Definition: Linear regression is a statistical method for modeling the relationship between a dependent variable and one or more independent variables using a linear equation. The aim is to find the best-fitting line through the data points.

Concepts:

Dependent Variable: The variable you're trying to predict (Y).

Independent Variables: The variables used to predict the dependent variable (X).

Line of Best Fit: The line that minimizes the sum of the squared differences between the observed values and the values predicted by the line (least squares method).

Intercept: The value of Y when all Xs are zero.

Slope: The rate of change in Y for a one-unit change in X.

Advantages:

Simplicity and ease of interpretation.

Fast computationally.

Works well with linearly separable data.

Provides insights into the relationship between variables.

Disadvantages:

Assumes linearity, which might not hold true for all datasets.

Sensitive to outliers.

Assumes independence of errors.

Doesn't handle multicollinearity well.

Applications:

Business: Forecasting sales, analyzing market trends.

Economics: Estimating economic indicators like GDP.

Healthcare: Predicting disease progression.

Real Estate: Estimating property prices.

Social Sciences: Studying relationships between variables in sociological data

Techniques for accuracy :

1. Mean Sqaured Error

2. R2\_score