**Deduce the Linear Regression model in the form of Normal Equation for two parameter predictors and then proceed to solve the following assignments.**

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

| Y | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|------|------|------|------|------|------|------|------|------|------|------|
| Rev. In billion Rupees | 200.5 | 198.5 | 187.4 | 189.2 | 188.9 | 183.5 | 189.1 | 284 | 292.3 | 396 | 397 |

a) Draw a least square line fitting the data.
b) What is the expected revenue in 2023
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 | 89 | 95 | 96 | 69 | 87 | 70 | 98 | 69 | 85 | 87 |
|-----|----|----|----|----|----|----|----|----|----|----|
| HRI | 81 | 86 | 98 | 72 | 93 | 82 | 96 | 75 | 89 | 84 |

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 | 55.3 | 62.5 | 72.4 | 89.7 | 118.6 | 194 |
|---|------|------|------|------|-------|-----|
| P | 63.2 | 50.5 | 37.5 | 38.4 | 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
     $Y = W_0 + W_1 X + W_2 X^2$

| 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:  **Ref Lecture videos 1-8** :      **22-10-2022**                    **10**

5. Download the COVID -19 data of India for the month of May, 2022 and design a predictor for the number of infections on a particular day. Hence, predict the number of infections on  April 20, 2022 and August 10th , 2022. Verify your prediction with the actual number of deaths and hence calculate the accuracy of prediction.

Submission Deadline: **24th October, 2022.**                          **20**

**[Hints: Collect data from my website or from the web for the month of May, 2022. 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 :    5th November, 2022**

                                                                                    **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 :    5th November, 2022**

                                                                                    **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 :    8th November, 2022**

                                                                                    **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**