

12.1

Zbiór *kwadraty.txt* zawiera sztucznie wygenerowane dane dwuwymiarowe, składające się z czterech niezależnych prób (każda o liczebności 20) z rozkładów jednostajnych na 4 kwadratach o środkach usytuowanych na przekątnej większego kwadratu. Na podstawie wykresu rozproszenia danych można zaproponować prawdopodobną liczbę skupień $k = 4$.

a) Zastosuj metodę k -średnich. Na wykresie rozproszenia zaznacz kolorami otrzymane skupienia. Oblicz: sumę kwadratów odległości między punktami w skupieniach. Podaj liczbę elementów w poszczególnych skupieniach.

b) Zastosuj metodę hierarchiczną dla odmienności:

- typu najbliższego sąsiada (ang. single linkage),
- typu najdalszego sąsiada (ang. complete linkage),
- typu średnia (ang. average linkage).

Dokonaj przycięcia dendrogramu na poziomie skupień $k = 4$. Sporządź wykres rozproszenia z zaznaczeniem kolorami skupień. Sporządź dendrogram.

12.2

Dane *wrecord.dat* zawierają rekordy krajowe w wybranych konkurencjach biegowych- są to zmienne: **100m**, **200m**, **400m**, **800m**, **1500m**, **3000m**, **marathon**. Czasy dla trzech najkrótszych biegów podane są w sekundach, pozostałe czasy w minutach.

Dokonaj zamiany minut na sekundy dla czasów biegów długich. Następnie wartości każdej zmiennej podziel przez odchylenie standardowe tej zmiennej.

a) Zastosuj metodę hierarchiczną z odmiennością najdalszego sąsiada. Przytnij dendrogram aby otrzymać 3 skupienia. Wyznacz środki 3 skupień.

b) Zastosuj metodę k -średnich, podając jako wartości początkowe parametru `centers` środki skupień wyznaczone w punkcie (a).

c) Wyznacz dwie pierwsze składowe główne. Sporządź wykres (w pierwszych dwóch składowych); każde ze skupień oznacz innym kolorem.

12.3

Zastosuj metodę `Mclust` do danych *kwadraty.txt*.

12.4

Dane *congress.txt* zawierają macierz rozbieżności w głosowaniach dla 15 kongresmenów, dotyczących 19 głosowań (3 możliwe wyniki głosowania: za, przeciw, wstrzymanie się od głosu). Odległość między dwoma kongresmenami jest wyznaczona przez liczbę głosowań, w których głosowali różnie. Wśród głosujących znajdują się: demokraci (D) i republikanie (R).

a) Wykonaj hierarchiczną analizę skupień z odległością średnią, przedstaw wyniki za pomocą dendrogramu.

b) Dokonaj skalowania wielowymiarowego (funkcja `cmdscale`) i wyrysuj wykres na płaszczyźnie, podając nazwisko każdego z głosujących.