→ **Linear Regression**

- What is linear regression
Statistical method used to model the relationship between dependent variable (target) and one or more independent variables (features) by fitting a linear equation.
Assumes relationship between variables can be represented by a straight line.

- Why it is used
  - Predicts continuous numerical values
  - Quantifies the strength and direction of relationship
  - Simple, interpretable and fast
  - baseline model for complex algorithms

- How it works
  - Takes input features X and output y
  - Assumes a linear relationship between them
  - Fits a line that minimizes the error between predicted and actual values.
  - Uses optimization techniques like gradient descent to find best parameters.
  - Produces predictions for new unseen data.

- Where it is used
  - Data Science
  - Machine Learning
  - Economics
  - Engineering
  - Business Analytics
  - Scientific Research

- Formula

• Simple: $y = mx + c$
• Multiple: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$

y = dependent variable
x = independent variable
β_0 = intercept
$\beta_1, \beta_2, \dots$ = coefficients
m = slope
c = intercept

- Technical Details

• Cost Function (Mean Squared Error):

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

• Optimization methods: normal equation, gradient descent
• Evaluation metrics: R^2 score
Mean Squared Error (MSE)
Root Mean Squared Error (RMSE)
Mean Absolute Error (MAE)

- Pros

• Easy to understand and implement
• Highly interpretable results
• Computationally efficient
• Works well with small datasets
• Useful for feature importance analysis.

- Cons
  - Assumes linear relationship
  - Sensitive to outliers
  - Poor performance with non-linear data
  - Requires assumptions to be satisfied
  - Multicollinearity affects stability

- Real-World Applications (Where)
  - Business & Economics: Sales forecasting
    Price prediction
    Demand analysis
  - Healthcare: Predicting patient recovery time
    Medical cost estimation
  - Finance: Stock price trends
    Risk assessment
    Credit scoring
  - Engineering: System performance prediction
    Load and stress analysis
  - Social Sciences: Population growth modelling
    Income vs education studies.

- Assumptions
  - Linearity
  - Independence of errors
  - Homoscedasticity (constant variance)
  - No multicollinearity
  - Normally distributed errors.

- Purpose (When)
  - Predict a continuous numerical value based on one or more input features
  - Model and quantify the relationship between independent variables and dependent variables.
  - Estimate how changes in input variables affect the output.
  - Provide interpretable insights through coefficients (slope and intercept)
  - Serve as a baseline predictive model in regression problems.