ASSIGNMENT 4

Name: Mehul Jain

Roll No: 2018IMT-051

Course: Machine Learning Lab

Course Code: ITIT-4107

Aim: 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 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. Use a scatter plot to visualize the data, since it has only two properties to plot (profit and population).

2. Consider a simple linear model with two parameters and one input variable and mean square error cost function to implement the gradient descent algorithm to find the intercepts. Assume a suitable terminating condition.

3. Plot the model alongside the scatter plot to show the fit model.

4. Perform steps 1, 2, 3 in batch mode for varying values of alpha, learning rate and plot the results.

5. For each of the experiments performed above in steps 1,2,3,4 with varying learning rates visualize the cost function as a contour plot as well as plot the values of parameters to visualize the stepwise traversing of the parameters on this contour plot.

Procedure:

1. We define the linear relationship between the two variables as follows:

h(x) = (theta_0) * x + theta_1

We have to determine the value of theta_0 and theta_1, such that the line corresponding to those values is the best fitting line or gives the minimum error. Formally, we have to minimize the below cost function:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

2. Then we implement gradient descent algorithm are as follows:

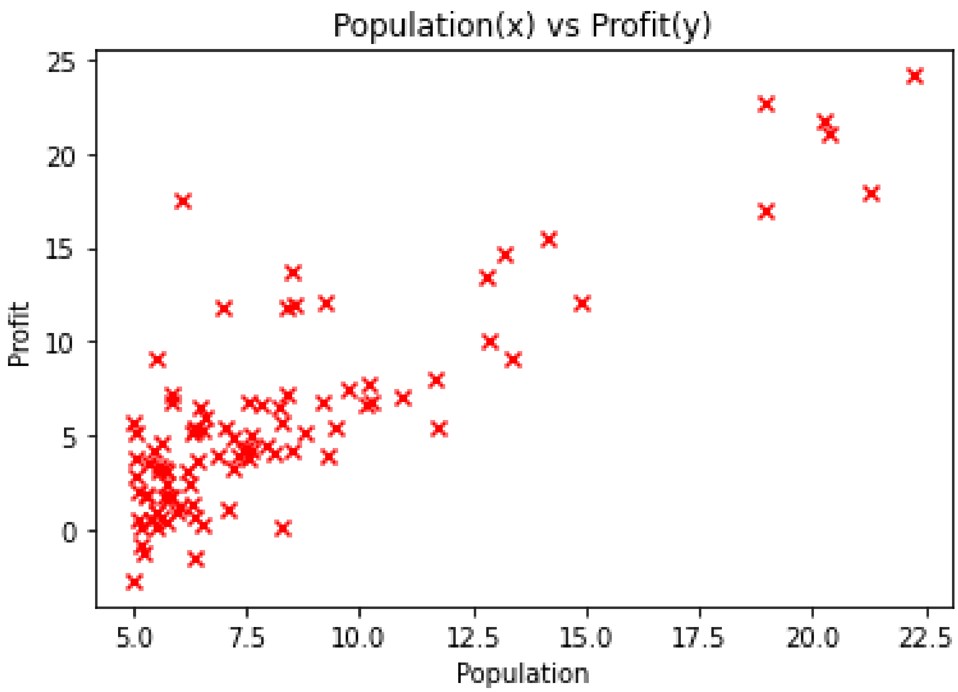
a. Start with theta_0 = 0 and theta_1 = 0. Vary the value of alpha i.e. learning rate (L) to obtain the least MSE.

b. Next, we do several iterations where each iteration updates the values of theta_0 and theta_1 with the following equation:

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right) x_j^{(i)} \qquad \text{simultaneously update } \theta_j \text{ for all } j$$

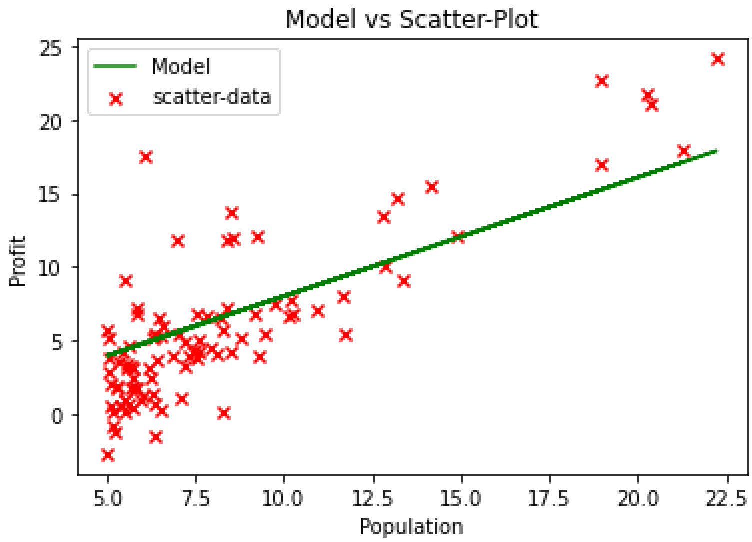
Output:

