**Machine Learning**

**Fall 2022**

**HW4**

**Due: Sept. 28 (Wed), 11:59pm, via Blackboard**

**Problem 1:** In this exercise, you are required to write a Matlab code of Linear Regression (LR) algorithm, following the example presented during the lecture. A data file with part shown in Table 1 is provided.

**Table 1**

|  |  |
| --- | --- |
| **x (variable)** | **y (response)** |
| 6.6994 | 18.1257 |
| 5.6825 | 8.8036 |

**.**

**.**

**.**

|  |  |
| --- | --- |
| 13.7473 | 9.5681 |
| 6.4135 | 0.4449 |

Based on file ‘**dataset.mat**’ (**Table 1**), you need to implement the LR algorithm to obtain the fitting line for the points in the dataset. Please follow the following requirements:

1. Load the dataset and plot the dataset in **figure 1** in matlab, Please name x axis ‘ x ‘ and y axis ‘ y ‘.



1. Write a self-defined function ‘gradientDescent( )‘ in matlab, which includes the gradient descent algorithm. You need to call the function ‘gradientDescent( )’ in you main code. Please name the self-defined function file as ‘**gradientDescent.m**’ and call this function in your main code file.
2. Write a self-defined function ‘computeCost( )‘ in matlab, which includes the computing of the cost function. You need to call the function ‘computeCost ( )’ in you main code. Please name the self-defined function file ‘**computeCost.m**’ and call this function in your main code file.
3. Set the total number of iterations to 2500 and the learning rate to 0.01.

Initialize the parameters and to 0.

Plot the dataset and the linear fitting line () in **figure 2.**

1. Plot the cost versus the iteration in **figure 3**

**Please Note**: When you submit the homework, you should make sure that the homework documents include below items:

**Three matlab files**:

maincode.m / maincode.mlx

gradientDescent.m

computeCost.m

**A report** that briefly addresses 1) a flow chart of your codes with brief text description, 2) all the results of requirements 1-4 should be clearly presented in your report.

If you make changes to the dataset file, such as adding the column name, please upload your dataset to BB.

**Hints**:

1. Follow the frame of the codes in the LR lecture.
2. You do not need to do the normalization or standardization since the data in this homework are relatively simple.
3. You do not need to split the dataset into training and test sets