

Wstęp do sztucznej inteligencji

Laboratorium 2

17 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 macierzy konfuzji, a także 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. Zainstaluj biblioteki `numpy`, `matplotlib` i `scikit-learn`.
2. Zapoznaj się z funkcjonalnościami dostępnymi w podmodule `sklearn.datasets` <https://scikit-learn.org/stable/datasets.html#datasets>. Zadania w tym laboratorium będą wykonywane na zbiorze danych Iris.
3. Zapoznaj się z dokumentacją klasy `SVC` oraz opisem metody SVM (SVC).
<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
4. Dla zbioru danych breast cancer wykonaj:
Opis zbioru danych: https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html
 - (a) Wczytaj zbiór danych.
 - (b) 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.
 - (c) Naucz klasyfikatora `SVC` na danych uczących.
 - (d) Oblicz dokładność klasyfikacji (accuracy) na zbiorze danych testowych. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html.
 - (e) Powtórz punkty 4c i 4d w celu dla znalezienia optymalnego parametru C oraz dla przebadania wpływu argumentu `kernel`, a mianowicie opcje `kernel='linear'` oraz `kernel='rbf'`.
 - (f) Naucz dwa klasyfikatory: jeden z `kernel='linear'` oraz drugi z `kernel='rbf'`, a także ustalonych dla nich wartości C , a następnie dla każdego z nich wykonaj:

- i. Stwórz macierz konfuzji, pokazującą dokładność klasyfikacji na danych testowych (https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.html#confusion-matrix).
 - ii. Wykonaj wizualizację całości danych za pomocą wykresu punktowego (scatter), pokazującego rzeczywisty podział klas, a następnie wykonaj taki sam wykres, ale zamiast rzeczywistego wektora klas użyj wyniku predykcji zbudowanego klasyfikatora. Punkty na wykresie muszą być pokolorowane według klas!
Uwaga: Ze względu na dużą liczbę atrybutów w tym zbiorze danych konieczne jest zastosowanie metod redukcji wymiarowości, na przykład PCA (<https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>).
5. W raporcie proszę umieścić wyniki poszczególnych etapów wykonywania zadania, jak to zrzuty ekranu części wygenerowanych/wczytanych danych, dobrana liczba N sąsiedzi, ewentualnie wyniki dokładności dla innych liczb N, wykresy pokazujące rzeczywisty podział klas oraz podział klas według predykcji nauczonego klasyfikatora.