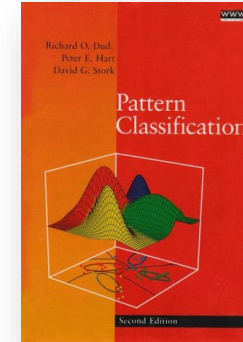
Danologia

Przykłady zadań uczenia maszynowego

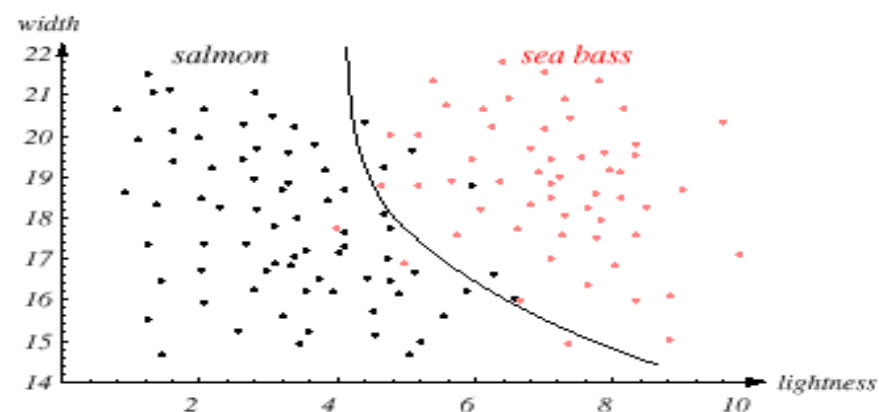
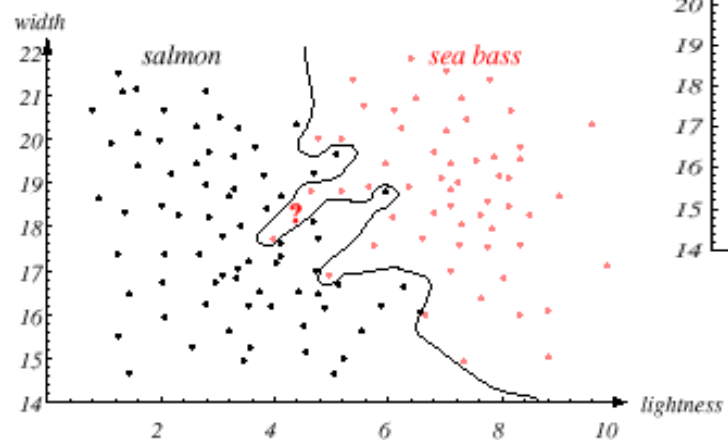
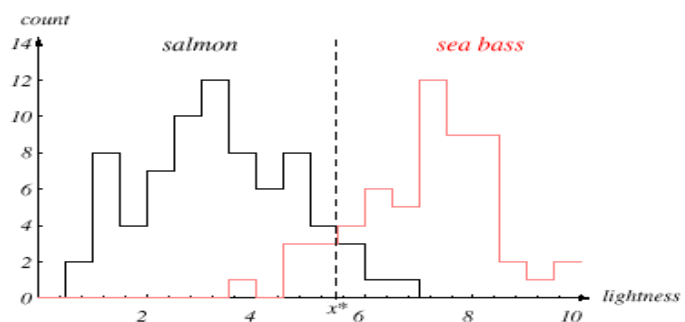
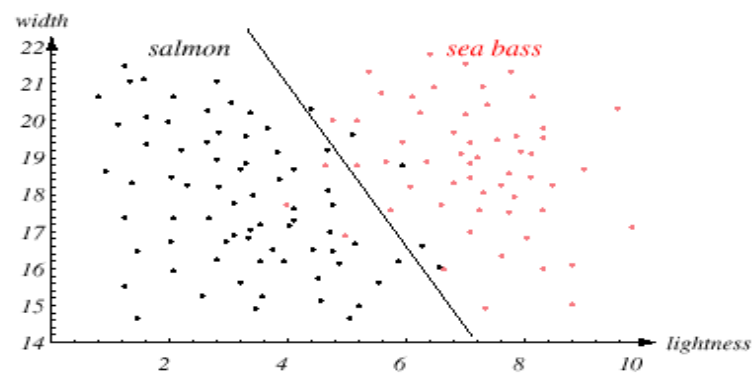
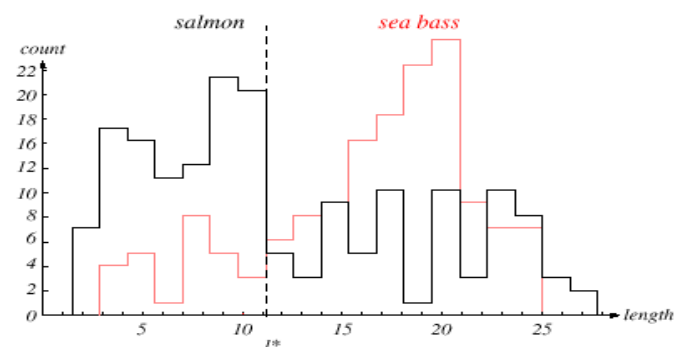
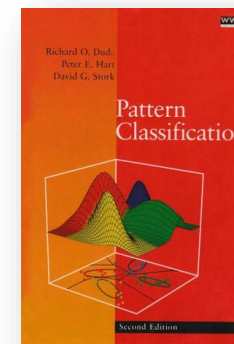
Klasyfikacja

Cechy:

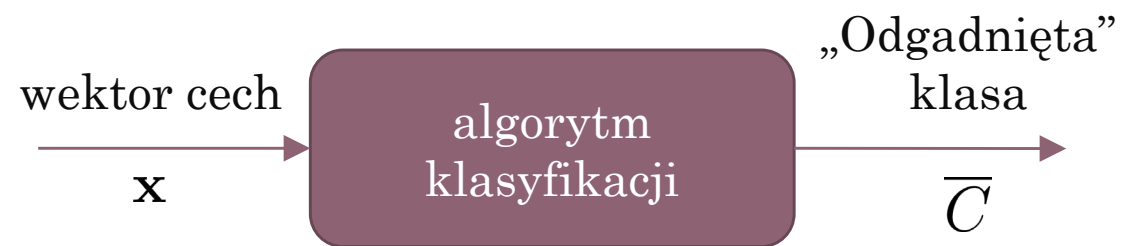
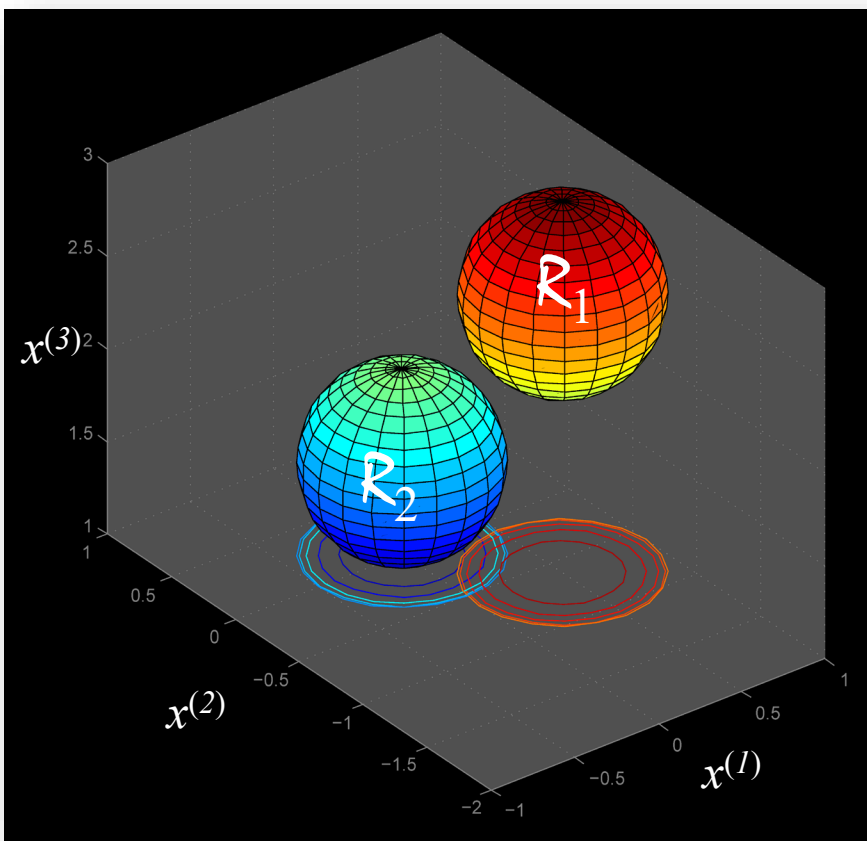
- ✓ długość
- ✓ jasność
- ✓ szerokość
- ✓ liczba i kształt płetw
- ✓ położenie otworu gębowego



