

Uczenie maszynowe

Wprowadzenie

semestr letni 2024/25

wykład: 10h

laboratorium: 14h

dr inż. Mateusz Gorczyca

mateusz.gorczyca@wroclaw.merito.pl

Konsultacje:

wkrótce pojawią się w planie

Plan wykładu

- Treść przedmiotu
- Literatura
- Sprawy formalne
- Część merytoryczna:
 - Wprowadzenie do uczenia maszynowego

Treść przedmiotu - wykład

- Wprowadzenie
 - definicja i zastosowania
 - typy i algorytmy uczenia maszynowego
 - podstawowe elementy uczenia maszynowego

Treść przedmiotu - wykład

- Uczenie nadzorowane
 - Klasyfikacja
 - drzewa decyzyjne
 - naiwny klasyfikator bayesowski
 - sieci neuronowe
 - Regresja
 - metody gradientowe
 - sieci neuronowe

Treść przedmiotu - wykład

- Uczenie bez nadzoru
 - Grupowanie
 - k-średnich
- Uczenie ze wzmacnieniem

Treść przedmiotu - ćwiczenia

- Praktyczne wykorzystanie treści wykładu
- Jupyter/Python w środowisku Anaconda lub Colab
- biblioteki uczenia maszynowego: scikit-learn, ...
- elementarne przetwarzanie i struktury danych: NumPy, pandas
- wizualizacja: matplotlib

Literatura

- Aurélien Géron

Uczenie maszynowe z użyciem
Scikit-Learn, Keras i TensorFlow
. Wydanie III

Helion, 2023

Rozdziały 1-9, 18

Github

- scikit-learn.org



Sprawy formalne

- Obecność na ćwiczeniach (dopuszczalna jedna nieobecność)
- Oceny z list zadań / projektu z uwzględnieniem pracy w trakcie ćwiczeń
- Przekroczenie terminu oddania listy zadań powoduje obniżenie oceny proporcjonalne do wielkości opóźnienia

Sprawy formalne

- Minimalne wymagania do zaliczenia:
 - obecności na ćwiczeniach,
 - praca na ćwiczeniach,
 - oddanie minimum z listy zadań oraz projektu,
 - zaliczenie egzaminu (ostatni zjazd).

Sprawy formalne

- Plagiaty – w zależności od skali: ocena niedostateczna, komisja dyscyplinarna
- Ustawa o prawie autorskim
- Prace wykonujemy samodzielnie, chyba że wprost zalecono współpracę
- Nie można wykorzystywać w pracy sztucznej inteligencji, chyba że wprost to zalecono

Sprawy formalne

- Ocena końcowa: średnia z oceny z kolokwium oraz oceny z ćwiczeń
- Ocena z ćwiczeń: średnia oceny z listy zadań i z projektu
- Przy wysyłaniu prac mailem lub na Teams/Moodle:
 - nie kompresujemy plików,
 - tytuł maila jest istotny,
 - nazwa pliku jest istotna,
 - format pliku jest istotny.

Sprawy formalne

- Przy pisaniu zapytań lub informacji za pośrednictwem **czatu na Teams** proponuję pomijać wszelkie standardowe zwroty rozpoczynające ("Dzień dobry", "Witam", "Szanowny Panie Doktorze", itp.) Uprości to i przyspieszy komunikację.

Uczenie maszynowe - wprowadzenie

Czym jest uczenie maszynowe

- **Sztuczna inteligencja** jest działem informatyki badającym i tworzącym oprogramowanie umożliwiające maszynom postrzeganie otaczającego świata oraz wykorzystywanie **uczenia się** i inteligencji do osiągnięcia wyznaczonych celów

Czym jest uczenie maszynowe

- **Uczenie maszynowe** skupia się na badaniu i tworzeniu **algorytmów statystycznych**, które mogą się **uczyć** z danych i **uogólniać** to na dane, których jeszcze nie widziały. Umożliwia to wykonywanie przez nie zadań, do których nie były jawnie zaprogramowane.

Problem klasyfikacji

Klasyfikacja

- **Problem klasyfikacji** (statystycznej)

- **dane** trenujące:

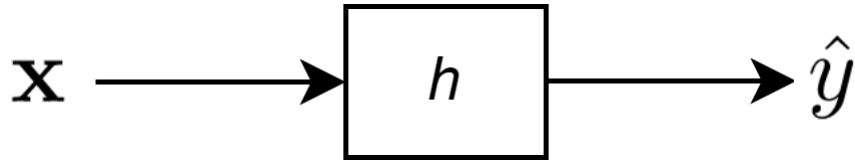
$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$$

gdzie:

- $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ik}]^T$ jest i -tym przykładem uczącym,
- x_{ik} jest wartością k -tej cechy i -tego przykładu uczącego,
- y_i jest klasą, do której należy i -ty przykład uczący.

