

SNS COLLEGE OF TECHNOLOGY

**An Autonomous Institution
Coimbatore-35**



DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

23ADT202 – FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

II YEAR IV SEM

UNIT II – LEAST SQUARES REGRESSION LINE

Introduction to Linear Regression

EMPATHY:



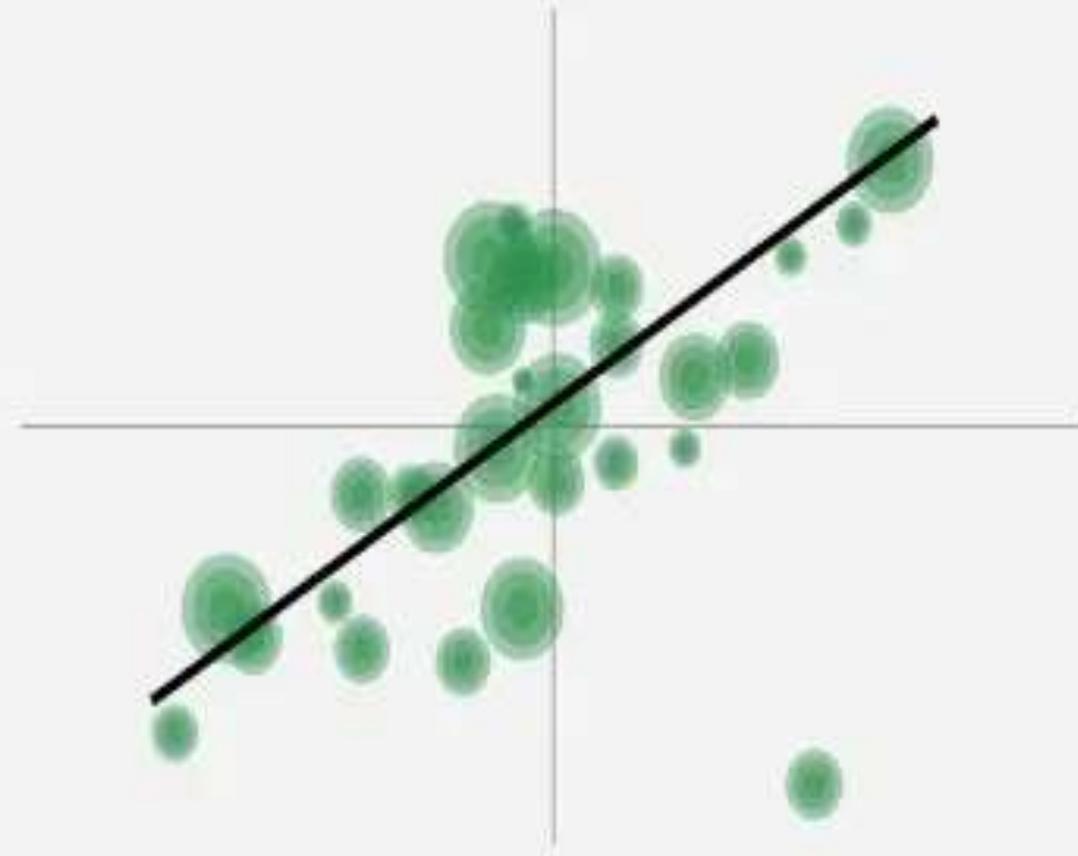
Simple model for predicting continuous values.



It finds the relationship between input features and output.

