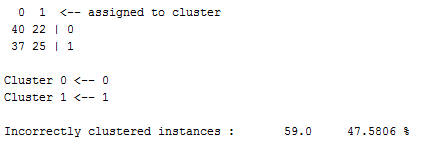
# Association analysis - Lab 2

Cluster data with two different cluster algorithms, SimpleKMeans and Hierarchical clustering. Two different number of clusters is tested for the SimpleKMeans clustering and two different links are tested for the hierarchical clustering.

## SimpleKMeans

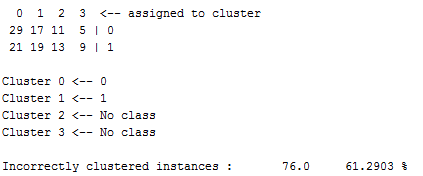
### K = 2 and seed = 10

Performs just marginally better than random guessing with 47.6 % of the observations misclassified.



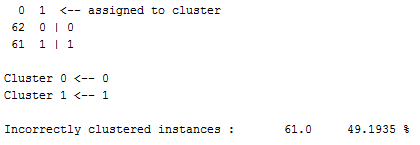
### K = 4 and seed = 10

The clustering with k=4 also performs badly and have clear problems with finding the existing class division. In almost every cluster around half of the observations has *0* as their true class and the other half has class *1* as their true class.

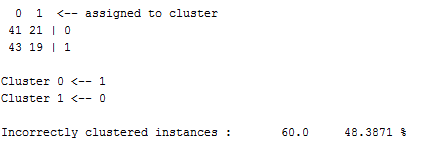


## Hierarchical clustering with k = 2 and single link

The hierarchical clustering with two clusters and single link as distance measure also fails with the objective to find the existing class division. All observations but one are in the same cluster and the misclassification rate is almost 50 %.



## Hierarchical clustering with k = 2 and complete link

The hierarchical clustering with complete link instead of single link produces a different result, but with a just slightly lower misclassification rate. In the obtained clusters are almost half of the observations belonging to either of the true classes. Hence, also for this clustering is the conclusion that it not was able to find the existing class division. 

## Association analysis

Why can the clustering algorithms not find a clustering that matches the class division in the database?

Finally, would you say that the clustering algorithms fail or perform poorly for the monk1 dataset? Why or why not?

Yes, I would say that the clustering algorithms performs poorly for the monk1 dataset. The misclassification rates for the performed clustering's are only slightly better than random guessing.