Klasyfikacja

- **Problem klasyfikacji** (statystycznej)
 - **szukane:** klasyfikator h odgadujący klasę y dla zadanego obiektu x
 - dla odróżnienia od rzeczywistej klasy y , wynik zwracany przez klasyfikator oznaczamy \hat{y}



Przykładowe dane trenujące – diagnostyka raka piersi:

```
from sklearn.datasets import load_breast_cancer
```

```
data = load_breast_cancer()
```

```
print(data.DESCR)
```

```
.. _breast_cancer_dataset:
```

```
Breast cancer wisconsin (diagnostic) dataset
```

```
-----
```

```
**Data Set Characteristics:**
```

```
:Number of Instances: 569
```

```
:Number of Attributes: 30 numeric, predictive attributes and the class
```

```
:Attribute Information:
```

- radius (mean of distances from center to points on the perimeter)
- texture (standard deviation of gray-scale values)
- perimeter
- area
- smoothness (local variation in radius lengths)
- compactness ($\text{perimeter}^2 / \text{area} - 1.0$)

Wszystkie instancje (przykłady) trenujące należą do jednej z 2 klas:

```
data.target_names
```

```
array(['malignant', 'benign'], dtype='<U9')
```

Wczytanie danych treningowych – wartości cech (X):

```
X, y = load_breast_cancer(return_X_y=True, as_frame=True)
```

X

	mean radius	mean texture	mean perimeter	...	worst concave points	worst symmetry	worst fractal dimension
0	17.99	10.38	122.80	...	0.2654	0.4601	0.11890
1	20.57	17.77	132.90	...	0.1860	0.2750	0.08902
2	19.69	21.25	130.00	...	0.2430	0.3613	0.08758
...
566	16.60	28.08	108.30	...	0.1418	0.2218	0.07820
567	20.60	29.33	140.10	...	0.2650	0.4087	0.12400
568	7.76	24.54	47.92	...	0.0000	0.2871	0.07039

instancja
danych
treningowych

cecha

569 rows × 30 columns

Wczytanie danych treningowych – etykiety klas (y):

y

0 0

1 0

2 0

..

