CPS 844 Lab 8: Cluster Analysis

Cluster analysis is the task of partitioning a set of objects into groups of similar instances. This lab has two parts:

Part 1: k-means clustering (50 points)

The k-means clustering method consists of the iterative assignment and update steps:

1. Form *k* clusters by assigning each instance to its nearest centroid.
2. Recompute the centroid of each cluster.

We use a simple and small dataset of movie ratings by 6 users. Your goal will be to identify people with similar movie preferences using k-means clustering.

Part 2: Hierarchical Clustering methods (50 points)

Using the data from the file 'vertebrate.csv', we apply the following hierarchical clustering methods:

1. Single link (MIN)
2. Complete link (MAX)
3. Group average

Write one Python script per part, that performs the tasks described in lab8_part1.py and lab8_part2.py. Submit the .py files on D2L. Please note that if you submit your file in some other format besides .py or (.txt should you meet an issue), then your mark will at most be 60%.