MG 265 Data Mining

Assignment 2 Due date: Thursday, October 27 2022

1. The objective of this assignment is to compare the performance of various classification algorithms. Use the attached Boston house price data set and predict if a given suburb has a crime rate above or below the median. Use LDA, QDA, decision tree, random forest, support vector machine and naive Bayes based classification models to predict if the crime rate of a suburb is above or below the median. Also, build a linear regression model to predict the crime rate and classify it as above or below the median based on the prediction to compare it with the classification models performance. Use a five fold cross-validation to estimate the test error.

2. Apply hierarchical clustering to the attached gene samples data and see if the genes separate into two groups. Vary the distance measure (eucledian/correlation/manhattan) used and linkage procedure (min/max/average) and document its impact on the clusters formed.

-x-x-x-x-