**Machine Learning**

**Fall 2022**

**HW5**

**Due: Oct. 6 (Thur), 11:59pm, via Blackboard**

**Problem 1:** In this exercise. you are required to write a Matlab code of a logistic regression model to predict whether a student will be admitted to a university. A data file with part shown in Table 1 is provided.

**Table 1 (0: not admitted, 1: admitted)**

|  |  |  |
| --- | --- | --- |
| **Exam 1** | **Exam 2** | **Status** |
| 36.8525 | 81.8471 | 0 |
| 29.229 | 47.3023 | 0 |

**.**

**.**

**.**

|  |  |  |
| --- | --- | --- |
| 55.9044 | 64.408 | 1 |
| 78.9105 | 88.0795 | 1 |

Based on file ‘**dataset.mat**’ (**Table 1**), you need to implement the logistic regression algorithm (generate decision boundary for the points in the dataset and predict new test samples). Please include following **requirements** in your code:

1. Load data, standardize the original dataset (do not standardize the status). In Matlab, plot the preprocessed dataset in **Figure 1**, name x axis as ‘Exam 1‘ and y axis as ‘Exam 2‘. Use marker ‘●’ to represent ‘ not admitted’ and marker ‘x’ to represent ‘ admitted’. Corresponding legend must be included in **Figure 1**.
2. Write a Matlab function ‘SigMoid‘ which computes . Please name this self-defined Matlab function file as ‘SigMoid**.m**’ and call this function in your main code file.
3. Write a Matlab function ‘costFunction‘. In this function, you are required to compute the cost value and the gradient, which means the output of this function is the cost and gradient. Please name the self-defined Matlab function file as ‘**costFunction.m**’ and call this function in your main code file. (please update the theta in the main code by using the gradient value from this costFunction() ).
4. Set iterations to 5000 and learning rate to 0.01.

Initialize , and to 1.

Train and update the , and

In **Figure 2**, plot the dataset points and decision boundary line () . Compute the accuracy on dataset (the probability threshold: 0.5, which means ‘>=0.5’, status=1; ‘<0.5’, status=0).

1. Plot the cost J over the iterations in **Figure 3**.
2. Predict the test sample [1 45 85] and show the probability.
3. With the trained model (theta), fix and plot a 3D figure, locating cost on z-axis, on x-axis, and on y-axis.

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

**Three matlab files:**

maincode.m / maincode.mlx

costFunction.m

SigMoid**.**m

**A report** that briefly addresses 1) a flow chart of your codes with brief text description, 2) all the results of requirements 1-6 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 framework of the codes in the lecture.
2. You need to standardize the data (please refer to the standardization in linear regression slides), otherwise the computational cost will very high. When you predict a test point, you should use the mean values and standard deviation values from the dataset to standardize the test point. Then you can compute the admission probability of the test point with the self-defined SigMoid( ) function.
3. Make sure your accuracy on the training set>=80%.
4. With the final , and , you can easily draw the decision boundary line: . You may not need to use the plotDecisionBoundary() in the lecture slides.

