Linear Regression

Theory

LINEAR REGRESSION

Linear Regression is a statistical and numerical technique used to model the relationship between two variables by fitting a straight line through a set of observed data points. The objective is to determine the best-fit line that minimizes the difference between actual and predicted values.

In this project, Simple Linear Regression is implemented using the Least Squares Method, where the relationship between an independent variable (x) and a dependent variable (y) is approximated by a straight line. The general form of the regression equation is: y = a + bx where a = intercept b = slope of the line

Least Squares Principle: The best-fit line is obtained by minimizing the sum of squared errors between the observed values and the predicted values.

Error for each data point: ei = yi – (a + bxi)

FORMULAS :

Let n be the number of data points. Slope **(b) = (nΣxy – ΣxΣy) / (nΣx^2 – (Σx)^2) Intercept (a) = (Σy – bΣx) / n**

ALGORITHM STEPS:

1.Read number of data points (n) 2.Read n pairs of x and y values from input file 3.Store x and y values in dynamic arrays (vectors) 4.Calculate **Σx, Σy, Σxy, and Σx^2** 5.Compute slope (b) using least squares formula 6.Compute intercept (a) 7.Display input data points 8.Print the best-fit linear equation

CONVERGENCE:

Produces a unique best-fit line if data variance exists Accuracy improves with more data points Sensitive to outliers in data

TIME COMPLEXITY:

O(n), where n is the number of data points

ADVANTAGES:

Simple and efficient implementation Works well for linear relationships Requires only basic arithmetic operations No need for iterative approximation Suitable for real-world data analysis

DISADVANTAGES:

Assumes linear relationship between variables Sensitive to outliers Cannot model non-linear patterns Accuracy depends on quality of input data