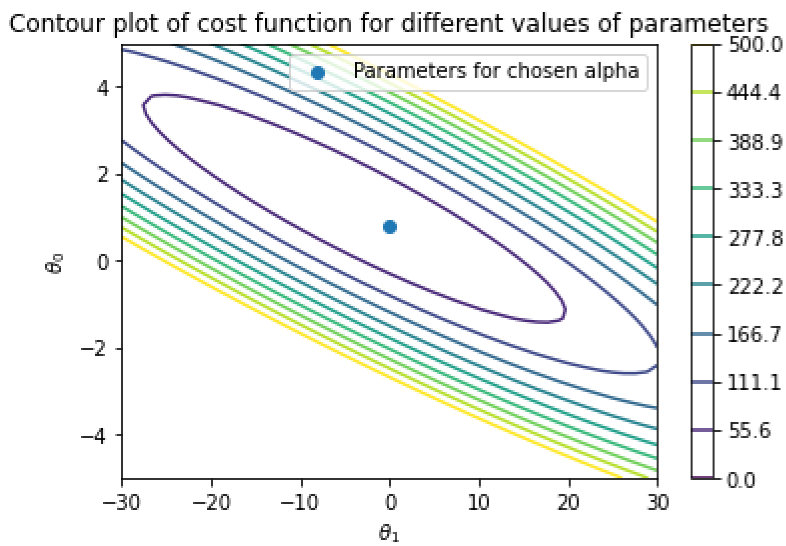
The given input data is visualized as follows-



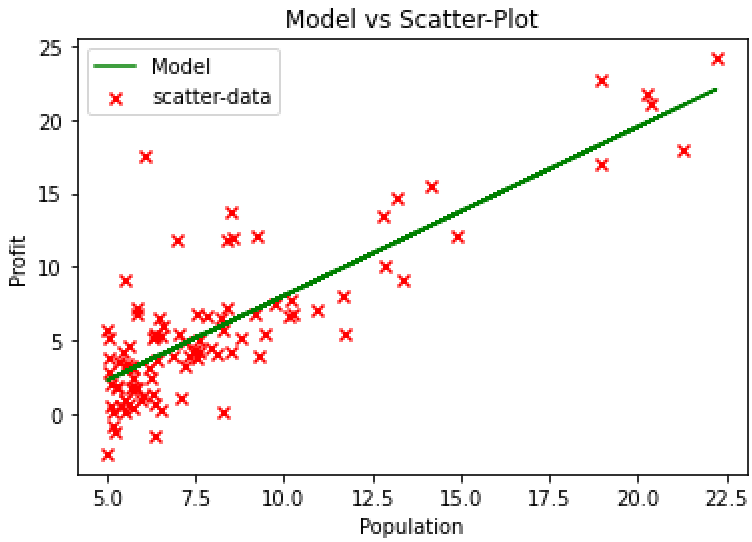
We implement the gradient descent algorithm on varying values of *alpha* i.e. learning rate to find the best fit model for the given data.

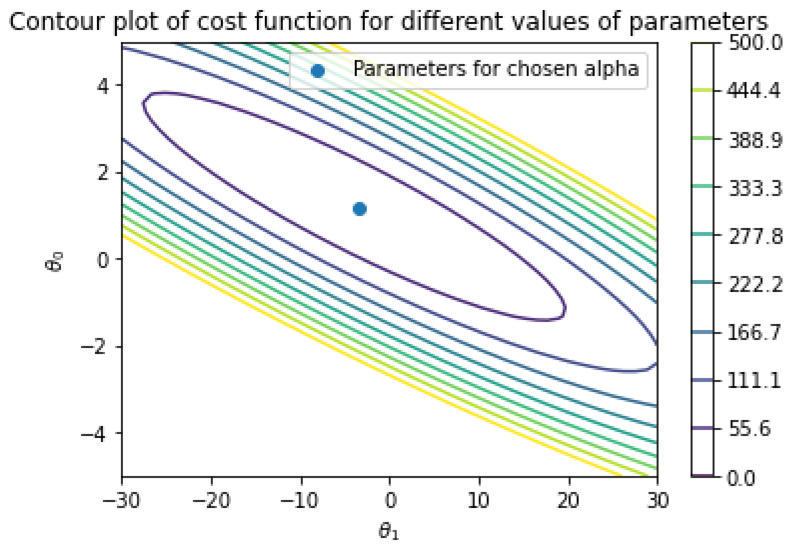
1. Learning rate: **0.0001**
   Hypothesis function: h(x) = 0.8077855783900526x + -0.060973353535627335
   Cost function value = 5.815774148036794