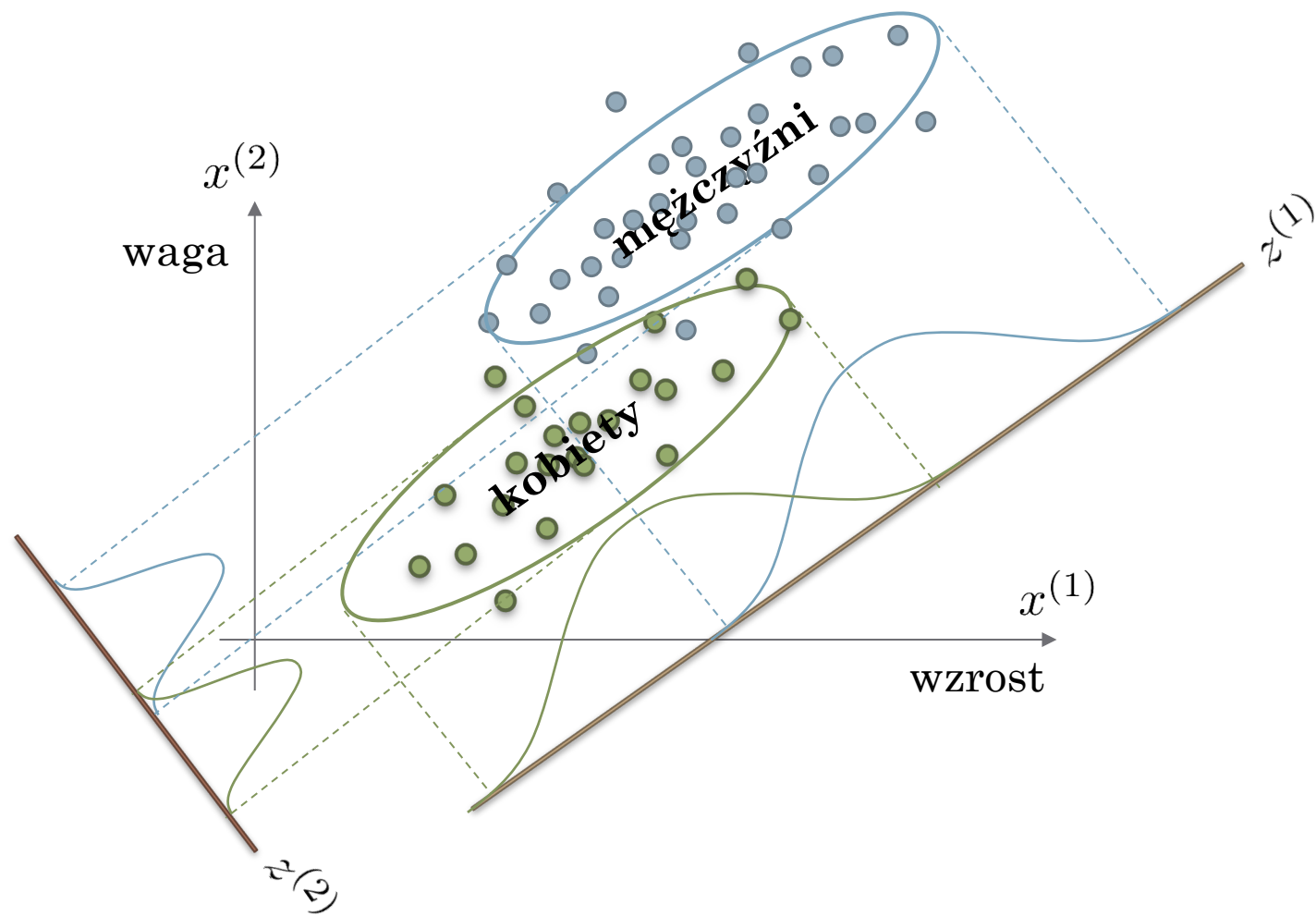
Klasyfikacja



Klasyfikacja



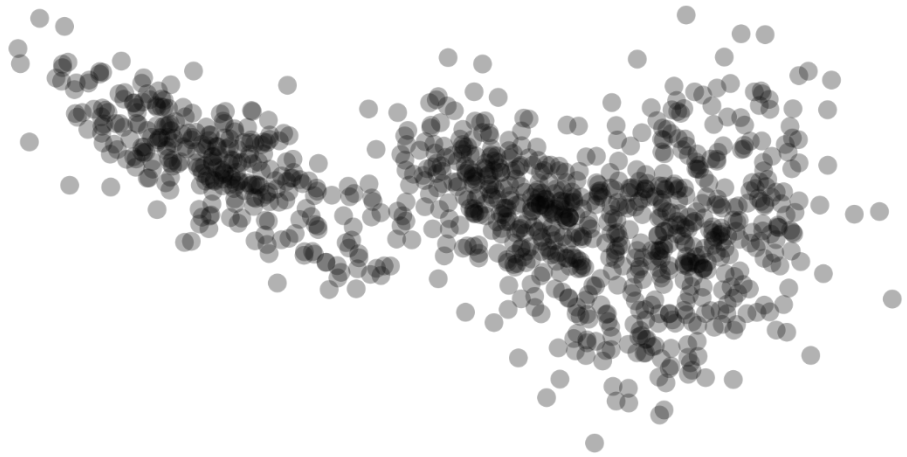
Transformacje



Przykłady algorytmów

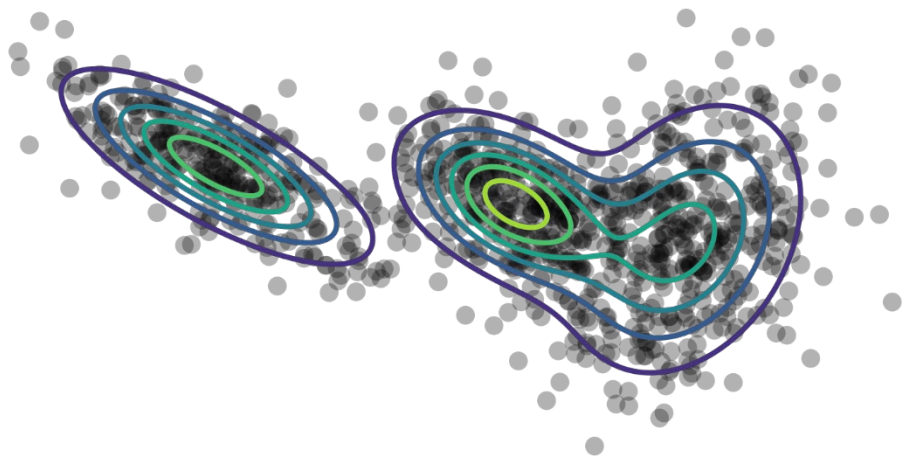
- k -najbliższych sąsiadów
- klasyfikator liniowy
- regresja logistyczna
- maszyny wektorów wspierających (ang. *Support Vector Machines*)
- drzewa decyzyjne (lasy)
- sieci neuronowe
- ...

Estymacja rozkładu



Mieszanina rozkładów Gaussa

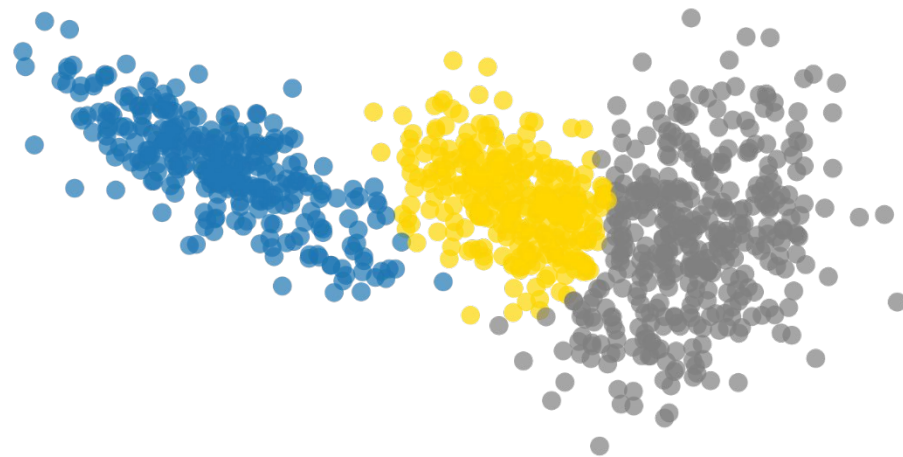
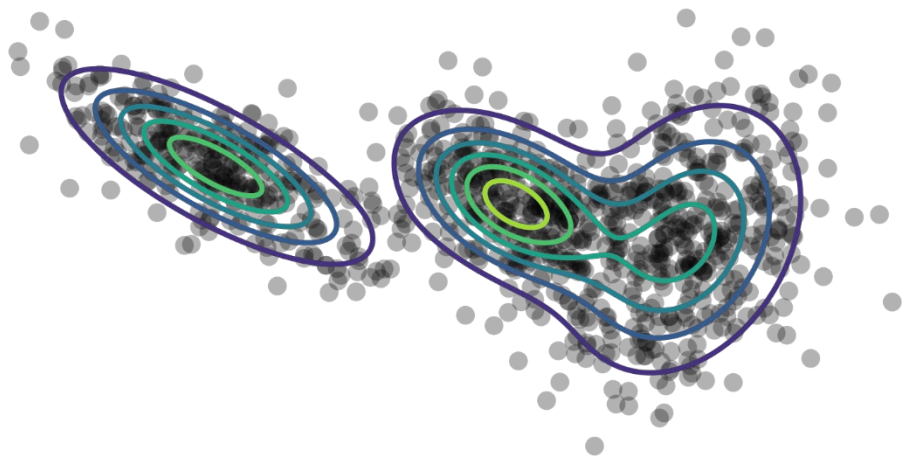
$$p(\mathbf{x}|\theta) = \sum_{m=1}^M \pi_m \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad 0 \leq \pi_m \leq 1, \quad \sum_{m=1}^M \pi_m = 1$$



Mieszanina rozkładów Gaussa

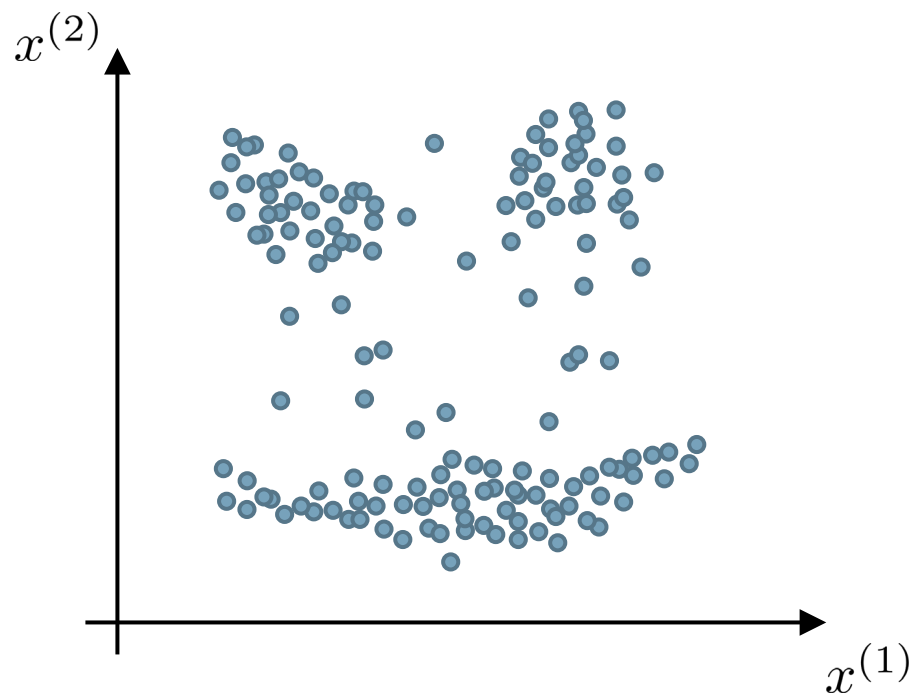
$$p(\mathbf{x}|\theta) = \sum_{m=1}^M \pi_m \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$$

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



Grupowanie

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



Przykłady algorytmów

- k -średnich (ang. *k-means*) / centroidów
- mieszanina rozkładów Gaussa
 - maksymalizacja oczekiwania (ang. *Expectation–Maximization*)

- DBSCAN

- widmowa analiza skupień

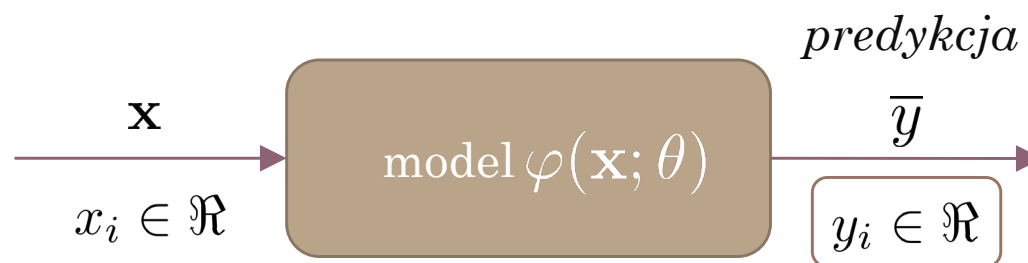
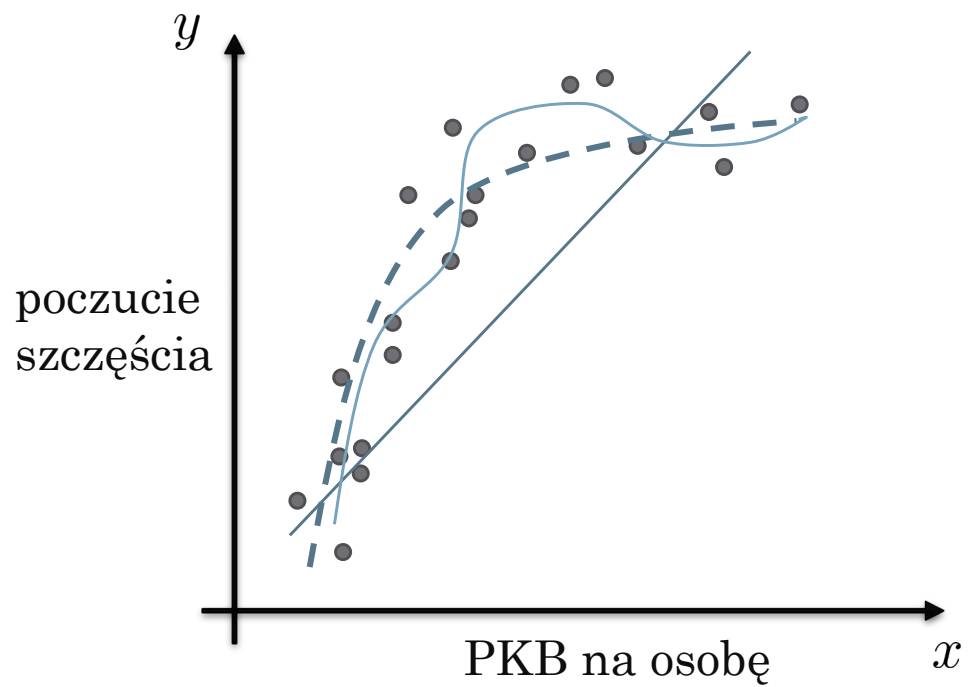
- grupowanie twarde i miękkie
 - rozmyte c-średnich (ang. fuzzy c-means)

- metody aglomeracyjne

```
sklearn.mixture.GaussianMixture  
                .BayesianGaussianMixture
```