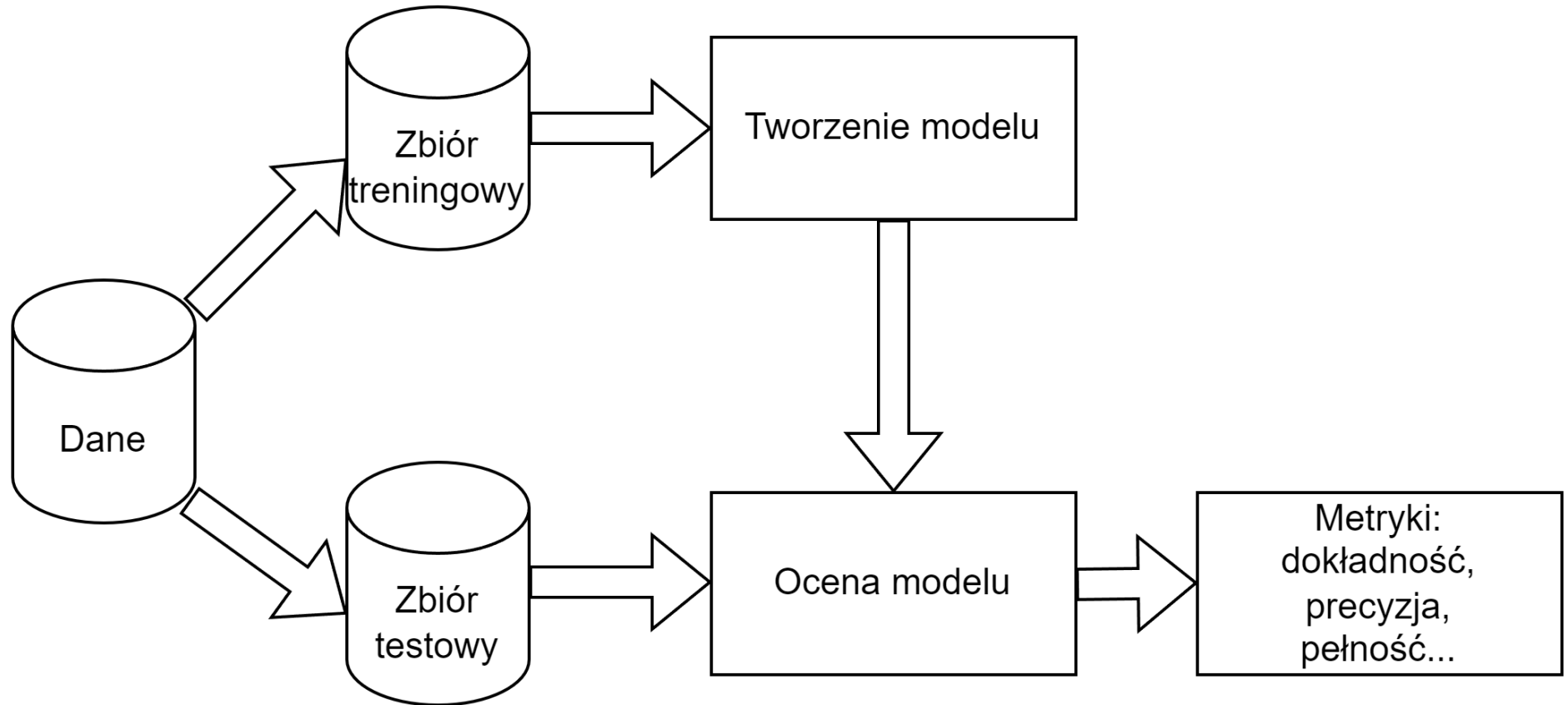
566 0

567 0

568 1

Name: target, Length: 569, dtype: int32

Klasyfikacja



Problem regresji

Regresja

- **Problem regresji**

- dane trenujące:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$$

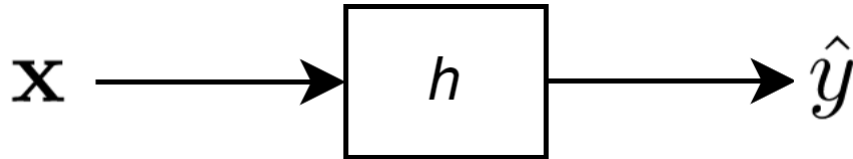
gdzie:

- $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ik}]^T$ jest i -tym przykładem uczącym,
- x_{ik} jest wartością k -tej zmiennej niezależnej i -tego przykładu uczącego,
- y_i jest wartością zmiennej objaśnianej.

Regresja

- **Problem regresji**

- szukane: model h pozwalający estymować y dla zadanego obiektu x
- dla odróżnienia od rzeczywistej wartości y , wynik zwracany przez model oznaczamy \hat{y}



Regresja

- Dane wejściowe – procentowa zawartość tłuszczu [428]:

```
from sklearn.datasets import fetch_openml
```

```
data = fetch_openml(name='bodyfat', version=1)  
print(data.DESCR)
```

```
**Author**: Roger W. Johnson
```

```
**Source**: [UCI (not available anymore)](https://archive.ics.uci.edu/ml)
```

```
**Please cite**: None.
```

Short Summary:

Lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men.

Classroom use of this data set:

This data set can be used to illustrate multiple regression techniques. Accurate measurement of body fat is inconvenient/costly and it is desirable to have easy methods of estimating body fat that are not inconvenient/costly.

```
[432]: from sklearn.datasets import fetch_openml
```

```
X, y = fetch_openml(name='bodyfat', version=1, as_frame=True, return_X_y=True)
```

```
[442]: X
```

```
[442]:
```

	Density	Age	Weight	...	Biceps	Forearm	Wrist
--	---------	-----	--------	-----	--------	---------	-------

0	1.0708	23	154.25	...	32.0	27.4	17.1
---	--------	----	--------	-----	------	------	------

1	1.0853	22	173.25	...	30.5	28.9	18.2
---	--------	----	--------	-----	------	------	------

2	1.0414	22	154.00	...	28.8	25.2	16.6
---	--------	----	--------	-----	------	------	------

3	1.0751	26	184.75	...	32.4	29.4	18.2
---	--------	----	--------	-----	------	------	------

4	1.0340	24	184.25	...	32.2	27.7	17.7
---	--------	----	--------	-----	------	------	------

...
-----	-----	-----	-----	-----	-----	-----	-----

247	1.0736	70	134.25	...	25.6	25.7	18.5
-----	--------	----	--------	-----	------	------	------

248	1.0236	72	201.00	...	35.2	28.6	20.1
-----	--------	----	--------	-----	------	------	------

249	1.0328	72	186.75	...	31.3	27.2	18.0
-----	--------	----	--------	-----	------	------	------

y

0 12.3

1 6.1

2 25.3

3 10.4

4 28.7

...

