BMIG 6201
Homework # 2: Regression and SVMs

Horacio Gómez-Acevedo, PhD.

February 16, 2021

1. Using the *Auto* dataset, provide a solution to the following questions
   (a) Create a binary variable, *mpg*01, that contains a 1 if *mpg* contains a value above its median, and a 0 if *mpg* containes a value below its median.
   (b) Explore the data graphically in order to investigate the associate between *mpg*01 and the other features using scatterplots and boxplots. Describe your findings.
   (c) Split the data into a training set and a test set.
   (d) Perform logistic regression on the training data in order to predict *mpg*01 using the variables that seemed most associated with *mpg*01. Calculate the test error of the model obtained
   (e) Fit a support vector classifier to the data with various values of *cost*, in order to predict whether a car gets high or low gas milage. Report the cross-validation error associated with different values of this parameter.
   (f) Repeat the previous item but this time using SVMs with radial and polynomial basis kernels with different values of *gamma* and *degree* and *costs*.
2. Using the *College* data set **you will predict the number of applications received using the other variables**, provide a solution for the following questions
   (a) Split the data set into a training set and a test set.
   (b) Fit a linear model using least squares on the training set and report the test error obtained.
   (c) Fit a ridge regression model on the training set, with λ chosen by cross-validation. Report the test error obtained.
   (d) Fit a lasso model on the trainint set, with λ chose by corss-validation. Report the test error obtained along with the number of non-zero coefficient estimates.