

Input breast cancer data from mlbench library that we often used (`data("BreastCancer")`)

- [1pt] select the data without the classification column and record number column (sample code number);
 - Fit the classless data using k-means
 - Compare the clusterization results with true classification given by class attribute by means of
 1. Contingency table
 2. Comparison of cluster centers with class centers.
 3. Plot the fit results w.r.t. PCA rotation
 - Are classes well separated on the plot?
- [1pt] Repeat the above steps with increased number of attempts (try 1000). Are the results any better? Scale data and repeat the procedure. Are the results better?
- [1pt] Do hierarchical clustering using hclust single link extracting 2 classes. Compare it with true classification using contingency table. Are the results any better than with k-means?
- [1pt] Repeat previous item using Ward metric methods; Are the results any better?
- [1pt] Explain your results with hierarchical clustering in terms of variability of data and inter-class distances.