**Programming Exercise I: Linear Regression**

**Due date: 3/25 Monday 11:59 PM**

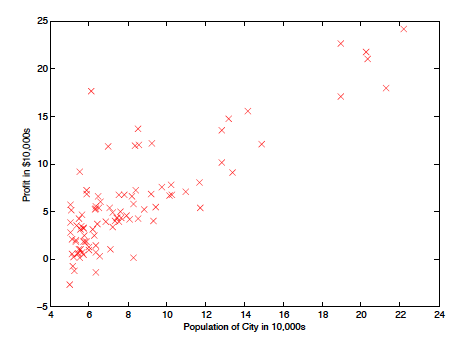
1. Linear Regression with One Variable [30 pts]

In this part of this exercise, you will implement linear regression with one variable to predict profits for a food truck. Suppose you are the CEO of a restaurant franchise and are considering different cities for opening a new outlet. The chain already has trucks in various cities and you have data for profits and populations from the cities. You would like to use this data to help you select which city to expand to the next.

The file ex1data1.txt contains the dataset for our linear regression problem. The first column is the population of a city and the second column is the profit of a food truck in that city. A negative value for profit indicates a loss.

* 1. Plotting the data with matplotlib [10 pts]

Before starting on any task, it is often useful to understand the data by visualizing it. For this dataset, you can use a scatter plot to visualize the data, since it has only two properties to plot (profit & population). A sample of this plot is provided as follows:



* 1. Train a single variable linear regression model in scikit-learn [20 pts]

Split the dataset in the file ex1data1.txt using a 0.75 vs 0.25 ratio. Train the model on the 0.75 split and test the trained linear regressor on the 0.25 split. Report the MSE losses on both train and test splits.

1. Linear Regression with Multiple Variables. [40 pts]

In this part, you will implement linear regression with multiple variables to predict the prices of houses. Suppose you are selling your house and you want to know what a good market price would be. One way to do this is to first collect information on recent houses sold and make a model of housing prices.

The file ex1data2.txt contains a training set of housing prices in Portland, Oregon. The first column is the size of the house (in square feet), the second column is the number of bedrooms, and the third column is the price of the house.

* 1. Feature normalization [20 pts]

Your task is to normalize each of the two features (i.e., size of the house in col 1, number of bedrooms in col 2). In particular,

* Subtract the mean value of each feature from the dataset.
* After subtracting the mean, additionally scale (i.e., divide) the feature values by their respective standard deviations.

* 1. Use the normalized features to train a multivariable linear regression model to predict the house price using a 0.7 vs 0.3 split. Report the MSE losses on both train and test splits. [20 pts]

**Submission**: please submit your source code or notebook files on Canvas following the assignment 1 link.