

Wstęp do sztucznej inteligencji

Laboratorium 3

28 marca 2023

Zasady oceniania

W ramach zajęć należy po kolei przerobić zadania umieszczone w dalszym ciągu tej instrukcji. W wyniku pracy ma powstać raport w formacie PDF, zawierający nastawienia metod, wyniki w postaci wykresów i wnioski. Dodatkowo, należy przesłać kod powstały w ramach zajęć. Akceptowalny jest też wydruk z Colab/Jupyter Notebook zapisany w formacie pdf.

Raport z wykonanych zadań w formacie PDF proszę przesłać na platformie Moodle wraz z kodami źródłowymi programów (pliki .py).

UWAGA: Termin oddania zadania jest ustawiony w systemie moodle. W przypadku nie oddania zadania w terminie, uzyskana ocena będzie zmniejszana o 0,5 za każdy zaczęty tydzień opóźnienia. Zadania oddawane później niż miesiąc po terminie ustawionym na moodle są oddawane i rozliczane w trybie indywidualnym na zajęciach lub po umówieniu się z prowadzącym.

UWAGA: W przypadku wysłania zadania w formie niezgodnej z opisem w instrukcji prowadzący zastrzega prawo do wystawienia oceny negatywnej za taką pracę. Przykład: wysłanie .zip lub .pdf tam, gdzie był wymagany plik tekstowy z rozszerzeniem .py.

1 Zadania do wykonania

1. Zapoznaj się z dokumentacją klasy `LinearRegression` oraz opisem metody regresji liniowej

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression

2. Dla zbioru danych california housing wykonaj:

Opis zbioru danych: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_california_housing.html#sklearn.datasets.fetch_california_housing

https://scikit-learn.org/stable/datasets/real_world.html#california-housing-dataset

- (a) Wczytaj zbiór danych.
- (b) Wykonaj 8 wykresów pokazujących zależność zależnej y od poszczególnych zmiennych x_i (kolumny ze zbioru danych X). Podpisz osi wykresów.
- (c) Podziel dane na zbiór uczący i testujący za pomocą funkcji `train_test_split` w proporcji 0.7 danych uczących i 0.3 testowych: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.
- (d) Naucz model `LinearRegression` na danych uczących.
Przykład użycia `LinearRegression` na innym zbiorze danych https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html.
- (e) Oblicz błędy MAE (mean absolute error) oraz MSE (mean squared error) na danych testowych oraz danych uczących, oceń jakość tego modelu.
- (f) Dla każdej z 8 zmiennych wykonaj:
 - i. Wybierz wybraną zmienną z macierzy `X_train` oraz `X_test`, tak by były to macierze jednokolumnowa.

- ii. Naucz model `LinearRegression` na tej wybranej zmiennej oraz na wektorze `y_train`.
 - iii. Narysuj wykres jak w punkcie (b) oraz umieść na nim linie regresji (proszę zobaczyć przykład użycia `Linear Regression` podany w punkcie (d)).
 - iv. Oblicz i wypisz także wartości `MAE` oraz `MSE` dla tego modelu.
 - (g) Przeanalizuj wyniki uzyskane w punktach (e) i (f). Czy uczenie modelu na pojedynczych zmiennych zmniejsza błąd czy zwiększa go?
3. W raporcie proszę umieścić wyniki poszczególnych etapów wykonywania zadania, jak to zrzuty ekranu części wygenerowanych/wczytanych danych, wykresy z danymi, wykresy z regresją i inne.