

14.1

Zbiór danych `wines` w pakiecie `kohonen` zawiera informację dotyczące win pochodzących z regionu Piemont we Włoszech. Każde z win jest opisane 13 atrybutami. Wina pochodzą od jednego z 3 plantatorów: Nebbiolo, Barberas, Grignolino (informacje o rodzaju plantatora znajdują się w obiekcie `vintages`). Dokonaj centrowania i standaryzacji danych (funkcja `scale`).

Wykonaj analizę skupień stosując algorytm map samoorganizujących (SOM). **a)** Stwórz siatkę punktów (dowolnie gęstą). Wyznacz współrzędne środków. Przedstaw graficznie otoczenia środków wraz z punktami im przyporządkowanymi. Wyznacz wektor opisujący przynależność do poszczególnych środków.

b) Korzystając z hierarchicznej analizy skupień dokonaj podziału danych na trzy skupienia na podstawie wyników z powyższego podpunktu. Zaznacz podział na skupienia na mapie.

c) Oceń skuteczność powyższego podejścia do identyfikacji oddzielnych plantatorów (zidentyfikuj, które skupienie odpowiada któremu plantatorowi i sprawdź, ile obserwacji zostało źle zaklasyfikowanych). Uwaga: skupienia pochodzące z hclust odpowiadają środkom siatki, potem trzeba jeszcze zidentyfikować, które punkty są przyporządkowane któremu węzłowi w siatce.

d) Porównaj powyższe podejście z hierarchiczną analizą skupień zastosowaną bezpośrednio do danych.

e) Korzystając z PCA, przedstaw obserwacje i środki siatki na płaszczyźnie. Zaznacz kolorem skupienie każdej z obserwacji.

f) Stwórz siatkę składającą się z 3 punktów. Wyznacz współrzędne środków. Przedstaw graficznie otoczenia środków wraz z punktami im przyporządkowanymi. Wypisz przyporządkowanie punktów do środków.

g) Porównaj środki wyznaczone w podpunkcie e) ze średnią wartością zmiennych w grupach.