**CS 5751 – Spring 2018 – Homework 5**
**Assigned: 03/13/2018**
**Due: 03/20/2018**
**Total points: 100 pts.**
**Submit a soft copy to canvas. Remember to write your name at the top of each file you submit.**

**Objectives:** The objectives of this homework are the following:

- Learn how to use R and Scikit-learn for linear regression.

**Notes:**

- This homework is to be done individually. You may discuss with your classmates, but the work that you write must be your own.

**Activity 1: (50 pts.) (Least-squares linear regression with Scikit-learn) Using Python and scikit-learn, do the following:**

a) Download the auto-mpg dataset from the UCI website.
b) (10 pts.) Do any pre-processing that may be needed.
c) (15 pts.) Split the dataset randomly into two portions: a training set and a test set. The training set will contain 80% of the rows of the dataset, while the test set will contain the remaining 20% of the rows.
d) (10 pts.) Write down an equation for your linear model. Choose the same variables that you chose for the last homework.
e) (15 pts.) Using scikit-learn's LinearRegression, train your linear regression model on your training set. Then, evaluate the performance of your linear model using the test set. Use the sum of the squared errors as a performance measure. Report the sum of squared errors that you obtained. Compare against the results you obtained in the previous homework.

For this activity, write a Jupyter notebook named yourLastName_hw5_q1.ipynb that implements 1a through 1e.

**Activity 2: (50 pts.) (Least-squares linear regression with R)** Repeat Activity 1, but instead of using scikit-learn and Python, use R. For this, either the function 'lm' or the package 'caret' might be useful.

For this activity, write a Jupyter notebook named yourLastName_hw5_q2.ipynb.