Least Squares Regression Method

edureka!

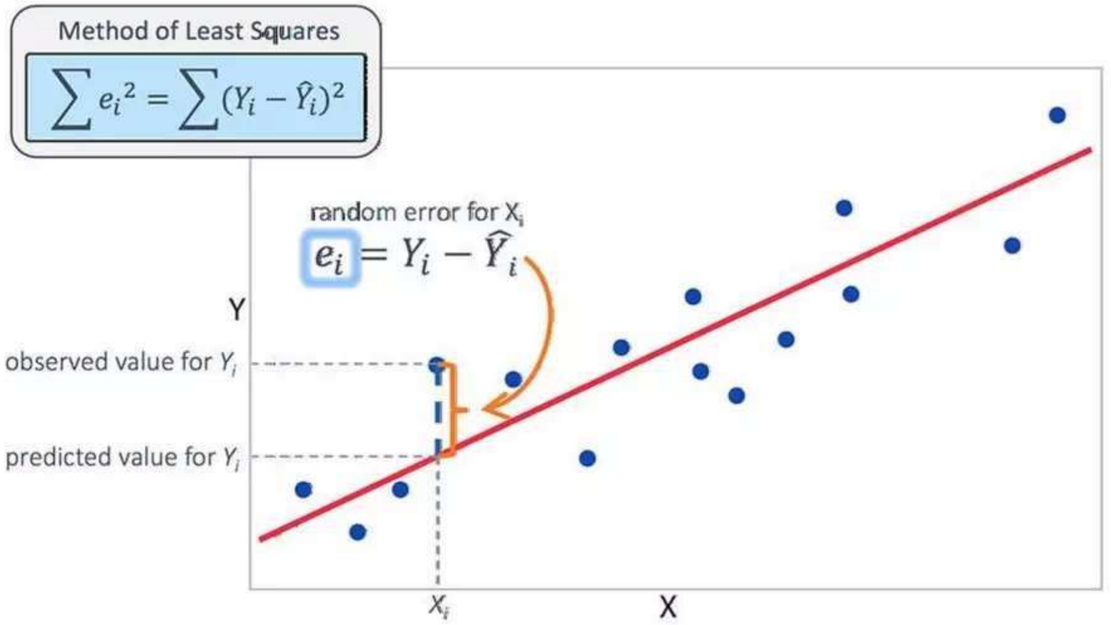
www.edureka.co/masters-program/machine-learning-engineer-training

AGENDA

WHAT IS THE LEAST SQUARES METHOD?
LINE OF BEST FIT
STEPS TO COMPUTE THE LINE OF BEST FIT
THE LEAST-SQUARES REGRESSION METHOD WITH AN EXAMPLE
REGRESSION ANALYSIS USING PYTHON

01

WHAT IS THE LEAST SQUARES REGRESSION METHOD?



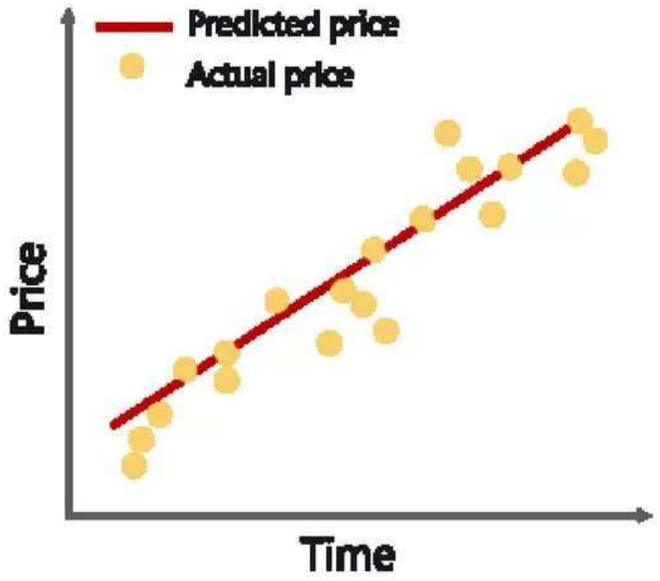
What is the Least Squares Regression Method?

The least-squares regression method is a technique commonly used in Regression Analysis. It is a mathematical method used to find the best fit line that represents the relationship between an independent and dependent variable in such a way that the error is minimized.

02

LINE OF BEST FIT

The least-squares method is one of the most effective ways used to draw the line of best fit. It is based on the idea that the square of the errors obtained must be minimized to the most possible extent and hence the name least squares method.