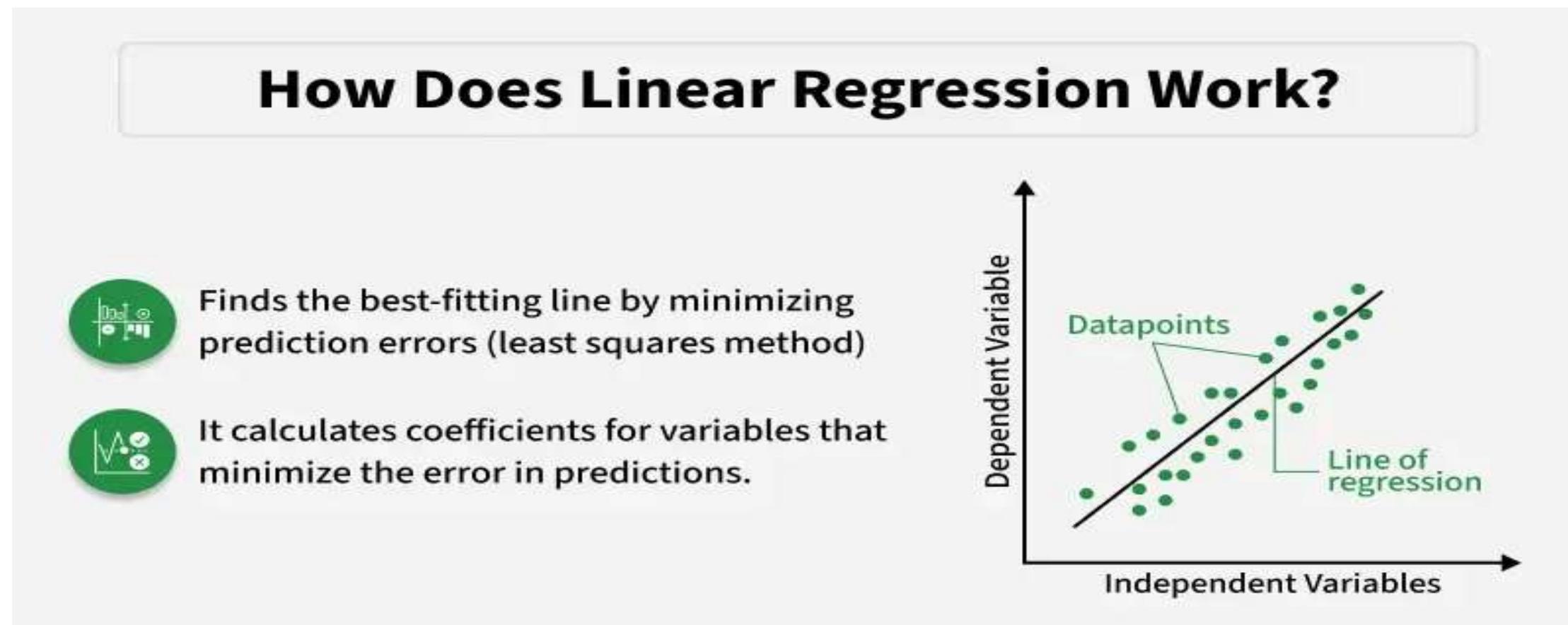


It is used in forecasting and predicting salaries based on experience.



DEFINE:

- ❖ Linear regression is a type of supervised machine-learning algorithm that learns from the labelled datasets and maps the data points with most optimized linear functions which can be used for prediction on new datasets.
- ❖ It assumes that there is a linear relationship between the input and output, meaning the output changes at a constant rate as the input changes. This relationship is represented by a straight line.



