

WSI - ćwiczenie 7.

Modele bayesowskie

1 czerwca 2023

1 Sprawy organizacyjne

1. Ćwiczenie realizowane jest samodzielnie.
2. Ćwiczenie wykonywane jest w języku Python.
3. Ćwiczenie powinno zostać oddane najpóźniej na ostatnich zajęciach. W ramach oddawania ćwiczenia należy zademonstrować prowadzącemu działanie kodu oraz utworzyć pull request (z kodem oraz raportem) który prowadzący będzie mógł komentować.
4. Rozwiązanie ćwiczenia powinno być zgodne z szablonem z repozytorium <https://gitlab-stud.elka.pw.edu.pl/jlyskawa/wsi-template>.
5. Dokumentacja powinna być w postaci pliku .pdf, .html albo być częścią notebooka jupyterowego. Powinna zawierać opis eksperymentów, uzyskane wyniki wraz z komentarzem oraz wnioski.
6. Na ocenę wpływa poprawność oraz jakość kodu i dokumentacja.
7. Można korzystać z pakietów do obliczeń numerycznych, takich jak *numpy*
8. Można skorzystać z pakietu *scikit-learn* w celu załadowania zbioru danych oraz wykonania walidacji krzyżowej
9. Implementacja powinna być ogólna, w szczególności powinna być możliwa do zastosowania do dowolnego zbioru danych o ciągłych wartościach atrybutów.

2 Ćwiczenie

Celem ćwiczenia jest implementacja naiwnego klasyfikatora Bayesa. *Uwaga: dane w zbiorze są ciągłe, należy to uwzględnić zakładając ciągły rozkład, np. normalny.*

Następnie należy wykorzystać stworzony algorytm do stworzenia i zbadania jakości klasyfikatorów dla zbioru danych Breast Cancer Wisconsin Dataset (https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_breast_cancer.html#sklearn.datasets.load_breast_cancer).

Należy sprawdzić jakość zarówno różnymi podziałami na zbiór trenujący i walidacyjny jak i różnymi uruchomieniami k-krotnej walidacji krzyżowej. Należy porównać obie metody pomiaru.