CS 5751 – Spring 2018 – Homework 4
Assigned: 02/20/2018
Due: 03/01/2018 (1 Week and 2 days)
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 implement 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.
- If you choose Python, then you need to make sure that you have the following packages: numpy, jupyter. See the slides posted on canvas to learn how to install Python and R. If you are using Linux or Mac, you can install these packages by typing the following at the command prompt: pip install numpy, jupyter.

Activity 1: (50 pts.) (Least-squares linear regression) Using either Python or R, do the following:

a) Download the auto-mpg dataset from the UCI website.
b) (20 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.
c) (40 pts.) Implement the least-squares linear regression algorithm using the normal equations. You cannot use either scikit-learn's regression algorithm, nor R's *lm* function. You need to do this on your own.
d) (10 pts.) Write down an equation for your linear model. You are free to choose the variables you like.
e) (30 pts.) 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.

For this activity you will write a Jupyter notebook named yourLastName_hw4_q1.ipynb that implements 1a through 1e.