## Experiment 5

**Objective:** Implement simple linear regression to analyze relationships between variables and make predictions.

**Explanation:**

**Linear Regression** is a statistical method to model the relationship between a dependent variable (target) and one or more independent variables (features). It is commonly used for prediction and finding trends.

$$y = mx + c$$

Where:

- y = Dependent Variable (Target)
- x = Independent Variable (Feature)
- m = Slope of the line
- c = Intercept (value of y when x=0)

Formula for Slope:

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

Formula for Intercept:

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

**Exercise 1:**

A company wants to predict **monthly sales** based on its **advertising budget** using simple linear regression.

| Advertising Budget (in $1000) | Monthly Sales (in $1000) |
|---|---|
| 1 | 4 |
| 2 | 5 |
| 3 | 7 |
| 4 | 8 |
| 5 | 11 |

i. Calculate the Line of Best Fit (on paper)

ii. Predict sales based on an advertising budget of $6000 (on paper)

iii. Perform a linear regression to predict sales based on the advertising budget using R.

**Exercise 2:**

**Predict House Prices Using Linear Regression**

In this exercise, build a linear regression model to predict house prices based on features like:

- Area (in square feet)
- Number of Bedrooms
- Number of Bathrooms
- Age of the House (in years)

i. Load the CSV file into R using read.csv().

ii. Perform exploratory data analysis (EDA) by plotting scatter plots and checking correlations.

iii. Fit a linear regression model using lm() where Price is the dependent variable.

iv. Analyze the model summary using summary().

v. Predict house prices using the fitted model.

vi. Evaluate model performance using RMSE or R-squared.