Assignments on Linear Regression #3 (Ref Video Lectures 1-12)

(Timely submission of assignments is essential. Copying/plagiarised submission from others will fetch fail (F) grade on this subject)

1. Annual Revenue data for a company is given as,

Y	2004	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Rev. In billion Rupees	61.2	58.3	67.1	69.2	68.9	83.5	89.1	80	92.3	93	97

a) Draw a least square line fitting the data.
b) What is the expected revenue in 2019
c) Analyze expected error in predictions.   10

2. The following table shows the final semester marks obtained by 10 students selected at random.

M L	75	80	93	65	87	71	98	68	84	77
HUR	82	78	86	72	91	80	95	72	89	74

Find least square line fitting the above data using
a) X as independent variables (regression of Y on X)
b) Y as independent variable (regression of X on Y)
c) If a student receives a mark 96 in ML, what is her/his expected marks in HUR.
d) If a student receives 95 in HUR. What is her/his expected marks in ML.
e) After plotting a) and b) what conclusions can you draw?   10

3. Experimental results of pressure (P) for a given mass of gas corresponding to various values of volume (V) is given as:

V	54.3	61.8	72.4	88.7	118.6	194
P	61.2	49.5	37.5	28.4	19.2	10.1

Assume PV^n = const =c
a) Find the parameters n and c
b) Write the equation connecting P and V.
c) Estimate the value of P when V=100

                                                                10

4. Find the least square parabola which fits the data
$Y= W0 + W1X+ W2 X^2$

X	0	1	2	3	4	5	6
Y	2.4	2.1	3.2	5.6	9.3	14.6	21.9

Submission Deadline: **CLOSED**   **10**

5. Download the COVID -19 data of India for the month of May, 2020 and design a predictor for the number of deaths on a particular day. Hence, predict the number of deaths on April 20, 2020 and June 10th , 2020. Verify your prediction with the actual number of deaths and hence calculate the accuracy of prediction.

Submission Deadline: **29 Th August, 2020.**   **20**

[Hints: collect data from my website or from the web for the month of May, 2020. Use 22 days data for training and rest 09 days data for testing. Use both straight line fitting and quadratic curve fitting for predictor design and compare their performance in predicting accuracy]

6. Download the housing price data set of Windsor City of Canada (provided on my website link). Design a housing price predictor taking only floor area (plot size), number of bedrooms, and number of bathrooms into considerations. Out of total 546 data , you may take 70% for designing the predictor and 30% for validating the design. The predictor design should be done using the following methods:
:
a) Normal equations with and without regularization and compare their performances in terms of % error in prediction. (only allowed to use NumPy library of Python.no other functions/libraries are allowed)
 (Ref Lecture-8 and 12)
Submission deadline : August 31st , 2020.   20

b) Design Predictor using Batch Gradient Descent Algorithm, Stochastic Gradient Algorithm and mini batch Gradient Descent algorithms (determining minibatch size is your choice- here it could be 10, 20, 30 etc.) with and without **feature scaling** and compare their performances in terms of % error in prediction.(only allowed to use NumPy library of Python, no other functions/libraries are allowed)
(**Ref Lecture- 10 and 11)**
Submission deadline : Sept 4th , 2020.   30

c) Design Predictor using Batch Gradient Descent Algorithm, Stochastic Gradient Algorithm and mini batch Gradient Descent algorithms (determining minibatch size is your choice- here it could be 10, 20, 30 etc.) with and without **regularization** and compare their performances in terms of % error in prediction.(only allowed to use the NumPy library of Python, no other functions/libraries are allowed) **(ref Lecture-10, 11 and 12)**

Submission deadline : Sept 6th , 2020.   30

d) Implement the LWR algorithm on the Housing Price data set with different tau values. Find out the tau value which will provide the best fit predictor and hence compare its results with a) , b) and c) above.   **50**

Submission deadline : Sept 7th , 2020.   (this assignment will earn additional bonus points of 50)

All assignments given before C-1 evaluation, will be counted towards C-1 marks.