Data Mining: Project 4

Tasks: Cluster Analysis

Assigned: November 2nd, 2017

Due: November 23rd, 2017

Points: 70

For this project we will use faithful data set from datasets package. The data is about waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. The data set contains 272 observations on 2 variables.

1) eruptions - numeric - Eruption time in mins

2) waiting - numeric - Waiting time to next eruption (in mins)

Write a report covering in detail all steps of the project. **The results have to be reproducible using your report.** Carefully describe every assumption and every step in your report. Also mention any program/code that you are using for your analysis.

1. Modeling [70 points]

1.1 Perform cluster analysis using k-means clustering. [10 points]

- Find a suitable number of clusters.
- Explain how you determine a suitable number of clusters.

1.2 Perform cluster analysis using Hierarchical clustering (with single link, complete link, group average, centroid, and ward's method). [30 points]

- Find the best method for hierarchical clustering.
- Explain how you determine the best method.

1.3 Perform cluster analysis using DBSCAN. [20 points]

- Find suitable values of eps and MinPts.
- Explain how you determine suitable values of eps and MinPts.

1.4 Which method is the best clustering method for the faithful data set? [10 points]

Please submit project 4 report together with your power point presentation.