Heritage Institute of Technology
Department of Computer Sc. & Engg. (Data Science)
B. Tech 3rd Year 6th Semester Group 1/2
Paper: Machine Learning Lab (DSC 3252)
Session 2025-2026
Least Squares and Regression

Problem Statement:

Objective:

- Introduction to the basic concept of the least squares method and its application in fitting a linear model to a set of data points.
- Importing data from .csv file and implementing the least squares method and its application in fitting a linear model to a set of data points.

Problem Description:

You are provided with a dataset representing the relationship between an independent variable 'x' and a dependent variable 'y'. The objective is to use the least squares method to find the parameters of a linear model that best fits the given data.

Dataset:

x	1	2	3	4	5
y	3	7	9	13	15

Assignment:

Part – 01:

1. Based on the form $f(x) = mx + b$, formulate the linear model for the given dataset.
2. Calculate the Slope (m) and Y-Intercept (b):Apply the least squares method to calculate the Slope (m) and Y-Intercept (b) using the formulas:

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{(\sum y) - m(\sum x)}{n}$$

3. Plot the Linear Regression Model $f(x) = mx + b$ using *matplotlib* for the initial values of x and y as given in the dataset.
4. Evaluate the Model:
   a. Calculate the predicted values of respective $f_{new}(x)$ based on the linear model and compare them with the actual values.
   b. Calculate the sum of squared differences between the actual and predicted values.
5. Generate $f_{new}(x)$ using $f_{new}(x) = mx + b$. Plot the Linear Regression Model $f_{new}(x) = mx + b$ using *matplotlib* for the values of x and $f_{new}(x)$.
6. Take an unknown value of $x = 4.5$ and find the value of y. Plot it in the graph using a different colour.

Part – 02:

1. Create a .csv file having 100 random datapoints between the values 40 and 50. Read the .csv file and then plot the graphs as mentioned in point 3 and point 5 in Part – 01.
2. Take an unknown value of $x = 43.5$ and find the value of respective $f_{new}(x)$. Plot it in the graph using a different colour.