Linear Regression Model

1. Introduction

Linear Regression is one of the simplest and most widely used machine learning algorithms.
It establishes a **linear relationship** between **independent variable(s) (X)** and a **dependent variable (Y)** by fitting a straight line to the data.

The general equation of a linear regression model is:

$$Y = \beta_0 + \beta_1X + \epsilon$$

Where:

- Y → Dependent variable (target)
- X → Independent variable (feature)
- β_0 → Intercept (bias term)
- β_1 → Coefficient (slope)
- ϵ → Error term

---

2. Types of Linear Regression

1. **Simple Linear Regression** – One independent variable, one dependent variable.

$$Y = \beta_0 + \beta_1X$$

2. **Multiple Linear Regression** – More than one independent variable.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_nX_n$$

---

3. Assumptions of Linear Regression

- **Linearity**: Relationship between X and Y is linear.
- **Independence**: Observations are independent of each other.
- **Homoscedasticity**: Constant variance of residuals.
- **Normality**: Residuals should be normally distributed.
- **No Multicollinearity** (for multiple regression): Independent variables should not be highly correlated.

---

4. Steps to Build a Linear Regression Model

1. Import necessary libraries.
2. Load and preprocess dataset.

3. Split dataset into training and testing sets.
4. Fit the model using training data.
5. Predict on test data.
6. Evaluate performance.

---

5. Evaluation Metrics

- **Mean Squared Error (MSE)**

MSE=1n∑(yi−y^i)2MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2MSE=n1∑(yi−y^i)2

- **Root Mean Squared Error (RMSE)**

RMSE=MSERMSE = \sqrt{MSE}RMSE=MSE

- **R-Squared (R2R^2R2)**
  Explains how much of the variance in Y is explained by X.

R2=1−SSresSStotR^2 = 1 - \frac{SS_{res}}{SS_{tot}}R2=1−SStotSSres

---

6. Applications

- Predicting house prices based on features like area, location, etc.
- Sales forecasting.
- Risk analysis in finance.
- Medical research (e.g., relation between dosage and recovery).