School of Computer Science Engineering and Technology

Course- BTech
Course Code- 301
Year- 2022
Date- 20-01-2022

Type- Core
Course Name-AIML
Semester- Even
Batch- 4th Sem (SPL)

2 - Lab Assignment # No. (2.2)

Objective: To gain a deeper understanding of Simple Linear regression by implementing it from scratch.

1. Download the Students performance dataset available on
UCI repository (https://archive.ics.uci.edu/ml/datasets/student+performance) which consists of a total of 32 attributes. (5)

2. Read the dataset (use **read_csv()** from **pandas**) into some variable. Take the last two columns (G2 and G3) into XY. (5)

3. Print the different statistical values of data contained in XY using **describe()** function from **pandas**. (5)

4. Divide XY into X consisting of G2 and Y consisting of G3. Print the shape of both. (10)

5. Add a column at position 0 with all values=1. (5)

6. Print some of the rows from XY. (5)

7. Complete the following functions given in the provided Ipython Notebook to implement a Linear Regression model between X and Y (Y = mX + C). (40)

- Write code to predict G3 for a given set of weights and input G2.
- Write a function to calculate the loss (mean squared error) for given set of weights, input G2 and actual output G3
- Write a function to calculate the gradient for given set of weights, input G2 and actual output G3
- Write a function to perform gradient decent for given set of input G2 and actual output G3

8. Play with different values of max_iterations and the learning rate. (15)

Additional fun (will not be evaluated)

9. Split the data in X_train, X_test, Y_train, Y_test (sklearn.model_selection.train_test_split function)

10. Calculate mean squared error on both X_train and X_test.

11. Generalize the code for multivariate(multiple) linear regression.