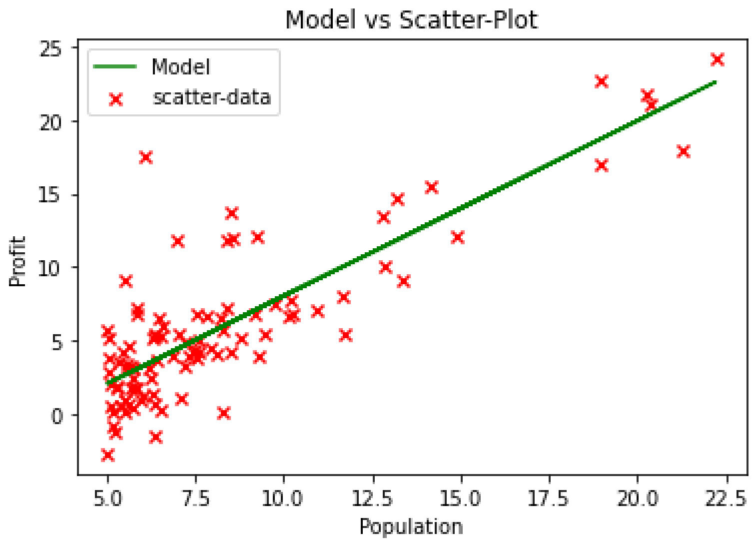


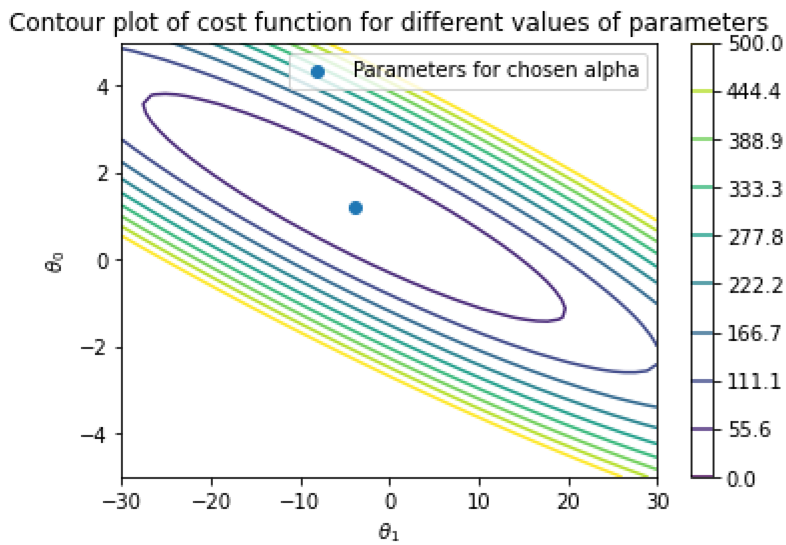
2. Learning rate: **0.0061**
   Hypothesis function: h(x) = 1.1487486962170874x + -3.454962894241694
   Cost function value = 4.494662200232537





3. Learning rate: **0.0181**
   Hypothesis function: h(x) = 1.1924525693532309x + -3.8899967857547533
   Cost function value = 4.476974421764978





Inference:

We varied the values of alpha i.e. learning rate from 0.0001 to 0.018. After analyzing the results, it was evident that the optimal learning rate is between 0.012~0.018. For alpha = 0.018, the linear model is represented by:

$$h(x) = 1.19x + -3.88$$

The value of the cost function for alpha = 0.018 is minimal among all learning rates, equal to 4.476.

Code:

https://github.com/MehulJain-831/ITIT-4103-2021/blob/main/Assignment%204/2018IMT_051_Assignment_4.ipynb