Data Mining I
Homework 1
30 points

Directions: Submit all source codes with write up.

1) (10 points) Consider the Auto dataset in the ISLR package. Suppose that you are getting this data in order to build a predictive model for mpg (miles per gallon). Using the full dataset, investigate the data using exploratory data analysis such as scatterplots, and other tools we have discussed. Pre-process this data and justify your choices in your write up. Submit the cleaned dataset as an *.RData file.

2) (10 points) Perform a multiple regression on the dataset you pre-processed in question one. The response variable is mpg. Use the lm() function in R.
   a) Which predictors appear to have a significant relationship to the response.
   b) What does the coefficient variable for "year" suggest?
   c) Use the * and : symbols to fit models with interactions. Are there any interactions that are significant?

3) (10 points) ISLR textbook exercise 2.10 modified: This exercise concerns the boston housing data in the MASS library (>library(MASS) >data(Boston)).
   a) Make pairwise scatterplots of the predictors, and describe your findings.
   b) Are any of the predictors associated with per capita crime rate?
   c) Do any of the suburbs of Boston appear to have particularly high crime rates? Tax rates? Pupil-teacher ratios? Comment on the range of each predictor.
   d) In this data set, how many of the suburbs average more than seen rooms per dwelling? More than eight rooms per dwelling? Comment on the suburbs that average more than eight rooms per dwelling.