

Deadline: week 22-26.04 (on tutorials)

K-means

Implement k-means algorithm and use it on iris data. When calculating clusters ignore the label (type of iris) as it is unsupervised algorithm. It is not allowed to hard-code the number of attributes, program should work for any number of attributes.

After each iteration your program should print:

- sum of distances between observations and their centroids (one value)
- the purity of each cluster, use percent of each iris label in each cluster