

# SWD cwiczenia

**Zadanie:** Stworzyć klasyfikator dla zbioru danych IRIS używający algorytmu KNN. Przeprowadzić klasyfikację dla różnych wartości K.

**Dane wejściowe:** Iris Data Set : <http://archive.ics.uci.edu/ml/datasets/Iris>

## Wynik działania programu:

Przykładowych kilka klasyfikacji wypisywanych do pliku/na konsole z podaną faktyczną klasą decyzyjną i tą przydzieloną przez klasyfikator dla jednego wybranego K (lub dla najlepszego w przypadku wariantu na DB i wyżej).

Lista wartości K, dla których testowano klasyfikator. Dla każdej wartości podać miarę **precision** (dokładności) klasyfikatora.

## Skala ocen:

DST - działający klasyfikator dla jednej ustawionej na sztywno wartości K

DB - to co na DST plus dodatkowo kilka wartości K, podanie miary **precision**, wybranie tego K dla którego są najlepsze wyniki.

BDB - to co DB plus dodatkowo 3 segmentowa walidacja w każdym przebiegu klasyfikatora.

Forma zaliczenie: prezentacja działającego programu. Odpowiedź na kilka pytań związanych z projektem.

Implementacja w dowolnym wybranym języku programowania.

## Pomocne wzory:

Odległość dwóch punktów od siebie:

$$d(A, B) = \sqrt{(x_{1A} - x_{1B})^2 + (x_{2A} - x_{2B})^2 + \dots + (x_{nA} - x_{nB})^2} = \sqrt{\sum_{i=1}^n ((x_{iA} - x_{iB})^2)}$$

**Precision = true positive / (true positive + false positive)** , gdzie:

- true positive - ilość przypadków z klasa X poprawnie sklasyfikowanych jako należące do klasy decyzyjnej X
- false positive - ilość przypadków sklasyfikowanych jako należące do klasy decyzyjnej X które w rzeczywistości mają inną klasę decyzyjną
- true positive + false positive - wszystkie przypadki sklasyfikowane jako należące do klasy decyzyjnej X