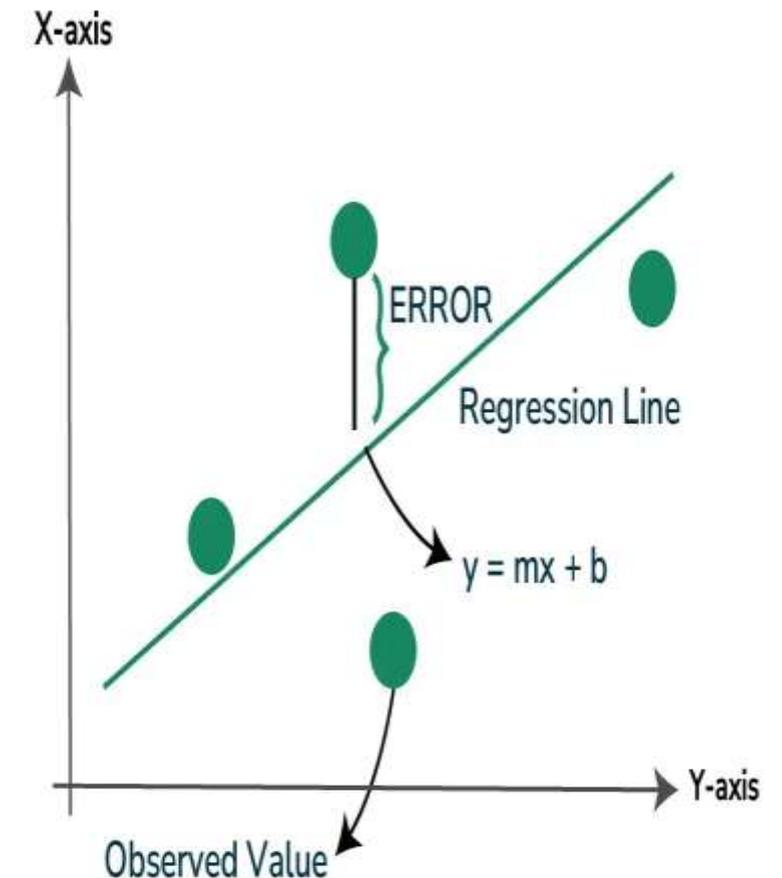
IDEATE:

Least Square Method



Least Square Method

- The Least Square method is a popular mathematical approach used in data fitting, regression analysis, and predictive modeling.
- It helps find the **best-fit line or curve** that minimizes the sum of squared differences between the observed data points and the predicted values.
- This technique is widely used in statistics, machine learning, and engineering applications.



- Least Squares Method is used to derive a generalized linear equation between two variables.
- When the value of the dependent and independent variables they are represented as x and y coordinates in a 2D Cartesian coordinate system.
- Initially, known values are marked on a plot. The plot obtained at this point is called a scatter plot.
- Then, we try to represent all the marked points as a straight line or a linear equation.
- The equation of such a line is obtained with the help of the Least Square method.
- This is done to get the value of the dependent variable for an independent variable for which the value was initially unknown.
- This helps us to make predictions for the value of a dependent variable.

PROTOTYPING:

Least Square Method Formula

- The Least Square Method formula finds the best-fitting line through a set of data points.
- For a simple linear regression, which is a line of the form $y=mx+c$,
- where y is the dependent variable, x is the independent variable, a is the slope of the line, and
- b is the y -intercept, the formulas to calculate the slope (m) and intercept (c) of the line are derived from the following equations:

$$\text{Slope (m) Formula: } m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$\text{Intercept (c) Formula: } c = \frac{(\sum y) - a(\sum x)}{n}$$

Where:

- n is the number of data points,
- $\sum xy$ is the sum of the product of each pair of x and y values,
- $\sum x$ is the sum of all x values,
- $\sum y$ is the sum of all y values,
- $\sum x^2$ is the sum of the squares of the x values.

Steps to find the line of Best Fit by using the Least Squares Method :

Step 1: Denote the independent variable values as x_i and the dependent ones as y_i .

Step 2: Calculate the average values of x_i and y_i as X and Y .

Step 3: Presume the equation of the line of best fit as $y = mx + c$, where m is the slope of the line and c represents the intercept of the line on the Y-axis.

Step 4: The slope m can be calculated from the following formula:

$$m = [\Sigma (X - x_i) \times (Y - y_i)] / \Sigma (X - x_i)^2$$

Step 5: The intercept c is calculated from the following formula:

$$c = Y - mX$$

Thus, we obtain the line of best fit as $y = mx + c$, where values of m and c can be calculated from the formulae defined above.