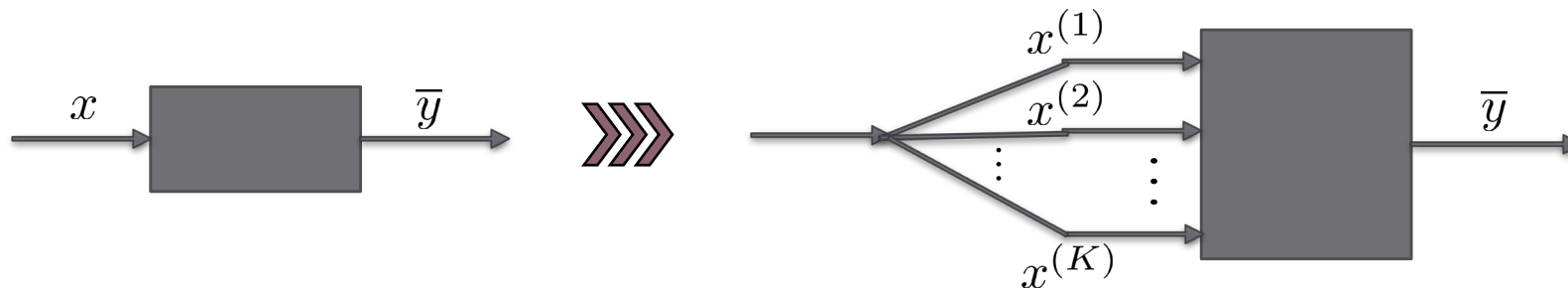
```
sklearn.cluster.Kmeans  
                .DBSCAN  
                .AgglomerativeClustering  
                .SpectralClustering  
scipy.cluster.hierarchy.linkage  
                      .dendrogram
```

Regresja



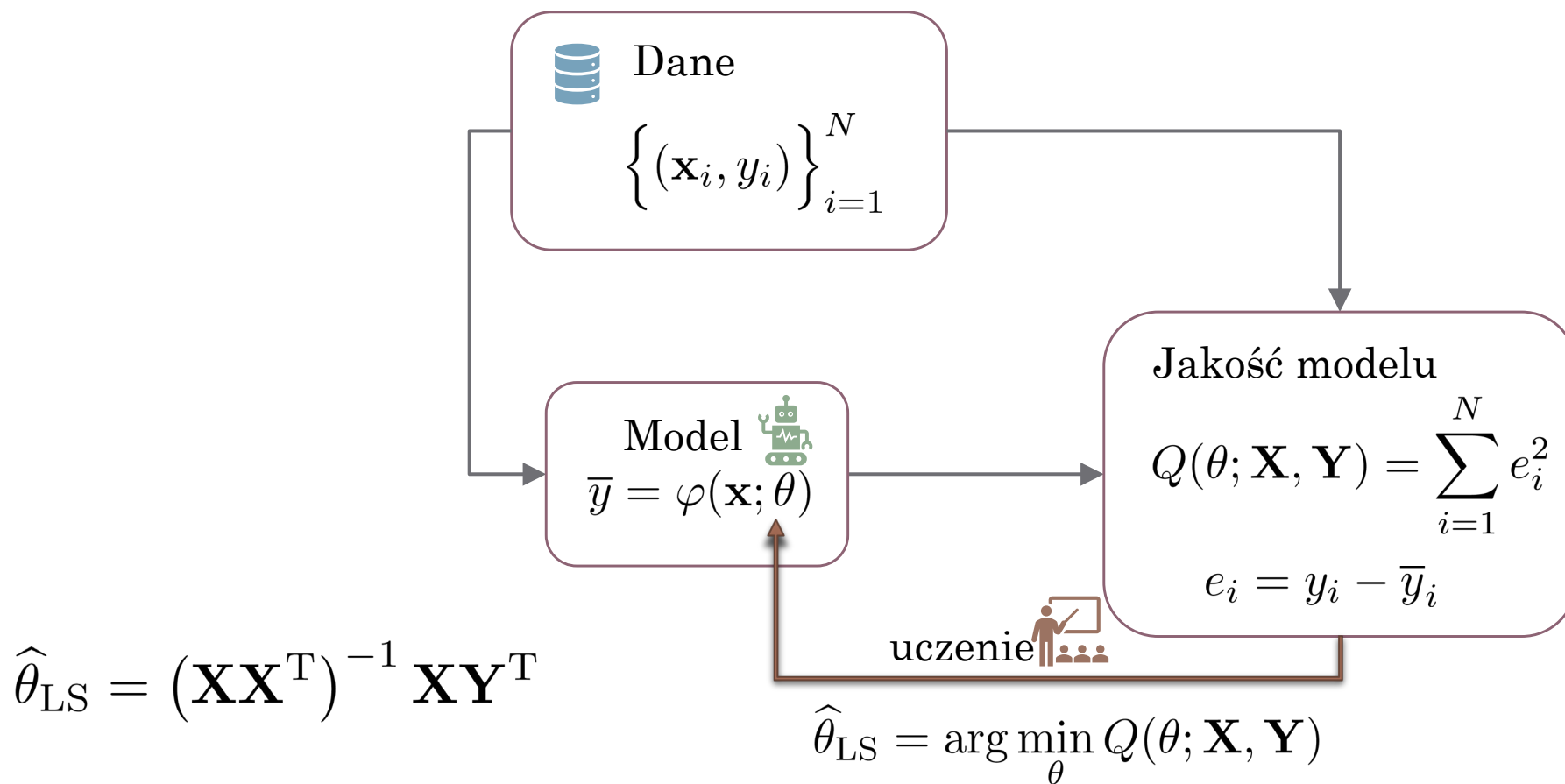
→ *Model odniesienia / naiwny*

Uogólnione modele liniowe



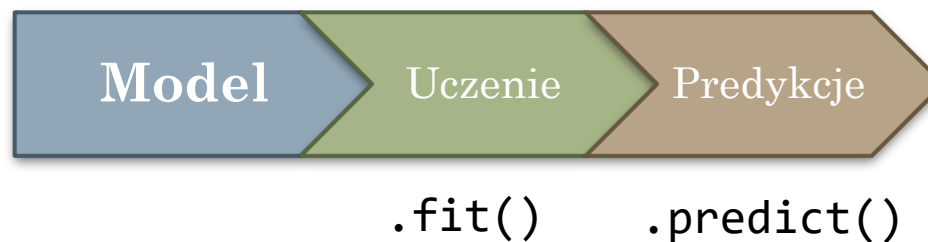
Regresja

Dopasowanie
modelu do
danych = **uczenie**

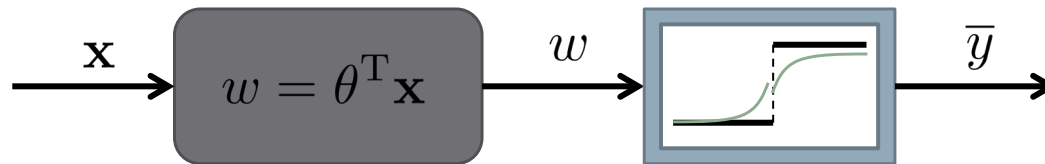


Przykłady algorytmów

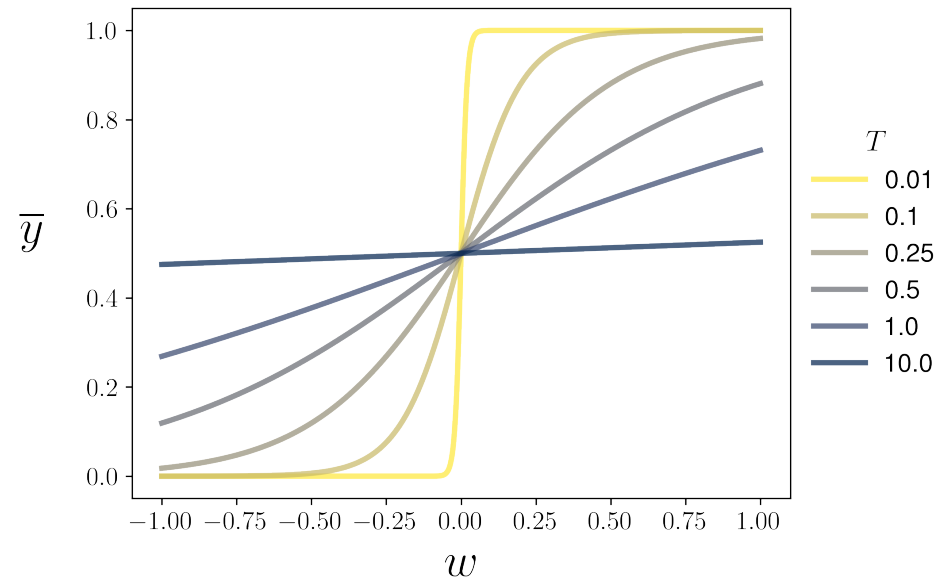
- regresja liniowa
- uogólnione modele liniowe (ang. *Generalized linear models*)
- drzewa decyzyjne (lasy)
- regresja bayesowska
- maszyny wektorów wspierających (ang. *Support Vector Machines*)
- sieci neuronowe
- ...



Model regresji jako klasyfikator

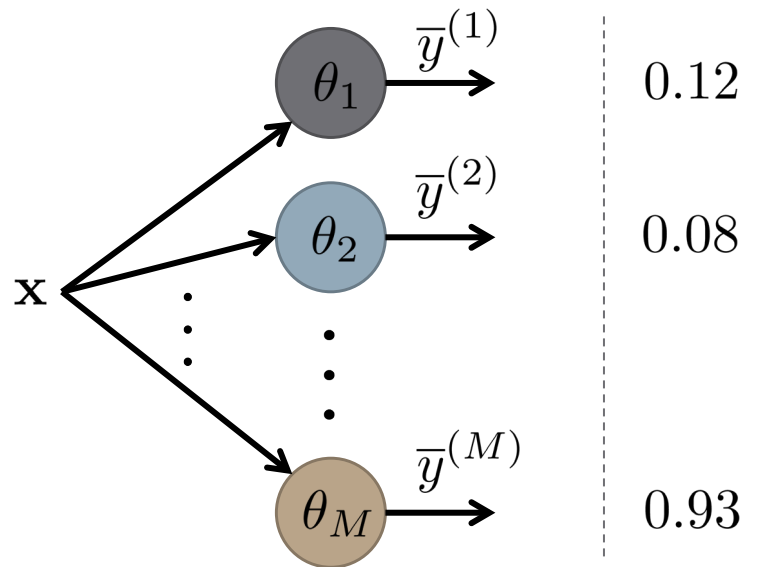


$$\bar{y} = \frac{1}{1 + e^{-w}}$$



Model regresji jako klasyfikator

Wiele klas



$$\bar{y}^{(m)} = \frac{1}{1 + e^{-\theta_m^T \mathbf{x}}} \quad (m = 1, 2, \dots, M)$$

vs

Softmax

$$\bar{y}^{(m)} = \frac{e^{\theta_m^T \mathbf{x}}}{\sum_{l=1}^M e^{\theta_l^T \mathbf{x}}} \quad (m = 1, 2, \dots, M)$$

Ujęcie probabilistyczne klasyfikacji

wektor cech

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

etykiety klas

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

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

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

$p(\mathbf{x})$ – rozkład cech

$p(C)$ – rozkład klas

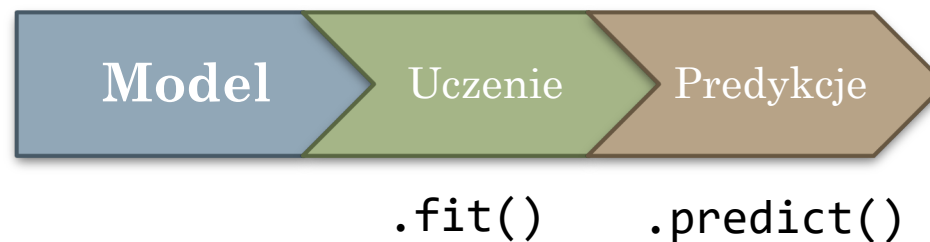
$p(\mathbf{x}|C)$ – rozkład cech w klasach

