Assignment 3

Objective

In this assignment, you are required to implement two of the clustering approaches introduced in the class.

Detailed Requirement

We have introduced the following two clustering approaches in the class.

- K-means
- Hierarchical clustering

In your assignment, you will implement and test one of these clustering techniques on two data sets from different problem domains. You do not need to create your own data set, since there are many such data sets available on the Web. One of these is the UCI Machine Learning Repository:

http://archive.ics.uci.edu/ml

We shall focus on those data sets with numerical attribute values.

Assignment Submission

Program (40%)

You should submit your program and a readme file with instructions on how to run the program.

Report (60%)

You should summarize your work in the form of a report which should include descriptions of:

1. The data sets you have chosen and their main characteristics.

2. The data preprocessing operations you have applied to improve data quality.

3. The design of the different modules of your program. In describing each module, you should relate your design to the theories you have learnt in the class.

4. The clustering results and your interpretation of the results, based on how well the clustering results match with the actual class memberships of the data records.

5. Limitations and possible improvements of the program.