

# Laboratorium: Regresja

March 17, 2022

## 1 Cel/Zakres

- Przeprowadzenie regresji.
- Porównanie regresorów.

## 2 Przygotowanie danych

Dany jest zbiór argumentów ( $X$ ) oraz wartości ( $y$ ) funkcji jednej zmiennej w postaci obiektu DataFrame `df`:

```
import numpy as np
size = 300
X = np.random.rand(size)*5-2.5
w4, w3, w2, w1, w0 = 1, 2, 1, -4, 2
y = w4*(X**4) + w3*(X**3) + w2*(X**2) + w1*X + w0 + np.random.randn(size)*8-4
df = pd.DataFrame({'x': X, 'y': y})
df.to_csv('dane_do_regresji.csv', index=None)
df.plot.scatter(x='x', y='y', height=600)
```

Podziel ww. zbiór na zbiory: uczący oraz testujący w proporcji 80:20.

## 3 Regresja

1. Wykonaj następujące regresje na ww. zbiorze:
  1. liniową,
  2. KNN, dla  $k = 3$  oraz  $k = 5$ ,
  3. wielomianową 2, 3, 4 i 5 rzędu.
2. Przeanalizuj działanie każdej z otrzymanych funkcji regresyjnych. Porównaj ich przebiegi z rozkładem zbioru danych.
3. Zapisz w osobnym DataFrame wartości MSE dla zbiorów uczących i testujących dla ww. regresorów; kolumny: `train_mse`, `test_mse`, wiersze: `lin_reg`, `knn_3_reg`, `knn_5_reg`, `poly_2_reg`, `poly_3_reg`, `poly_4_reg`, `poly_5_reg`. Zapisz ww. DataFrame do pliku Pickle o nazwie: `mse.pkl`  
  
7 pkt
4. Zapisz do pliku Pickle o nazwie `reg.pkl` listę krotek zawierających obiekty reprezentujące regresory: `[(lin_reg, None), (knn_3_reg, None), (knn_5_reg, None), (poly_2_reg,`

```
poly_feature_2), (poly_3_reg, poly_feature_3), (poly_4_reg, poly_feature_4),  
(poly_5_reg, poly_feature_5)]
```

6 pkt

## 4 Prześlij raport

Prześlij plik o nazwie `lab3.py` realizujący ww. ćwiczenia.

Sprawdzone będzie, czy skrypt Pythona tworzy wszystkie wymagane pliki oraz czy ich zawartość jest poprawna.