Assignment 3 Summer Analytics 2019

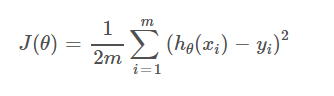
Linear Regression with One Variable

*Here we 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.*

The file data.txt contains the dataset for our linear regression exercise. 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.

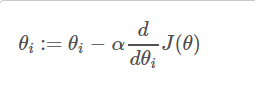
TASKS:

1. Load the dataset and perform EDA. 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 and population).
2. Initialize the parameters such as theta, alpha (learning rate), number of iterations with sensible values as taught in lectures.
3. Define the cost function:

As

Compute\_cost(X, y, theta), this function will compute and return J.

1. Optimize your parameters using Gradient Descent so define the corresponding function :



def gradientDescent(X, y, theta, alpha, iterations):  
 #write your code  
 return theta

1. Calculate the Cost function with optimized value of parameters and compare it with the value of Cost function initially before optimization.
2. Plot the scatter plot of the data along with best line fit i.e. your optimized hypothesis function h(theta).
3. Predict the values of prices for population of 35000 and 70000.
4. Using sklearn library perform Linear regression on the given dataset and predict prices for the same population as above and compare your results.