→ *Modele generatywne*

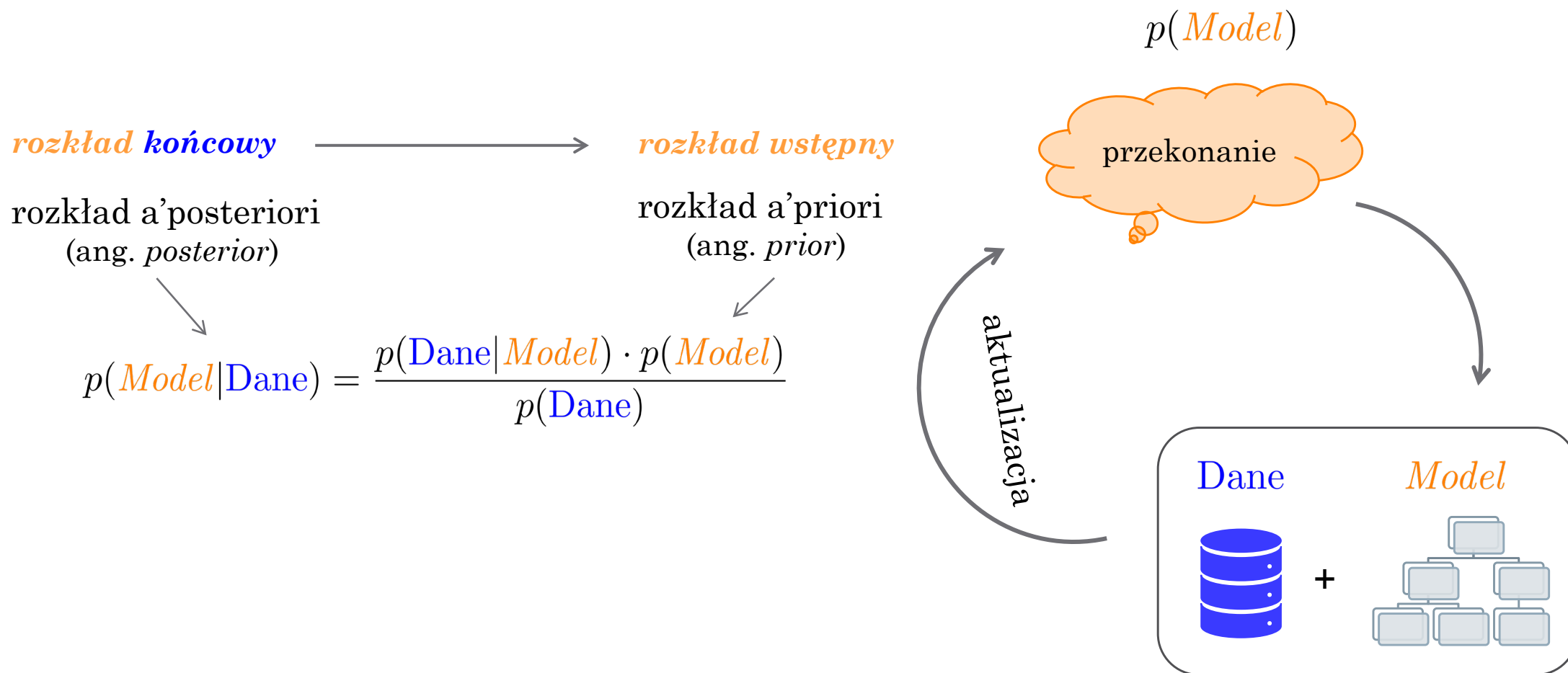
$$p(C|\mathbf{x}) = \frac{p(\mathbf{x}|C) \cdot p(C)}{p(\mathbf{x})}$$

Przykłady algorytmów

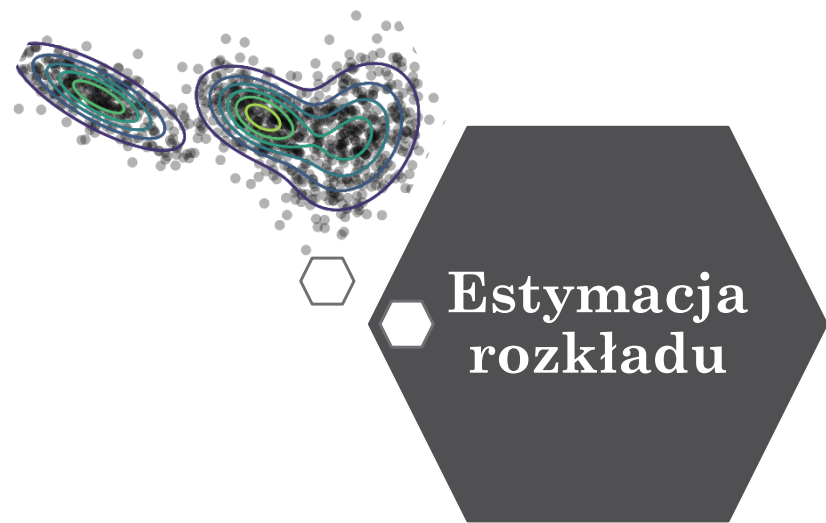
- optymalny klasyfikator Bayesa
- naiwny klasyfikator Bayesa (ang. *Naive Bayes Classifier*)
- ukryte modele Markowa (ang. *Hidden Markov Models*)
- ...



Metody bayesowskie



Podstawowe zadania uczenia maszynowego



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

uczenie nadzorowane
(ang. *supervised learning*)

etykieta
„prawdziwa
odpowiedź”

regresja
wyjście ciągłe

klasyfikacja
wyjście
kategorialne

uczenie
półnadzorowane
(ang. *semisupervised
learning*)

uczenie ze
wzmocnieniem
(ang. *reinforcement
learning*)

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

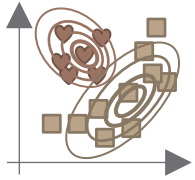
uczenie nienadzorowane
(ang. *unsupervised learning*)

brak etykiet

grupowanie

Uczenie maszynowe

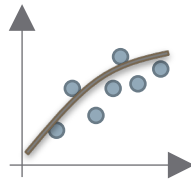
Estymacja rozkładu



- rozkład dedykowany
- estymator jądrowy
- mieszanina rozkładów Gaussa

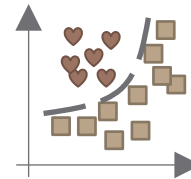
Uczenie nadzorowane

Regresja



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

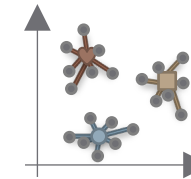
Klasyfikacja



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

Uczenie nienadzorowane

Grupowanie



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

Sposoby korzystania z bazy danych



Parametryczne

Ustalona liczba parametrów



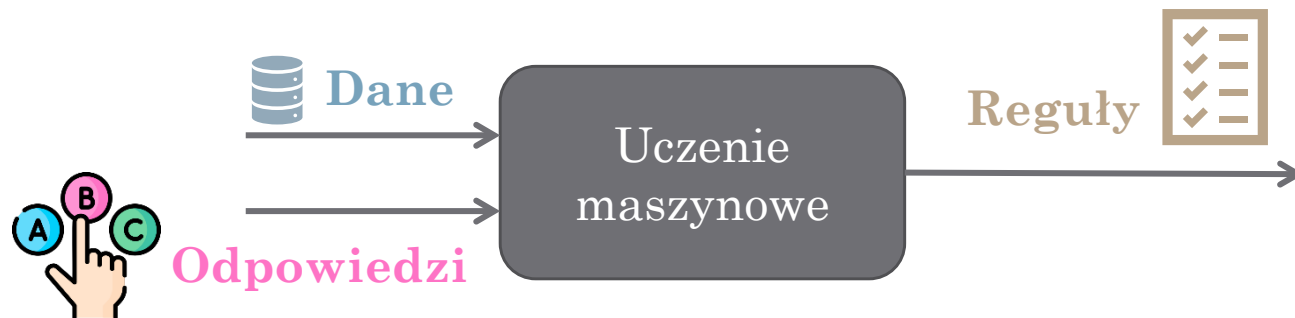
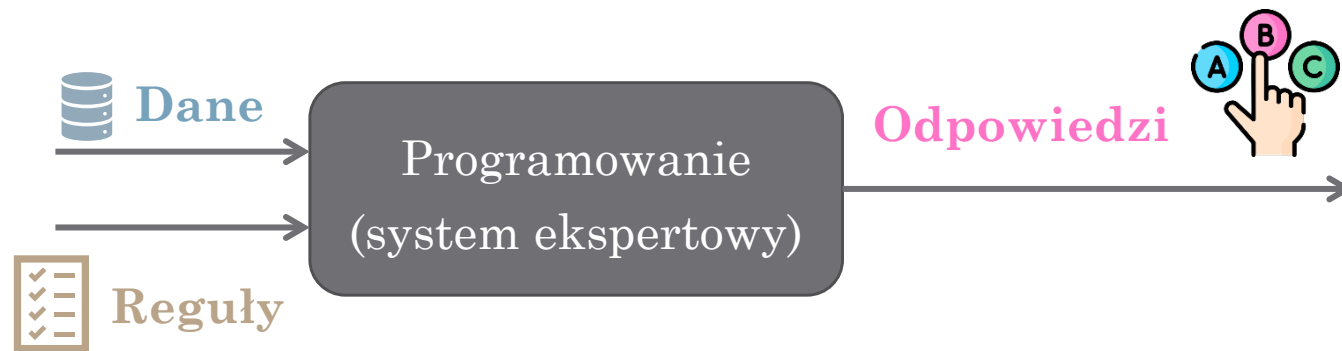
- ☐ klasyfikator liniowy
- ☐ regresja liniowa i logistyczna
- ☐ mieszanina rozkładów Gaussa
- ☐ liniowy SVM
- ☐ naiwny klasyfikator Bayesa
- ☐ ...

