

Worksheet 3

Problem 1 (Clustering)

We use again the `protein.txt` dataset, available from <https://www.math.uzh.ch/furrer/download/sta121/protein.txt>.

- (a) Conduct a cluster analysis, using all the three agglomeration methods of hierarchical clustering seen during the lesson. Discuss the figures.
- (b) Select one method. Is it possible to cut the tree such that the resulting classification can be interpreted in a meaningful way?
- (c) Compare the results obtained from the application of a PCA analysis and a clustering method. Can you see any similarities or dissimilarities in the analysis performed?

Problem 2

Write a function `my.kmean()`, which takes as inputs a vector x and the number of clusters k , does the kmeans clustering in one dimension, and returns a data.frame with (1) the original x as one column, and (2) the assigned cluster numbers as the second column.

For simplicity, you can specify the number of iterations after which the algorithm stops.

Apply your function with: `x <- c(1,2,1,3,2,6,5,7,6,12)`, `k = 3`