Kursus Bigdata F2017

Problem Week 4

Prepare an R script file generating and saving plots of examples on clustering processes as detailed in the following:

Q1. On the basis of the R code in [R_Intro_Clustering, 2017] carry out a single link clustering on the neutrient data using method = “single”. Compare the result to the “average” link clustering.

Q2. Using the reference [Charrad, 2014] Table 2 page 19 , select two indices and verify these using the R code in [R_Intro_Clustering, 2017]. Notice that not all indices can be applied on all clustering problems.

Q3. On the basis of the R code in [R_Intro_Clustering, 2017] carry out a PAM (Partitioning Around Medoids) for k=5 and 6 and compare to the k=3 and 4 results.

Q4. Following the example presented in line 144 – 150 in [R_Intro_Clustering, 2017] construct 4 clusters of 2D data using normal distributions. Verify one or two clustering methods on this dataset.

The resulting R codes from Q1, Q2, Q3 and Q4 are inserted into a Problem4_xxx.R script file where xxx are characters chosen from the persons name. Each participant uploads the script to the Campusnet and keeps a copy for later use.

Course material

[Charrad, 2014] Malika Charrad, et al. "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set", Journal of Statistical Software, October 2014, Vol. 61, Issue 6. http://www.jstatsoft.org/

[R_Intro_Clustering, 2017] 5_R_Intro_Clustering.R available at Campusnet.

2017.02.23/jaas