Nieparametryczne

Postać modelu i liczba parametrów zależą od danych



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



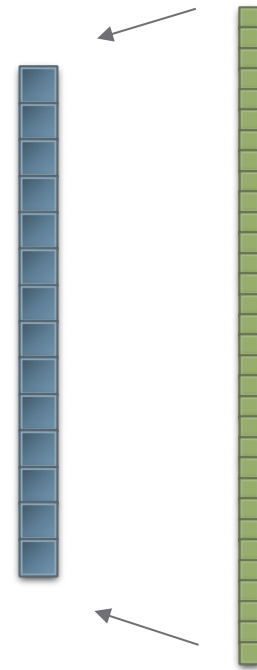
Uczenie głębokie

x



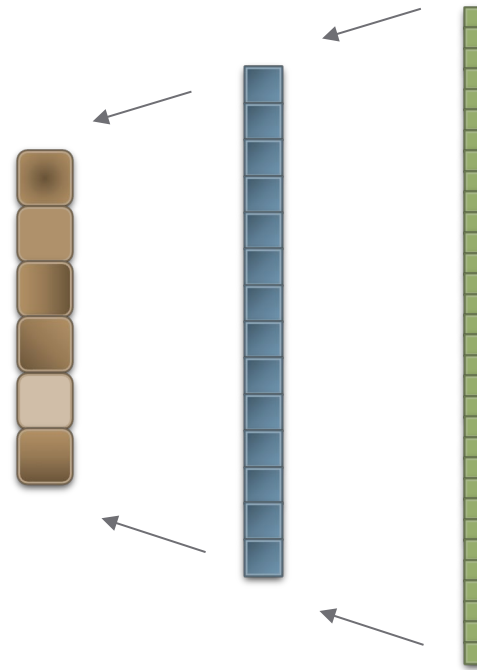
Uczenie głębokie

$$\varphi_0(\mathbf{W}_0 \cdot \mathbf{x})$$



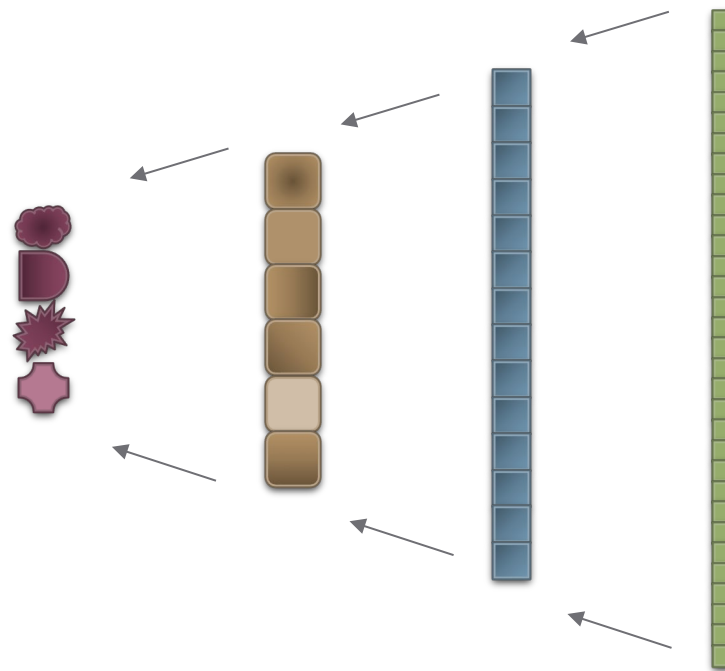
Uczenie głębokie

$$\varphi_1(\mathbf{W}_1 \cdot \varphi_0(\mathbf{W}_0 \cdot \mathbf{x}))$$



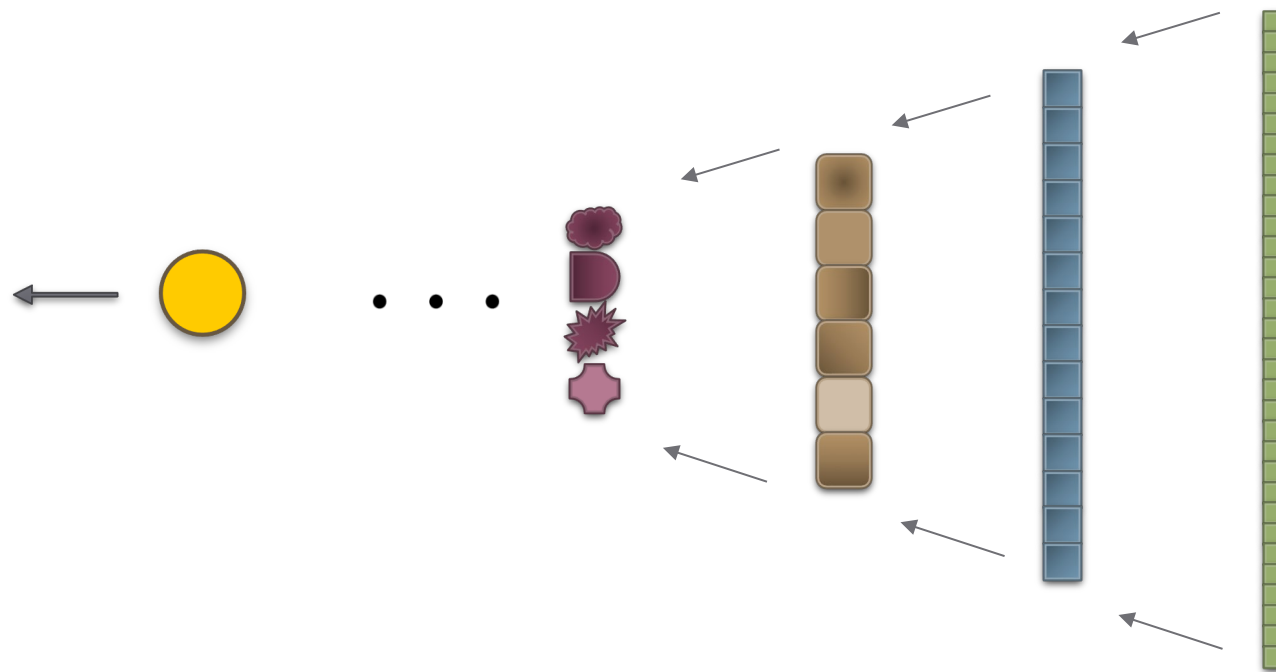
Uczenie głębokie

$$\varphi_2\left(\mathbf{W}_2 \cdot \varphi_1\left(\mathbf{W}_1 \cdot \varphi_0\left(\mathbf{W}_0 \cdot \mathbf{x}\right)\right)\right)$$



Uczenie głębokie

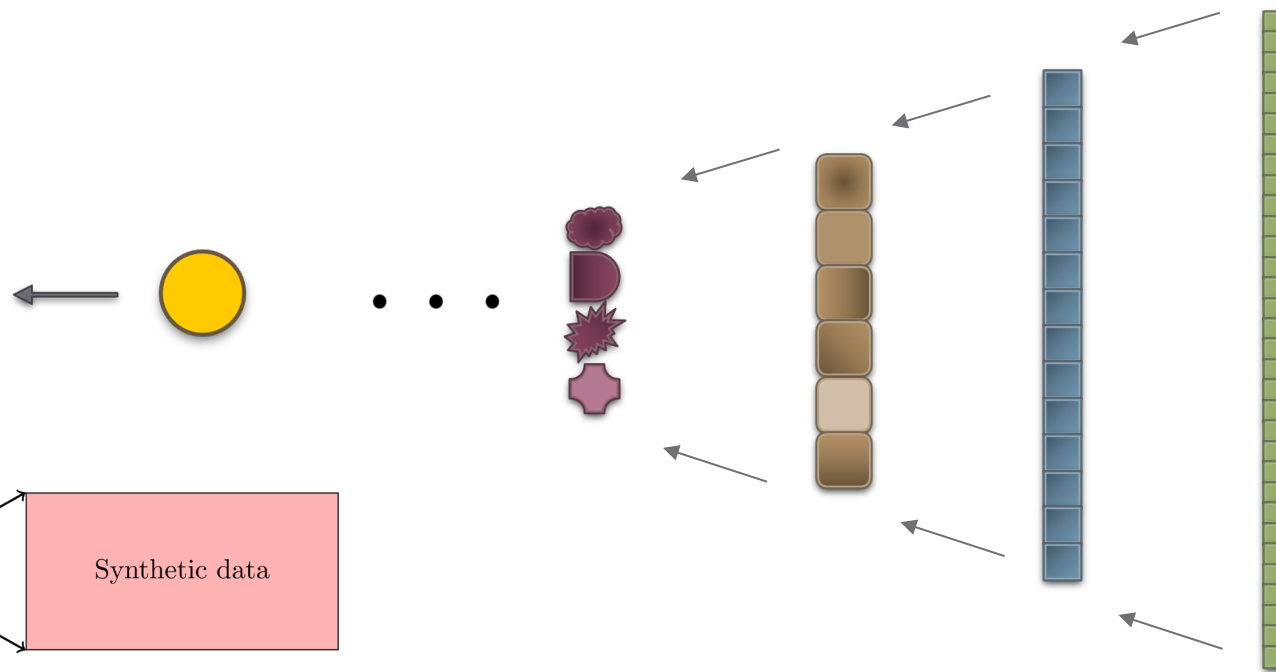
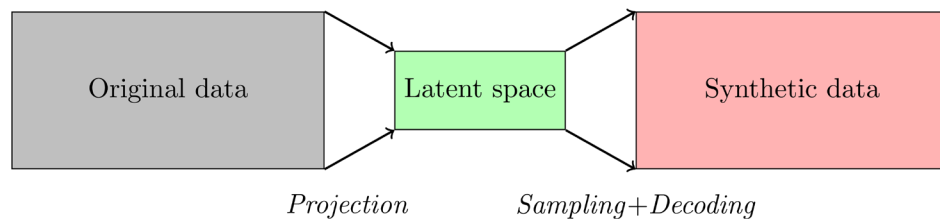
$$y = \varphi_K \left(\mathbf{W}_K \cdot \dots \cdot \varphi_2 \left(\mathbf{W}_2 \cdot \varphi_1 \left(\mathbf{W}_1 \cdot \varphi_0 (\mathbf{W}_0 \cdot \mathbf{x}) \right) \right) \dots \right)$$



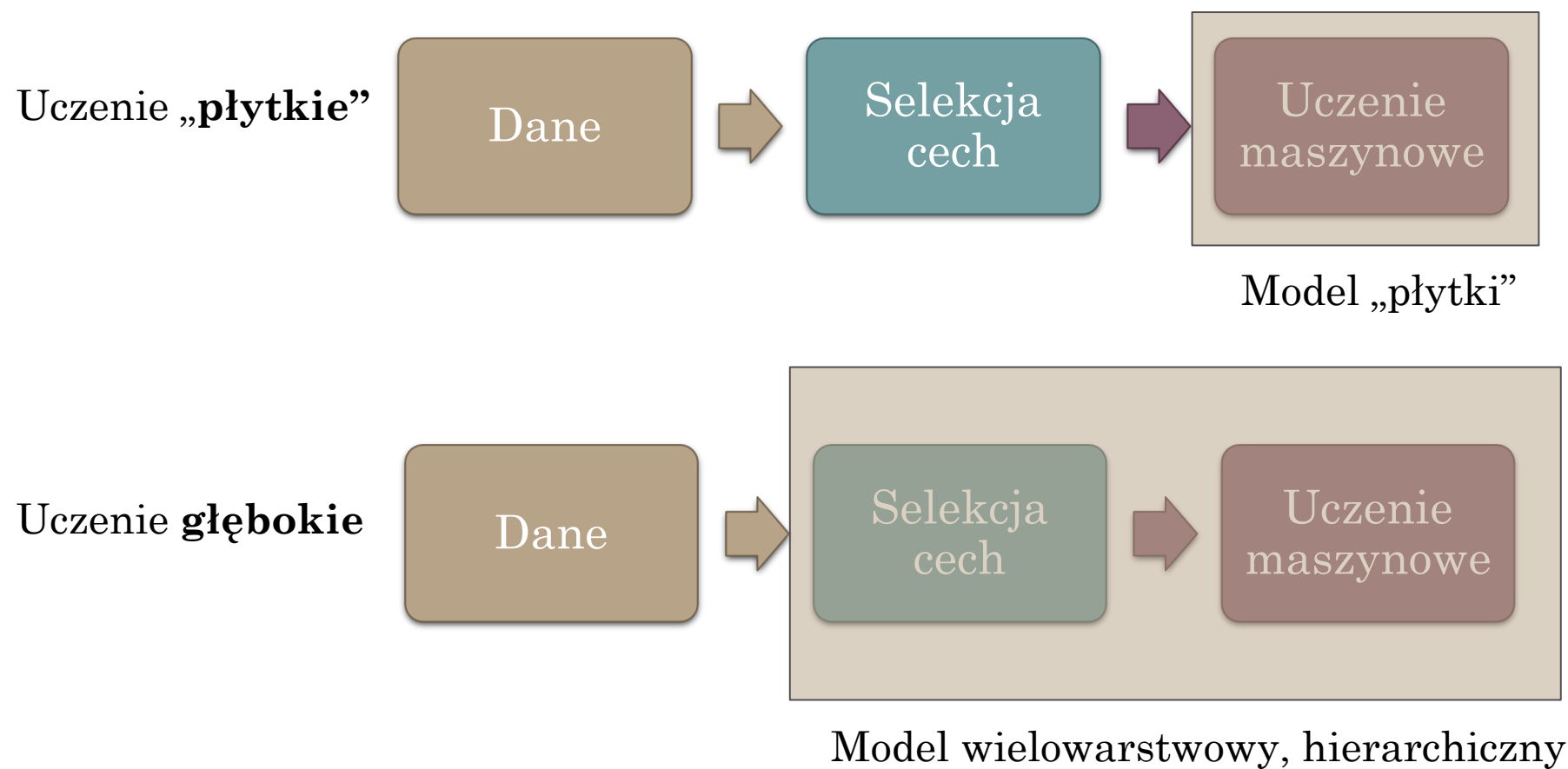
Uczenie głębokie

$$y = \varphi_K \left(\mathbf{W}_K \cdot \dots \cdot \varphi_2 \left(\mathbf{W}_2 \cdot \varphi_1 \left(\mathbf{W}_1 \cdot \varphi_0 (\mathbf{W}_0 \cdot \mathbf{x}) \right) \right) \dots \right)$$