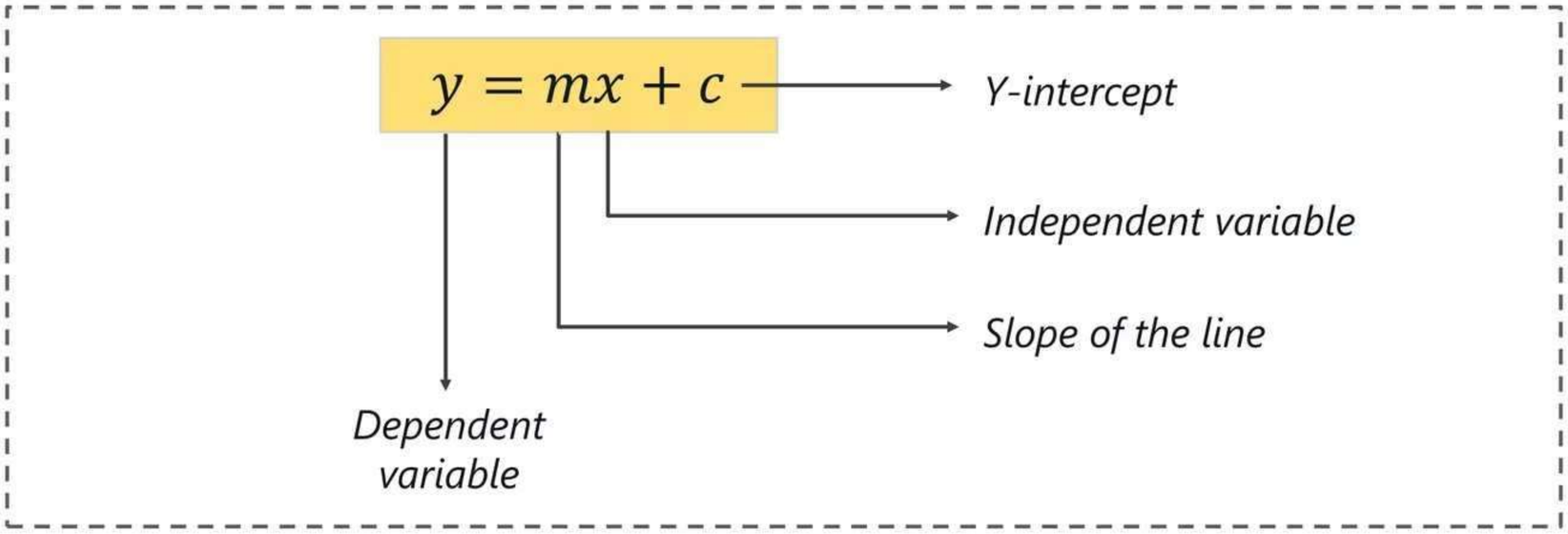


What is the Line Of Best Fit?

The Line of best fit line is drawn across a scatter plot of data points in order to represent a relationship between those data points.

03

STEPS TO COMPUTE THE LINE OF BEST FIT



A simple equation that represents a straight line along 2-Dimensional data, i.e. x-axis and y-axis.

Steps to calculate the Line of Best Fit

To start constructing the line that best depicts the relationship between variables in the data, we first need to get our basics right.