247 11.0

248 33.6

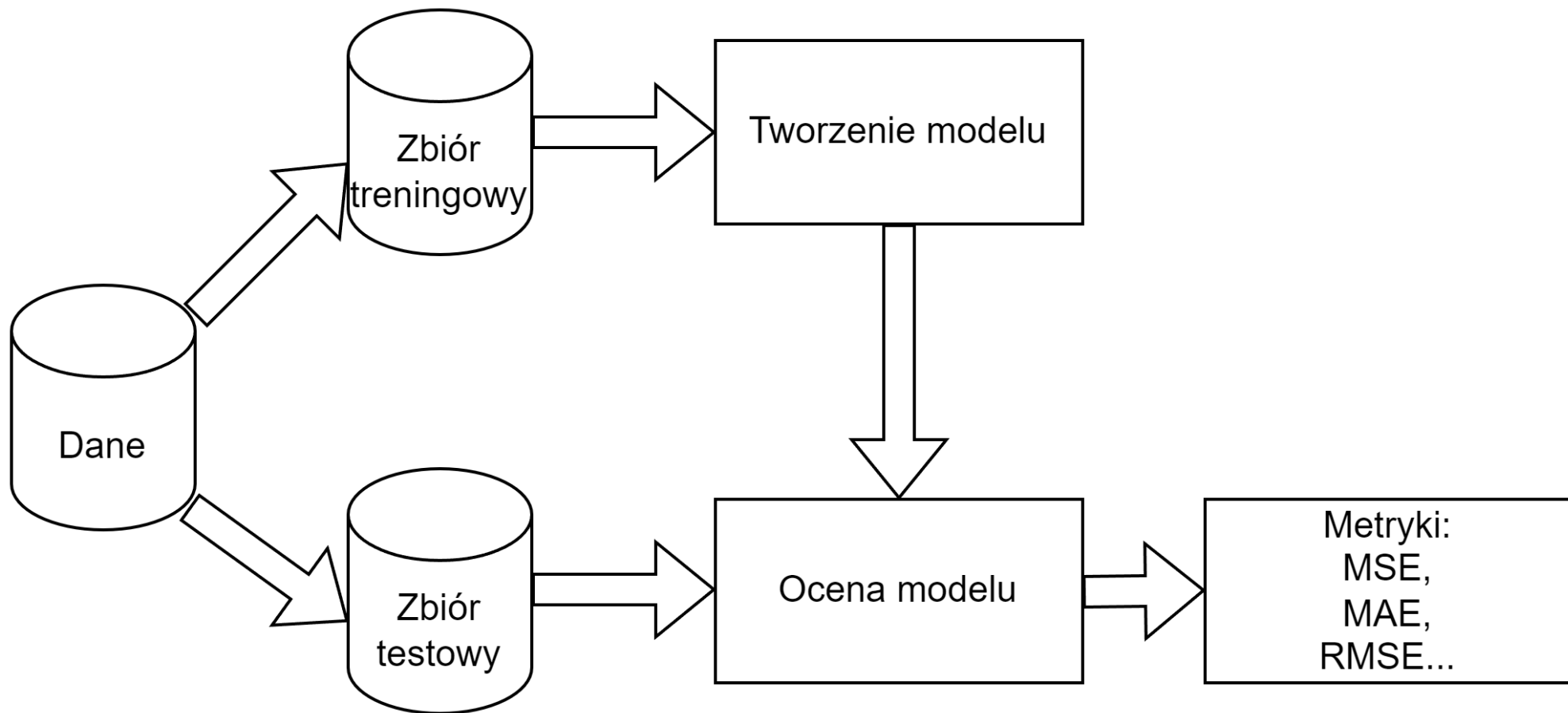
249 29.3

250 26.0

251 31.9

Name: class, Length: 252, dtype: float64

Regresja



Uczenie maszynowe

- Dalsza część wykładu jest w notatniku Jupyter „UM – Wykład1 lato 2025.ipynb”