→ *Modele generatywne*



Uczenie głębokie

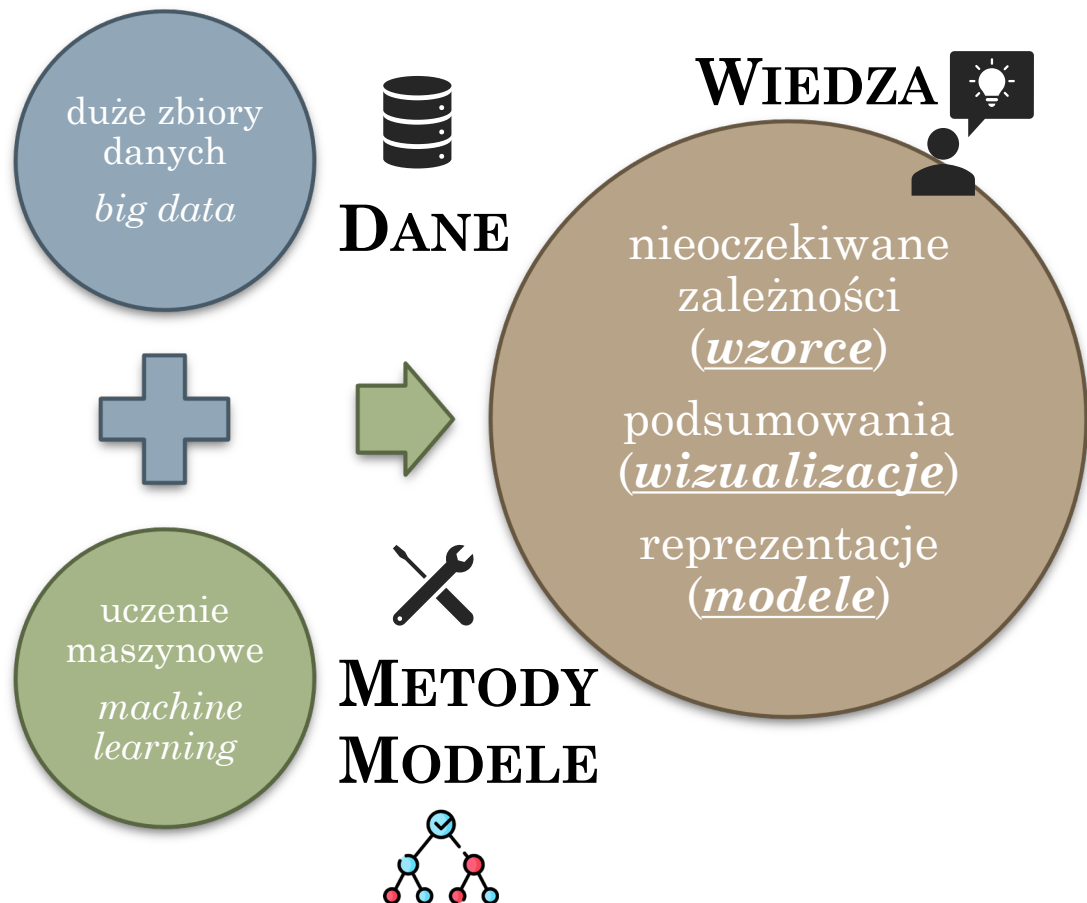


Uczenie głębokie

- 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



Dane, wiedza, modele i predykcje



- Dane gromadzone z różnych przyczyn
- Wydobywanie wiedzy z danych
- Rola modeli

Czym jest uczenie maszynowe

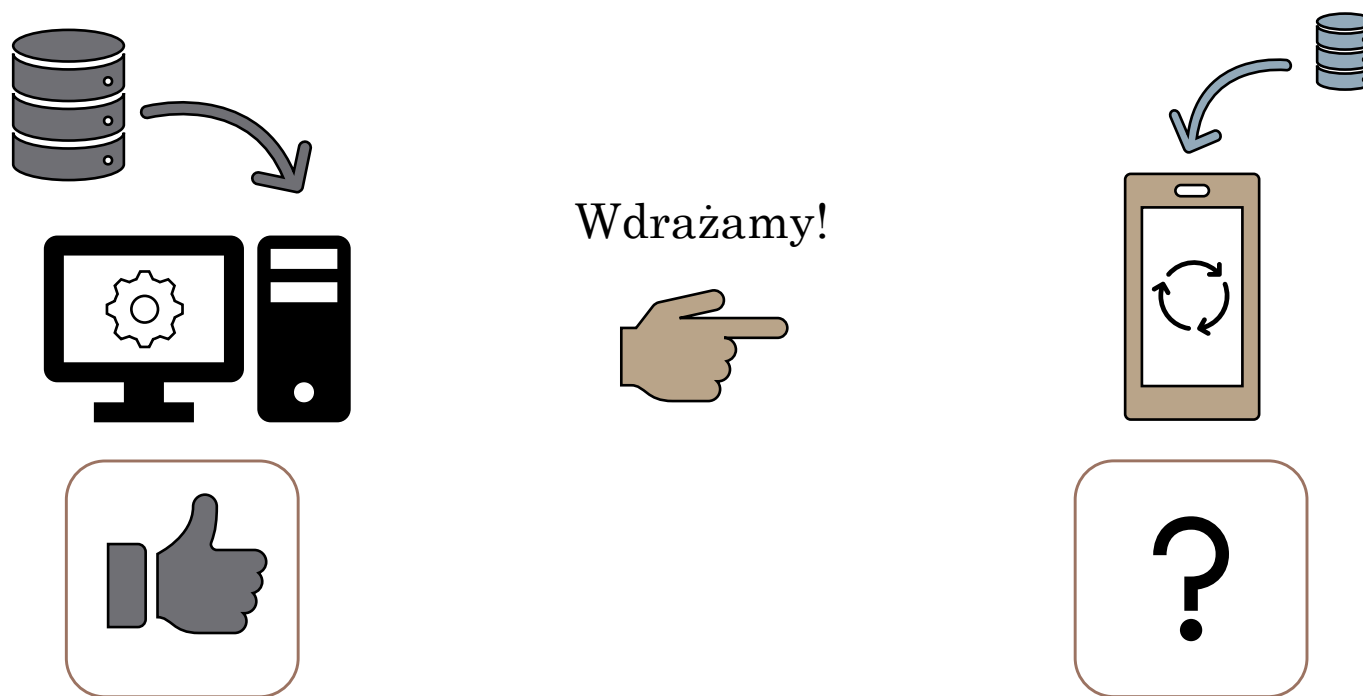
- ang. *machine learning* *learning* ≠ uczenie

I've learned that you are a confident person

- modelowanie statystycznych regularności w obserwacjach / pomiarach
- modelowanie danych vs 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**

Dobry model

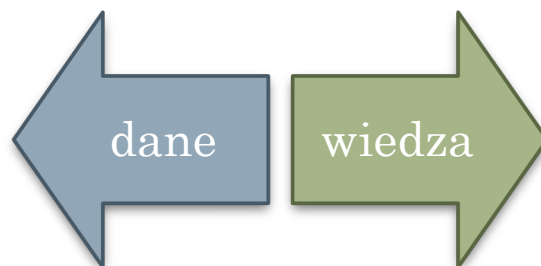
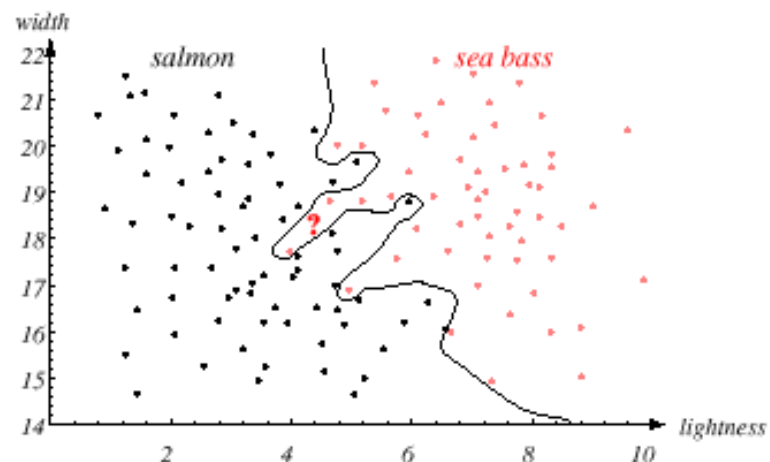
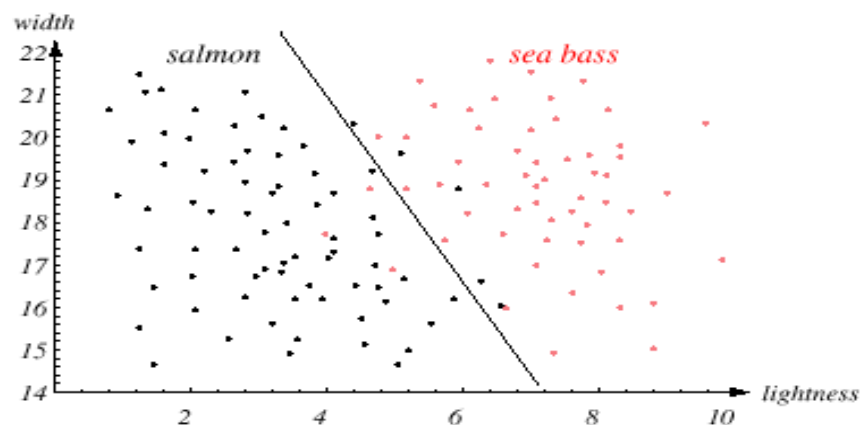
Jaki model jest dobry?



Jakość na danych treningowych $\implies$ Jakość na **nowych danych**

Złożoność modelu

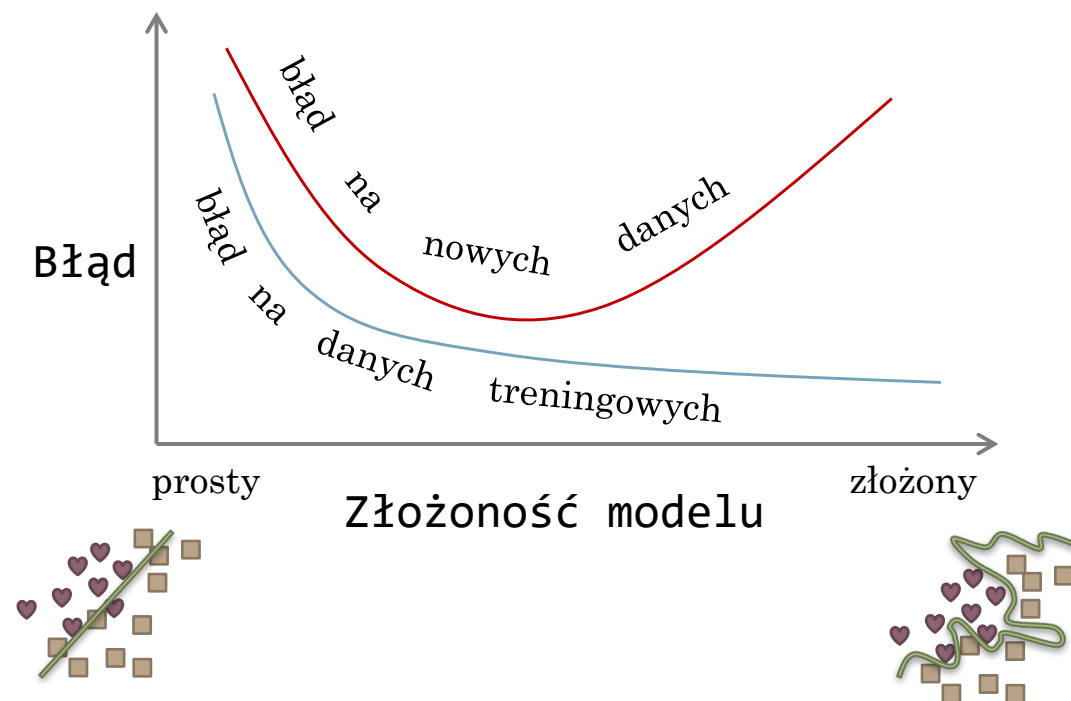
kompromis obciążenie – wariancja
(ang. *bias-variance dilemma*)