$$m = \frac{n\sum xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

m – slope of the line

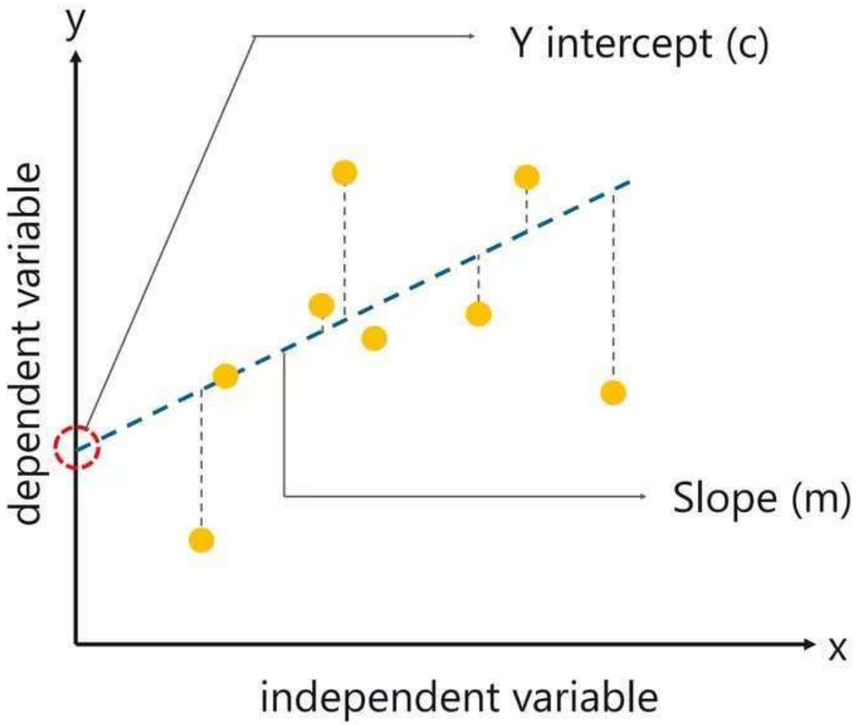
n – total number of data points

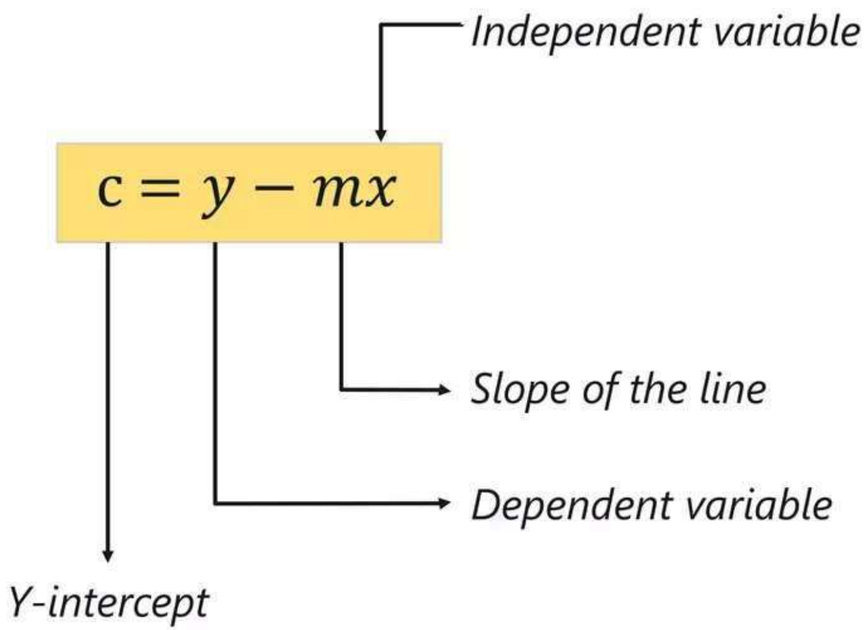
x – Independent variable

y – Dependent variable

Step 1: Calculate the slope 'm' of the line

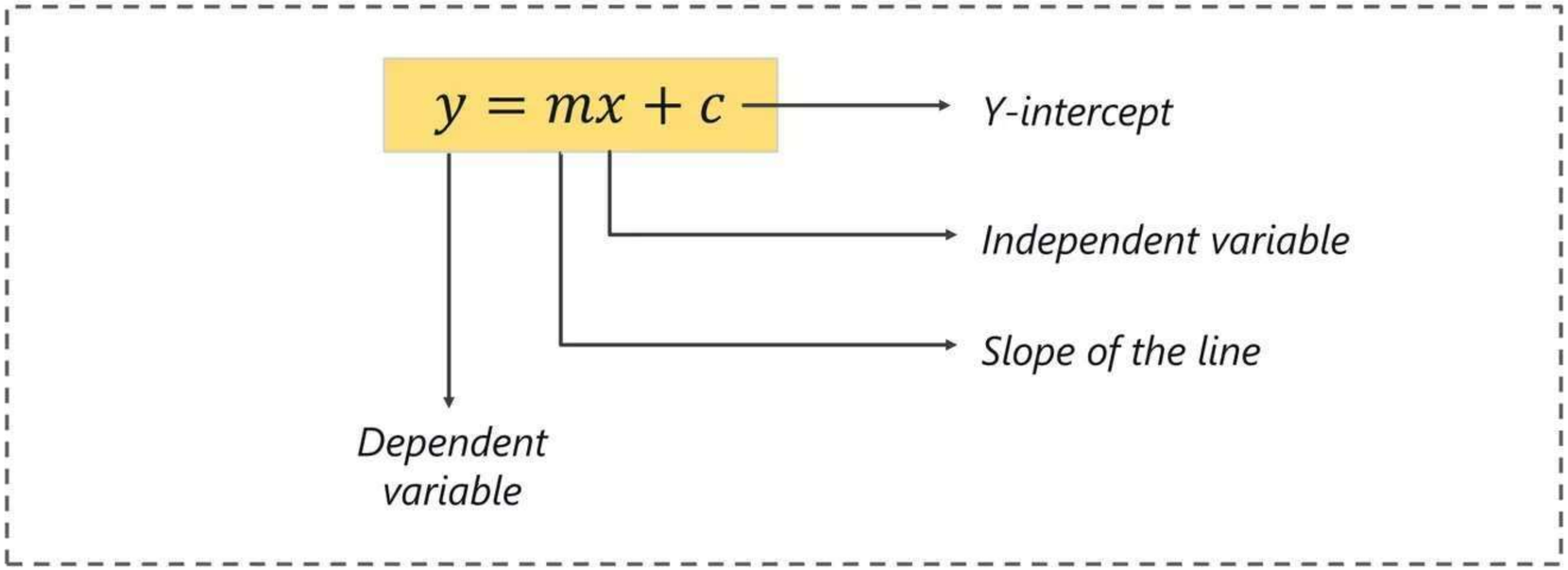
The slope of a line characterizes the direction of a line. To find the slope, you divide the difference of the y-coordinates of 2 points on a line by the difference of the x-coordinates of those same 2 points .





