Math 189: Homework 5

Auto Mileage

In this assignment you will develop a model to predict whether a given car gets high or low gas mileage based on the Auto data set. This data can be found in the ISLR package.

```
library(ISLR)
data(Auto)
```

The dataset has 392 observations on automobiles. For each automobile, we record the miles per gallon (MPG), among other variables such as horsepower and weight.

Tasks

Analyze the dataset according to the following steps:

1. Create a binary variable, mgp01, that contains 1 if mpg contains a value above its median, and a 0 if mpg contains a value below its median. You can compute the median using the median() function.
2. Explore the data graphically in order to investigate the association between mgp01 and the other features. Which of the other features seem most likely to be useful in predicting mgp01? Scatterplots and boxplots may be useful tools to answer this question. Describe your findings.
3. Split the data into a training set of size 300 and a test set of size 92.
4. Perform LDA on the training data in order to predict mgp01 using the variables that seemed most associated with mgp01.
5. Classify the test data. Discuss the results in terms of the proportion of correctly classified records.

Remarks

Your R Markdown Notebook report should have a introduction, body, conclusion (and optional appendix). Importantly, your code should run!