Clustering Methods

Exercises 1/7, 30.1.2017

1. Find one example clustering application and prepare two-page Power Point (PPT) presentation: one page to give overview of the application, and second page to demonstrate how clustering is applied. Define what is the input and wanted output of this application. Send your powerpoint to samisi@cs.uef.fi by Monday 12:00.

2. Prepare sample data of your application, and convert it into text format: one data object per line, each attribute as one number and separated by space or tabulators. If you do not have own data, select an interesting one from here: http://cs.uef.fi/sipu/datasets/. Then upload your selected data to the clusterator: http://cs.uef.fi/paikka/Radu/clusterator/. Perform clustering, visualize it (double click on data), play with the tool and take screenshot so that it represents the data and the clustering as well as possible.

3. Prepare image file of the output of your data and include it into the PPT file of task 1.

4. Study the *K-means* algorithm using the slides http://cs.uef.fi/pages/franti/cluster/notes.html and testing with the Animator http://cs.uef.fi/sipu/clustering/animator/. Drag the centroids and try to get as bad clustering as possible. What is the maximum CI-value you can reach? You can choose the dataset yourself among the ones available.

5. What improvements would you think these systems would benefit from? Are you able to implement it?

6. What is the time complexity of K-means?