Math 5366 Homework 37

1. Split the Auto MPG data set from the UCI Machine Learning Repository into 70% training data and 30% test data. Build regression models for predicting `mpg` based on the other variables in the data set, using the follow methods:

   (a) Ordinary Least Squares
   (b) LOESS
   (c) Weighted k-Nearest Neighbors
   (d) Support Vector Regression
   (e) Neural Networks
   (f) Random Forests

   To evaluate the performance of each technique, produce the following:

   - A plot of y vs. $\hat{y}$, where $y =$ mpg, and
   - the correlation coefficient between y and $\hat{y}$.

   Complications may result when building these models. For example, LOESS cannot be applied to a model with categorical predictors, and it can't be used with more than four predictors, so you may want to restrict attention to displacement, horsepower, weight, and acceleration for the LOESS model.

   There are also six missing values for horsepower, and given the size of the data set, it would be reasonable to discard them. Alternatively, another regression technique could be used to estimate these missing values.

   Finally, you may run into the error message, "predictors must all be numeric", even after removing all categorical variables. For a given data frame `mydata`, the command

   ```
   mydata=as.data.frame(sapply(mydata,as.numeric))
   ```

   will convert it to a numeric data frame.