**Computational Statistics**

**Lab 4 Assignment**

**Subash Kharel**

1. Answer to question number 11
   1. Answer is in the R file attached
   2. Answer: Looking at the scatter plot (Figure 1) and boxplots (Figure 2- 5), we can analyze the relation between mpg01 and other predictors in following ways:
      1. Cylinders: Not significant relation since the graph does not show any boundary values of cylinder that can predict 0 or 1 for cylinder. They are equally distributed on each side.
      2. Displacement: Though the graph seems better than the cylinders, value 1’s are completely overlapped with 0’s. This cannot be a good predictor.
      3. Horsepower: This can be considered as a good predictor since, we can draw somewhat vertical sigmoid line to separate points.
      4. Weight: Same case as horsepower.
      5. Acceleration: Same case as weight.
      6. Year: The points are equally distributed on each side, so not significant as a predictor.
      7. Origin: Somehow sparse but same case as Year. Cannot be a good predictor.
   3. Answer is in the R file attached. The test and training data are separated taking a value of year as a boundary.
   4. The test error for the model obtained in Linear Discriminant Analysis is 15.52%.
   5. The test error for the model obtained in Quadratic Discriminant Analysis is 12.94%.
   6. The test error for the model obtained in Logistic Regression Analysis is 2.04%.
   7. The test error for the model obtained in KNN Classification is 2.04%.  
      The error values for various values of K is given below:

K = 1, error = 25.86%   
K = 2, error = 32.76%   
K = 5, error = 29.31%   
K = 10, error = 22.41%   
K = 15, error = 17.24%  
K = 20, error = 15.51%

K = 25, error = 15.51%  
K = 50, error = 17.24%   
K = 50, error = 17.24%

K = 18 - 42, error = 15.51%

From the observation, we can say that the error value is least for K = 18 to 42, which is 15.51%