1. What is Linear Regression?

Linear Regression is a **supervised machine learning algorithm** used for predicting a continuous target variable based on one or more input features. It assumes a **linear relationship** between the input features (X) and the output variable (Y).

- **Simple Linear Regression**: Only **one input feature**

  $$Y=\beta0+\beta1X+\epsilon Y=\beta_0+\beta_1X+\epsilon$$

  where:

  - Y = Dependent variable (target)
  - X= Independent variable (feature)
  - $\beta0$ = Y-intercept (bias term)
  - $\beta1$ = Slope (weight of the feature)
  - ϵ = Error term
- **Multiple Linear Regression**: **Multiple input features**
  $$Y=\beta0+\beta1X1+\beta2X2+\cdots+\beta nXn+\epsilon Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n+\epsilon$$

---

2. Key Assumptions of Linear Regression

For reliable results, Linear Regression makes certain assumptions:

1. **Linearity**: The relationship between X and Y should be linear.
2. **Homoscedasticity**: Residuals (errors) should have constant variance.
3. **Independence**: Observations should be independent (no autocorrelation).
4. **Normality**: Residuals should be normally distributed (for inference).
5. **No Multicollinearity**: Features should not be highly correlated.

---

3. How Does Linear Regression Work?

The algorithm tries to find the **best-fit line** by minimizing the **sum of squared errors (SSE)** between predicted and actual values.

Cost Function (Mean Squared Error - MSE)

$$\text{MSE}=1n\sum i=1n(\text{Yi}-\text{Yi}^{\wedge})2 \quad MSE=n1_{i=1}\sum_{n}(Y_i-Y_i^{\wedge})2$$

where:

- $\text{Yi}Y_i$ = Actual value
- $\text{Yi}^{\wedge}Y_i^{\wedge}$ = Predicted value

Optimization (Gradient Descent or Normal Equation)

- **Gradient Descent**: Iteratively adjusts weights to minimize MSE.
- **Normal Equation**: Directly computes optimal weights (closed-form solution).

---

4. Evaluation Metrics

To assess model performance:

1. **R-squared (Coefficient of Determination)**:
   - Measures the proportion of variance in Y explained by X.
   - Ranges from **0 to 1** (higher is better).
2. **Adjusted R-squared**:
   - Adjusts for the number of predictors (penalizes unnecessary features).
3. **Mean Absolute Error (MAE) & Mean Squared Error (MSE)**:
   - Lower values indicate better fit.

---

5. Advantages & Limitations

✅ **Advantages**:

- Simple to implement and interpret.
- Works well when relationships are linear.
- Fast training and prediction.

❌ **Limitations**:

- Sensitive to outliers.
- Assumes linearity (fails for complex patterns).
- Can underfit if data is non-linear.

---

6. Practical Applications

- Predicting house prices based on features (size, location).
- Forecasting sales based on advertising spend.
- Estimating stock prices based on market trends.