Step 2: Compute the y-intercept

The y-intercept of a line is the value of y at the point where the line crosses the y axis.



Step 3: Substitute the values in the final equation

A simple equation that represents a straight line along 2-Dimensional data, i.e. x-axis and y-axis.

04

THE LEAST-SQUARES REGRESSION METHOD WITH AN EXAMPLE

edureka!

YouTube Video Link in the Description

REGRESSION ANALYSIS EXAMPLE

Tom who is the owner of a retail shop, found the price of different T-shirts vs the number of T-shirts sold at his shop over a period of one week.

Price of T-shirts in dollars (x)	# of T-shirts sold (y)
2	4
3	5
5	7
7	10
9	15

Step 1: Calculate the slope 'm'

$$m = \frac{n\sum xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2} = 1.518 \text{ approx}$$

Step 2: Compute the y-intercept value

$$c = y - mx = 0.305$$

Step 3: Substitute the values in the final equation

$$y = 1.518\,x + 0.305$$

REGRESSION ANALYSIS EXAMPLE

Tom who is the owner of a retail shop, found the price of different T-shirts vs the number of T-shirts sold at his shop over a period of one week.

Price of T-shirts in dollars (x)	# of T-shirts sold (y)	Y=mx+c	error
2	4	3.3	-0.67
3	5	4.9	-0.14
5	7	7.9	0.89
7	10	10.9	0.93
9	15	13.9	-1.03

Step 1: Calculate the slope 'm'

$$m = \frac{n\sum xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2} = 1.518 \text{ approx}$$

Step 2: Compute the y-intercept value

$$c = y - mx = 0.305$$

Step 3: Substitute the values in the final equation

Consider x = $8,

y = 1.518 x 8 + 0.305 = 12.45 = 13 T-shirts

- *The data must be free of outliers.*
- *Compute a line with the minimum possible squares of errors.*
- *Least Squares Regression works perfectly even for non-linear data.*
- *Technically, the difference between the actual value of 'y' and the predicted value of 'y' is called the Residual*

Points To Remember

A few things to keep in mind before implementing the least squares regression method

05

REGRESSION ANALYSIS USING PYTHON

Data Set Description: The data set contains the following variables:

- *Gender:* Male or female represented as binary variables
- *Age:* Age of an individual
- *Head size in cm^3:* An individuals head size in cm^3
- *Brain weight in grams:* The weight of an individual's brain measured in grams

	A	B	C	D
1	Gender	Age Range	Head Size	Brain Wei
2	1	1	4512	1530
3	1	1	3738	1297
4	1	1	4261	1335
5	1	1	3777	1282
6	1	1	4177	1590
7	1	1	3585	1300
8	1	1	3785	1400
9	1	1	3559	1255
10	1	1	3613	1355
11	1	1	3982	1375
12	1	1	3443	1340
13	1	1	3993	1380
14	1	1	3640	1355
15	1	1	4208	1522
16	1	1	3832	1208
17	1	1	3876	1405
18	1	1	3497	1358
19	1	1	3466	1292
20	1	1	3095	1340

Linear Regression Using Least Squares Regression Method In Python

Problem Statement: To apply Linear Regression using Least Squares Regression method and build a model that studies the relationship between the head size and the brain weight of an individual..

THANK YOU