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

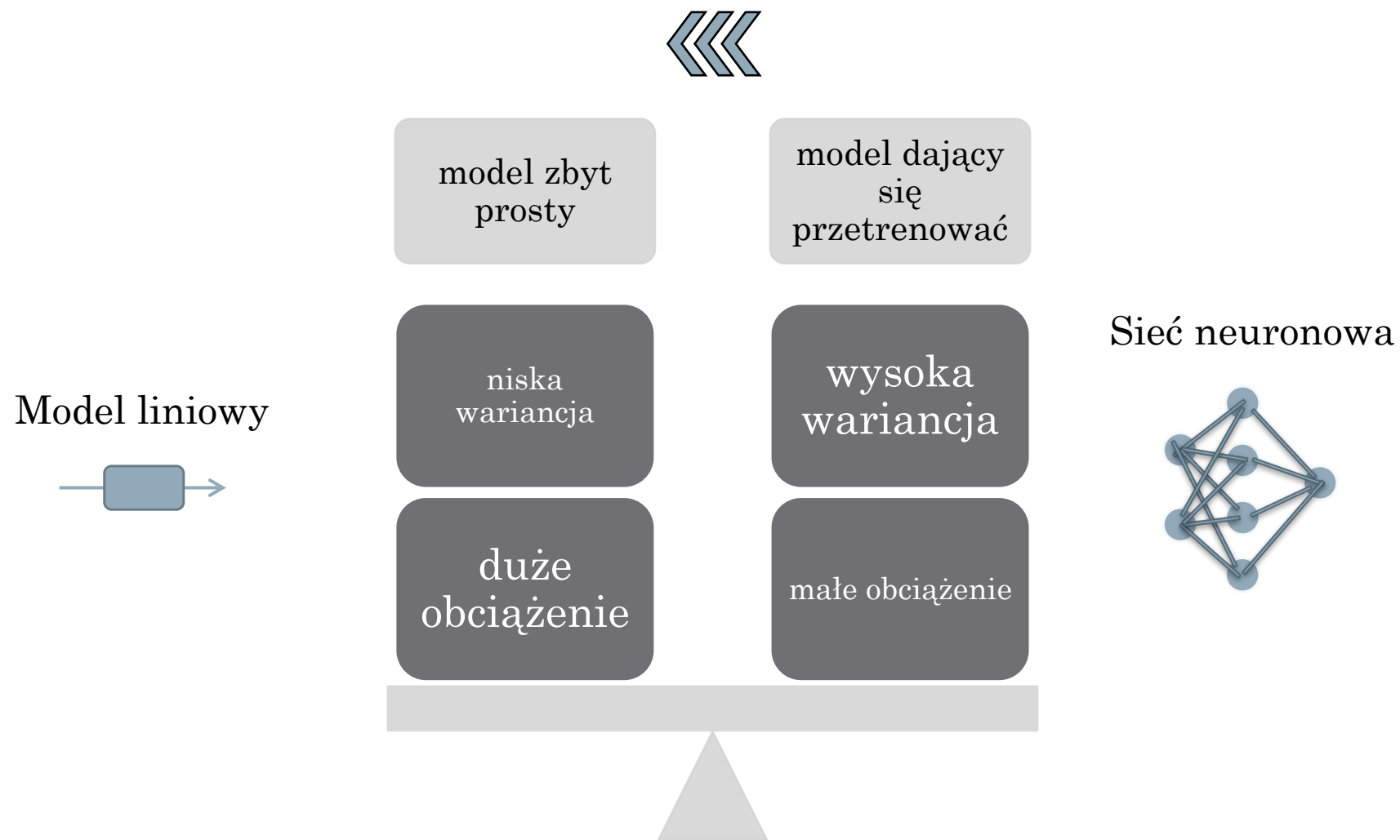
Złożoność modelu

- 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*)

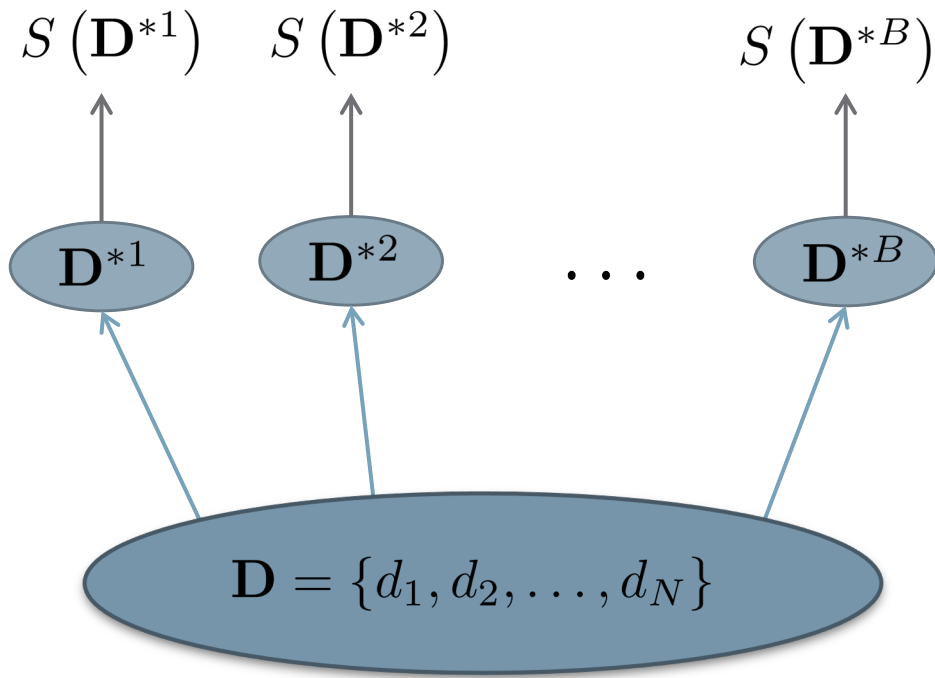


Regularyzacja

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



Metoda bootstrap



replikacje *bootstrapowe*

N – elementowe próby
bootstrapowe losowane
ze zwracaniem

Estymacja Monte-Carlo dokładności wyznaczenia

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

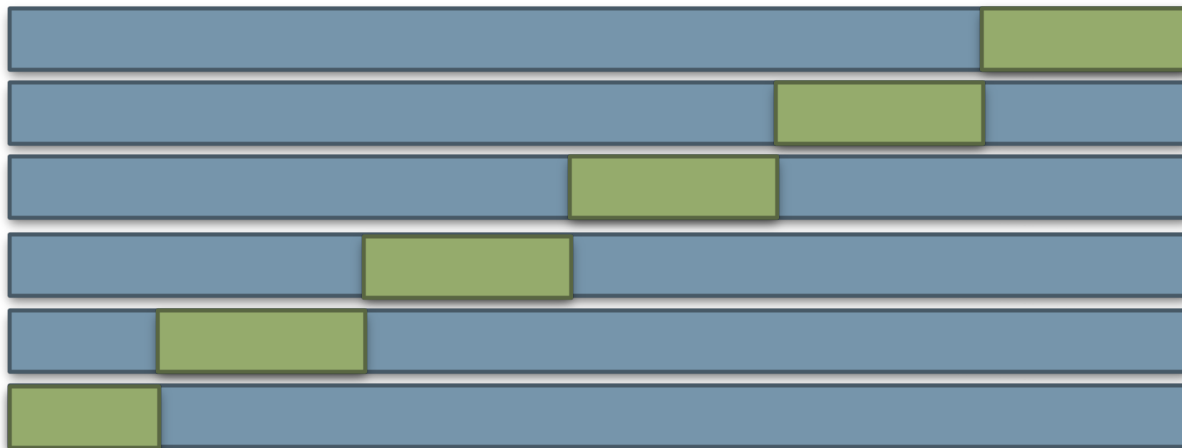
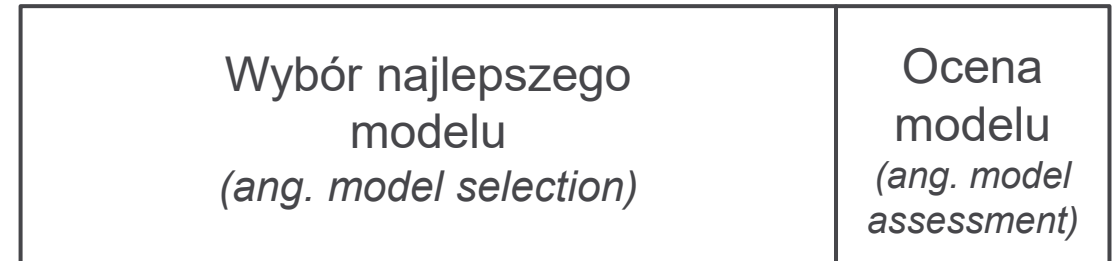
$$\widehat{\text{Var}}[S(D)] = \frac{1}{B-1} \sum_{b=1}^B \left(S(D^{*b}) - \bar{S}^* \right)^2$$



Walidacja krzyżowa

(ang. *Cross-Validation* – CV)

- Losowy podział zbioru uczącego
- Wielokrotny podział próby
- Oszacowanie błędu na zbiorze testującym



Potok wydobywania wiedzy z danych

Gromadzenie
danych

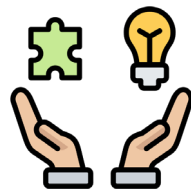


Potok wydobywania wiedzy z danych

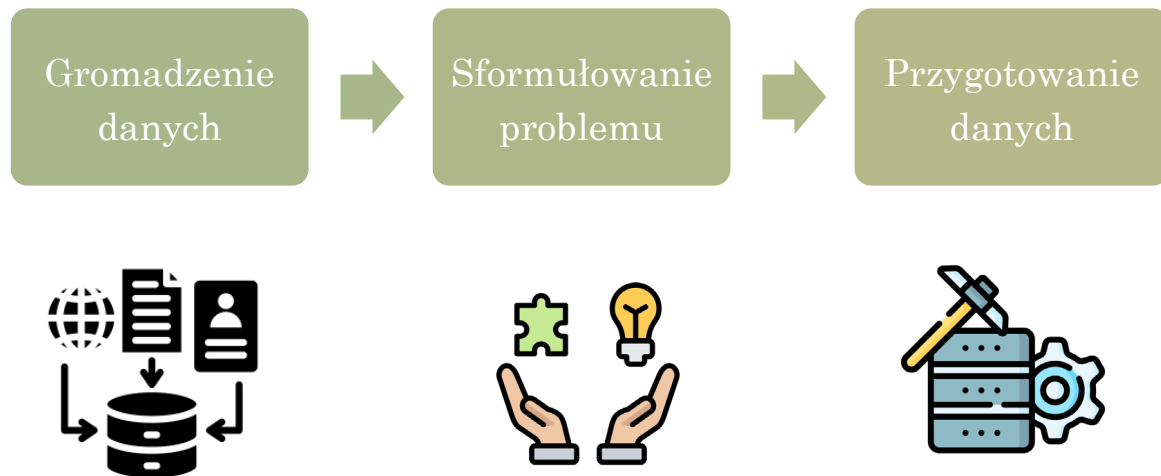
Gromadzenie
danych



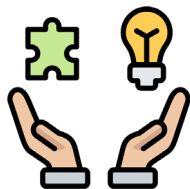
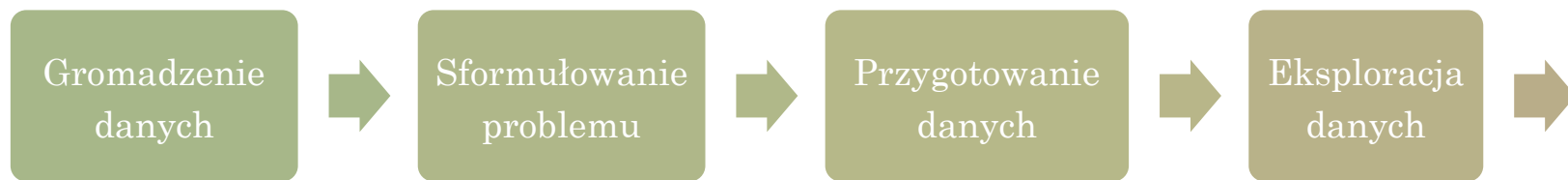
Sformułowanie
problemu



