## **Linear Regression Model**

## Linear regression is a statistical method used to model the relationship between a dependent variable (target) and one or more independent variables (predictors). It is used to predict continuous values by fitting a straight line to the data points.

## The general equation of a simple linear regression model (with one independent variable) is:

## Y=β0+β1X+ϵ

## Where:

## Y = Dependent variable (output)

## X = Independent variable (input)

## β0​ = Intercept (constant term)

## β1 = Slope of the regression line (coefficient)

## ϵ = Error term (difference between actual and predicted values)

## For **multiple linear regression** (with multiple independent variables), the equation becomes:

## Y=β0+β1X1+β2X2+...+βnXn+ϵ ​

## **Objective of Linear Regression Model**

## The main objective of the linear regression model is:

## **Predict Values** – Estimate the dependent variable based on independent variables.

## **Understand Relationships** – Identify the relationship between independent and dependent variables.

## **Minimize Error** – Reduce the difference between actual and predicted values using techniques like the least squares method.

## **Find Trends** – Identify trends and patterns in data for decision-making.

## **Step-by-Step Methods to Build a Linear Regression Model**

## **Step 1: Collect and Prepare Data**

## Gather relevant data containing dependent and independent variables.

## Remove missing or duplicate values.

## Handle outliers if necessary.

## **Step 2: Exploratory Data Analysis (EDA)**

## Visualize the relationship between variables using scatter plots and correlation matrices.

## Check for multicollinearity (correlation between independent variables).

## **Step 3: Split Data into Training and Testing Sets**

## Typically, split the dataset into **80% training** and **20% testing** using tools like train\_test\_split in Python.

## **Step 4: Choose the Regression Model**

## Simple Linear Regression (one independent variable)

## Multiple Linear Regression (multiple independent variables)

## Polynomial Regression (if the relationship is non-linear)

## **Step 5: Train the Model**

## Fit the model to the training dataset using methods like **Ordinary Least Squares (OLS)** to estimate the coefficients (β0,β1….).

## The OLS method minimizes the sum of squared residuals:

## ∑(Yactual​−Ypredicted​)2

## **Step 6: Evaluate the Model**

## Use performance metrics such as:

## **R² Score** – Measures how well the independent variables explain the variance in the dependent variable.

## **Mean Squared Error (MSE)** – Measures the average squared difference between actual and predicted values.

## **Root Mean Squared Error (RMSE)** – Square root of MSE for interpretability.

## **Step 7: Tune the Model**

## Remove irrelevant features to avoid overfitting.

## Use techniques like **Regularization (Lasso, Ridge Regression)** to improve generalization.

## **Step 8: Make Predictions**

## Once trained, use the model to predict new values.

## Apply it to unseen data and validate performance.

## **Step 9: Deploy and Interpret Results**

## Implement the model in real-world applications (e.g., sales forecasting, stock price prediction).

## Interpret coefficients to understand feature importance.