PAF-Karachi Institute of Economics and Technology
Course: Machine Learning (S2020)

Assignment – 1

Note: 1. Individual student submission
2. If your assignment fall in plagiarism, will not be graded
3. Possibility to conduct the viva of your code

Max marks : 33 **Due Date: 3-3-2020**

[Question – 3]

Consider the given data set and answer the following questions [5 + 5 + 2 + 3]

x	6.11	5.52	8.51	7.00
y	17.59	9.13	13.66	11.85

i. Calculate the two new values of θ's using gradient descent formula of Linear Regression. The initial θ's = [0, 0]. $h_\theta(x) = \theta_0 + \theta_1 X$

ii. Consider the boston_train.csv data set where "medv" is the target attribute and write the code for the following problems:

- **(a)** Write the python code for the cost function. Provide different θ's to your code and draw the cost function graph against provided θ's.
- **(b)** Write the code for linear regression. Is it possible to draw the regression line?
- **(c)** Calculate the MSE for the different train-test partition size (atleast 20). What is your observation in this regard?

[Question – 2] Binary logistic regression

[5 + 4 + 4]

Consider the given data of student exam results and our goal is to predict whether a student will pass or fail based on the number of hours slept and hours spent studying. We have two features (hours slept, hours studied) and two classes: passed (1) and failed (0).

Studied	Slept	Passed
4.85	9.63	1
8.62	3.23	0
5.43	8.23	1
9.21	6.34	0

(i) Calculate the one new values of θ's using gradient descent formula of Logistic Regression. The initial θ's = [0, 0, 0]. $h_\theta(x) = \theta_0 + \theta_1 X_1 + \theta_2 X_2$ for sigmoid function.

(ii) Consider the insurance_data.csv data set, write the code for the following problems

- **(a)** Write python code for scatter plot, and bar graph also attach both outputs
- **(b)** Write code for Logistic regression. Use different partitions size of train-test (atleast 20) and draw a graph between test accuracy vs partition size.