Eksploracja danych

Projekt 2: Klasyfikacja przy użyciu KNN

To zadanie pochodzi z dziedziny kryminalistyki. Celem jest sklasyfikowanie szeregu próbek szkła w celu określenia, jakiego rodzaju jest każda z nich (szkło budowlane, szyby samochodowe, szyby reflektorów samochodowych itp.).

Klasyfikacja ma być oparta na metodzie k najbliższych sąsiadów, którą należy wdrożyć w SciLab-ie.

Zestaw danych można znaleźć w pliku

`http://www.math.us.edu.pl/~pgladki/teaching/2021-2022/ed-p02.mat`

Kiedy załadujesz ten plik do SciLab-a, znajdziesz trzy macierze:

1. *GlassData:* macierz 163×9, w której każdy wiersz zawiera jedną obserwację, a każda kolumna odpowiada określonemu pomiarowi (tj. wartości współczynniku załamania światła i ilości magnezu),

2. *GlassClasses:* macierz 163×1, która zawiera klasy (od 1 do 6) odpowiadające każdej obserwacji w *GlassData*, oraz

3. *TestData* macierz 30×9 obserwacji, gdzie klasa obiektów jest nieznana.

Twoje zadanie jest następujące:

• Zaimplementuj algorytm k najbliższego sąsiada w programie SciLab. Użyj miary odległości Manhattan (taksówkowej)

• Przetwarzaj wstępnie dane, aby dopasować je do potrzeb algorytmu. Nie ma brakujących wartoąci i możesz założyć, że nie ma szumu.

• Podziel dane na odpowiednie zestawy testowe i treningowe i znajdź najlepszą wartość k. Należy pamiętać, że żadne praktyczne zasady nie są akceptowane jako wybór k.

• Użyj swojej implementacji, aby sklasyfikować niesklasyfikowane dane, które można znaleźć w macierzy *TestData*.