Data Science Lab (CS 356)

Assignment 3.1

Date: 27.01.2022

Instructions to submit the lab assignment

a. *Add proper comment lines for each important step of the code.*
b. *All the codes should be in same file.*
c. *Name each file as rollnumber_assignmentnumber.pdf.*
d. *Upload the program file in google classroom.*

---

1. Write a python program to implement the Simple Linear Regression(SLR) from scratch. Consider the dataset from the given link.
   Data Link: https://www.kaggle.com/karthickveerakumar/salary-data-simple-linear-regression
   Or
   Data Link: https://www.kaggle.com/prasadperera/the-boston-housing-dataset

   Find the relationship between the independent variable and dependent. Mention the variables considered for the respective dataset. Implement SLR for the given data as per the steps given below.
   [*Any one dataset can be considered as per your choice. Mention the dataset used before implementation*]
   a. Import the dataset
   b. Assign the values to independent variable and dependent variable
   c. Print the first few rows
   d. Separate the dataset into train and test data as 80% and 20% respectively.
   e. Plot the dataset (train dataset)(independent variable vs dependent variable)
   f. Calculate the regression line(train dataset):
      i. Compute the slope
      ii. Compute the intercept
      iii. Compute and Plot regression line with data points
   g. Predict the values using test data.
   h. Calculate the error / accuracy of the model using root mean squared error

2. Write a python program to implement the Multiple Linear Regression from scratch as per the steps given below and consider the dataset from the given link.
   Data Link: https://www.kaggle.com/prasadperera/the-boston-housing-dataset

   a. Import the required libraries and the data
   b. Scale the dataset [Standardization of the data]
   c. Separate the dataset into train and test data as 80% and 20% respectively.
   d. Implement the logic of the algorithm using Gradient Descent Function.
   e. Train the model and plot the data
   f. Predict the values using test data.
   g. Calculate the error / accuracy of the model using root mean squared error