Data Mining: Assignment 4

Census-Income

Tasks: Cluster Analysis

Assigned: November 13th, 2018

Due: December 4th, 2018

Points: 70 points

For this project we will look again at census-income data. The cleaned and preprocessed data set will be used as an input data set for this assignment.

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]

    1) Find a suitable number of clusters.

    2) 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]

    1) Find the best method for hierarchical clustering.

    2) Explain how you determine the best method.

1.3 Perform cluster analysis using DBSCAN. [20 points]

    1) Find suitable values of eps and MinPts.

    2) Explain how you determine suitable values of eps and MinPts.

1.4 Which method is the best clustering method for the census-income data set? [10 points]