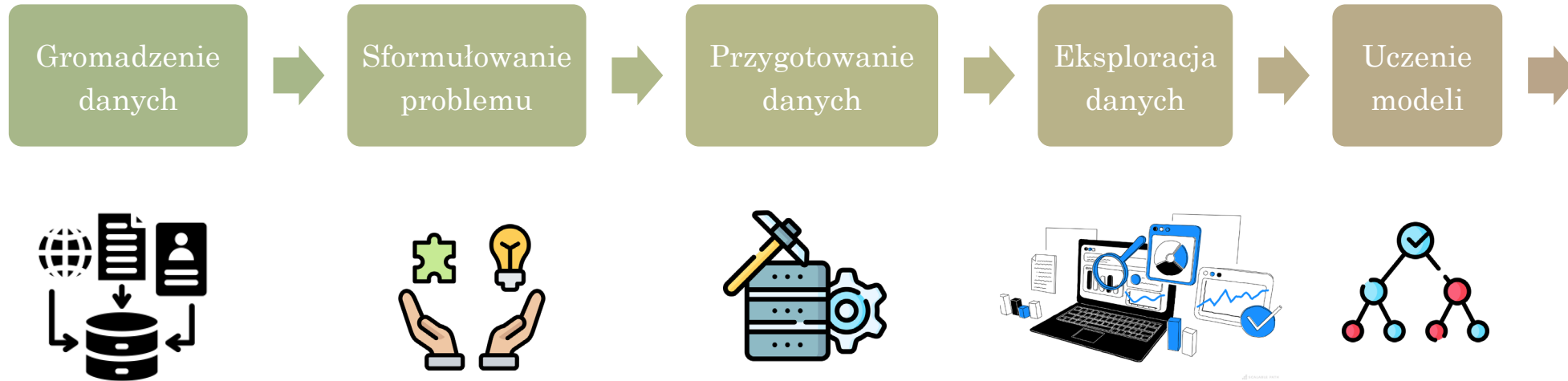
Potok wydobywania wiedzy z danych



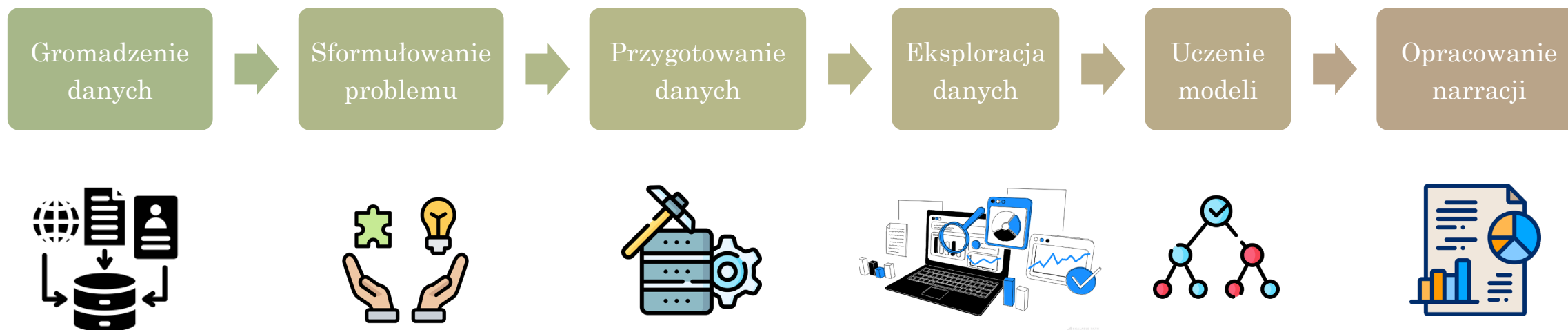
Potok wydobywania wiedzy z danych



Potok wydobywania wiedzy z danych



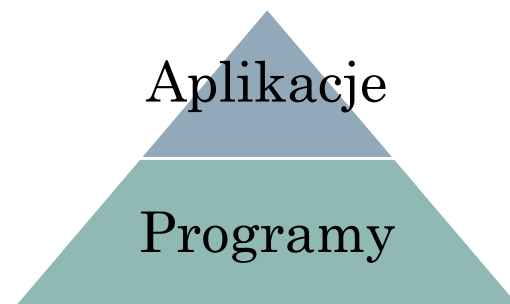
Potok wydobywania wiedzy z danych



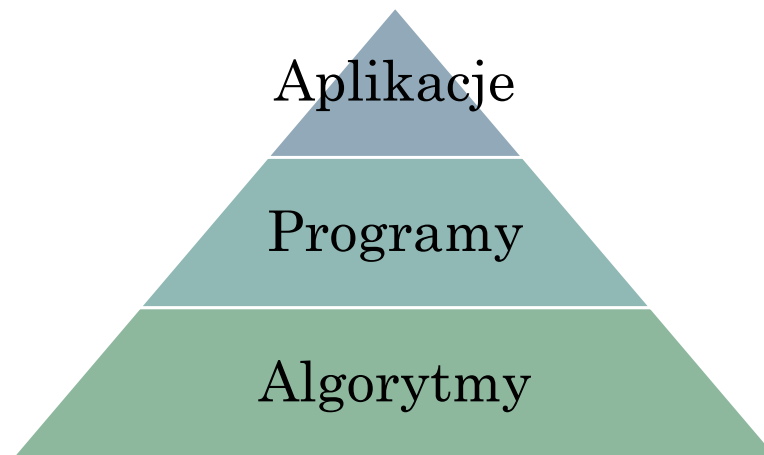
Skąd się biorą aplikacje?



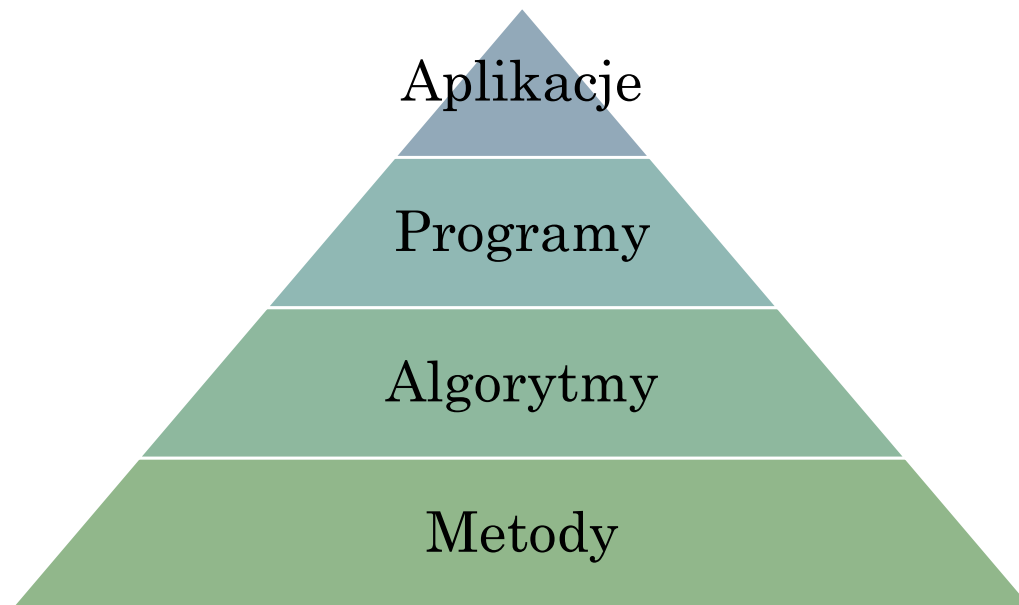
Skąd się biorą aplikacje?



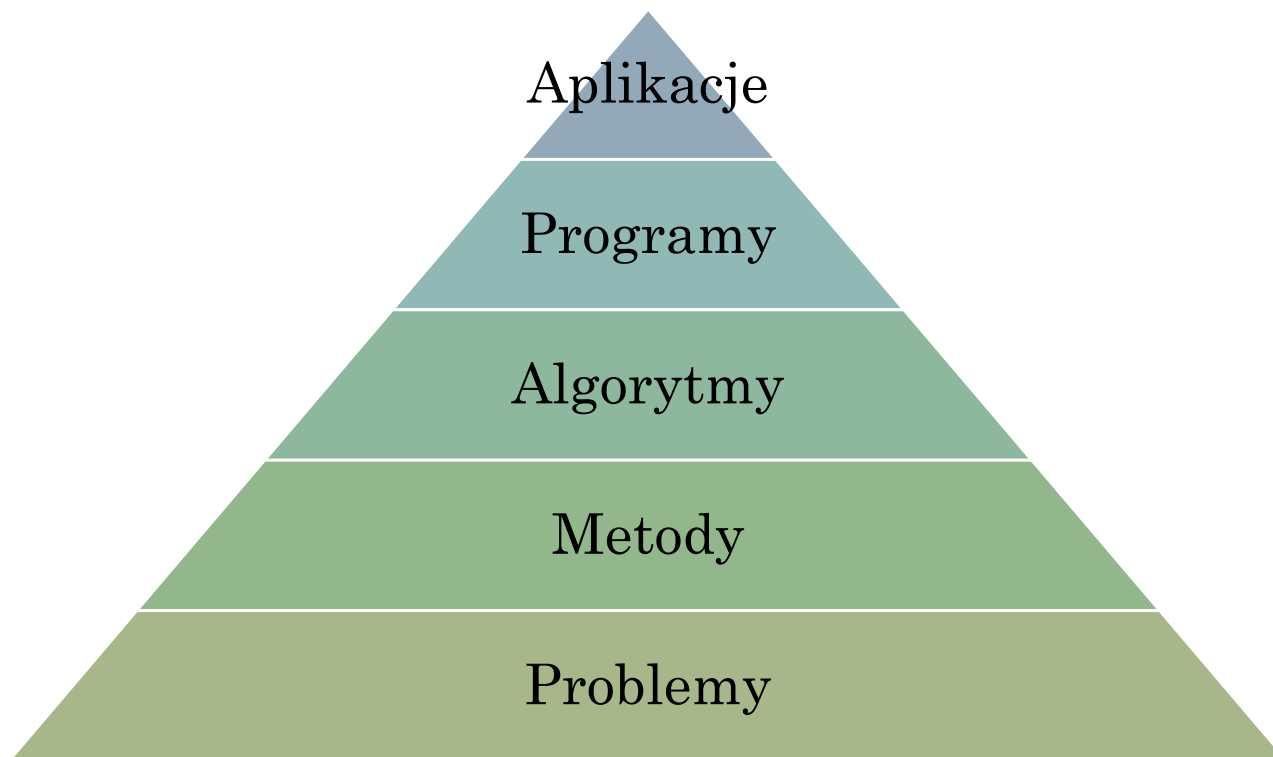
Skąd się biorą aplikacje?



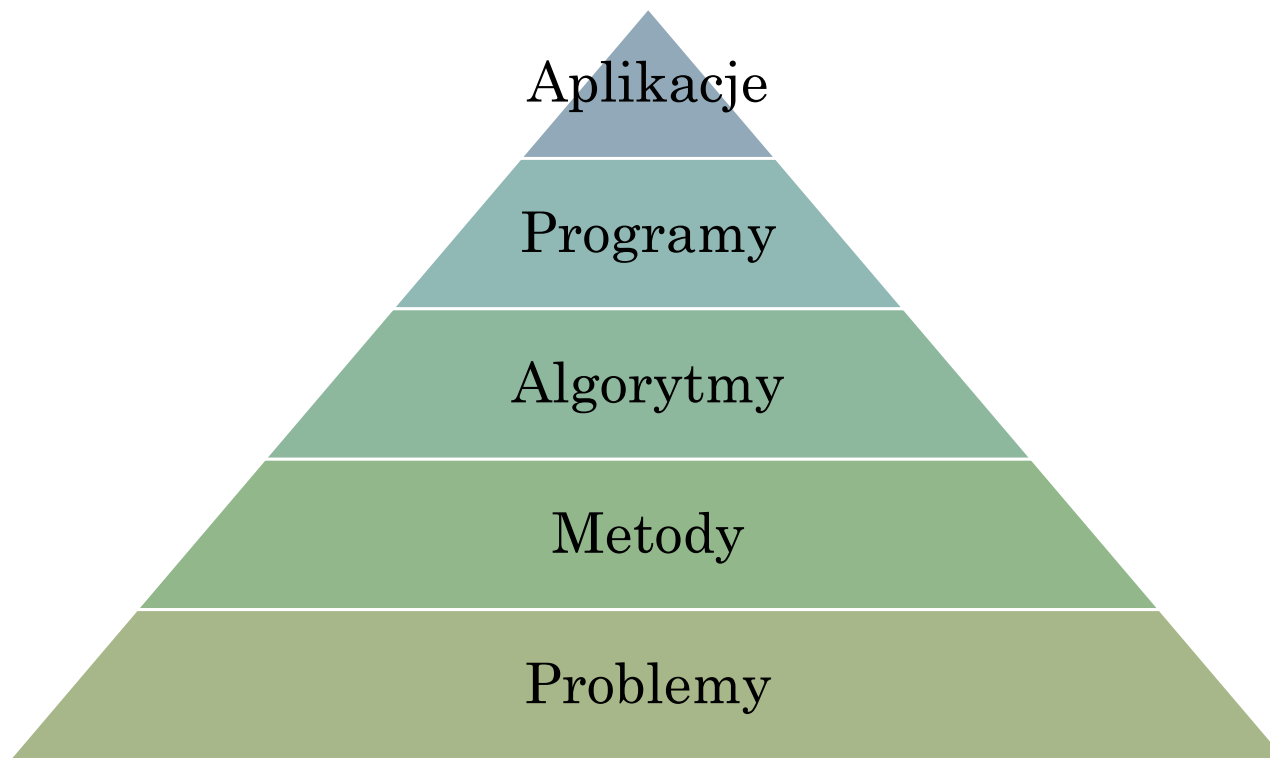
Skąd się biorą aplikacje?



Skąd się biorą aplikacje?



Skąd się biorą aplikacje?



analiza, algebra, statystyka, optymalizacja



<https://medium.com/@jaroslaw.drapala>

