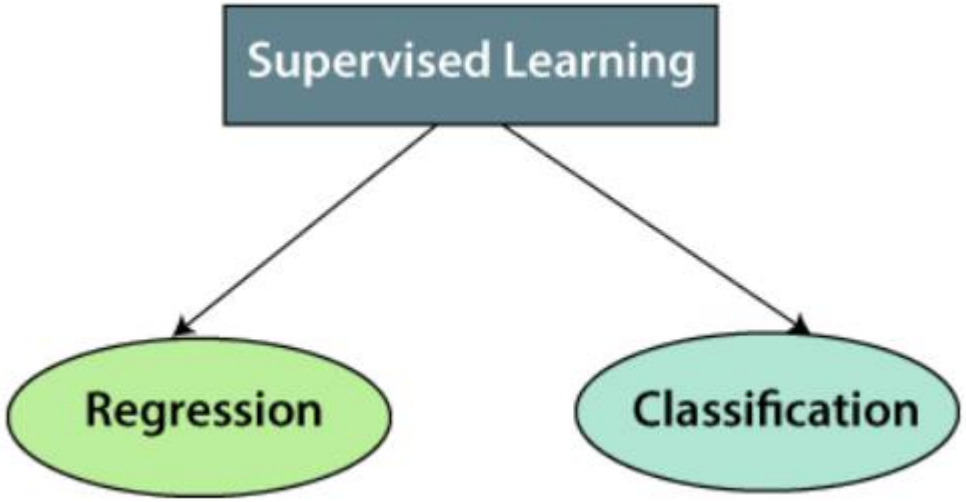
Simple Linear Regression

Supervised Learning :

Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.

Types: Supervised learning can be further divided into two types of problems:

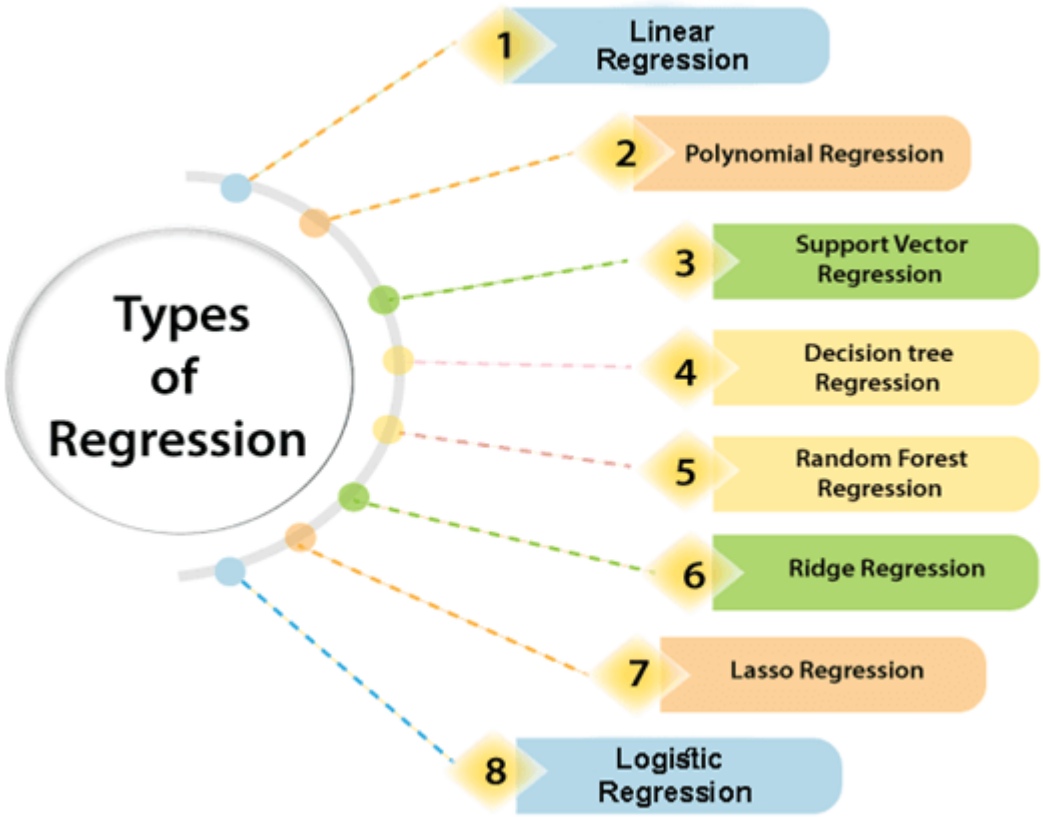


Simple Linear Regression

Regression:

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

Types:



Simple Linear Regression

Simple Linear Regression:

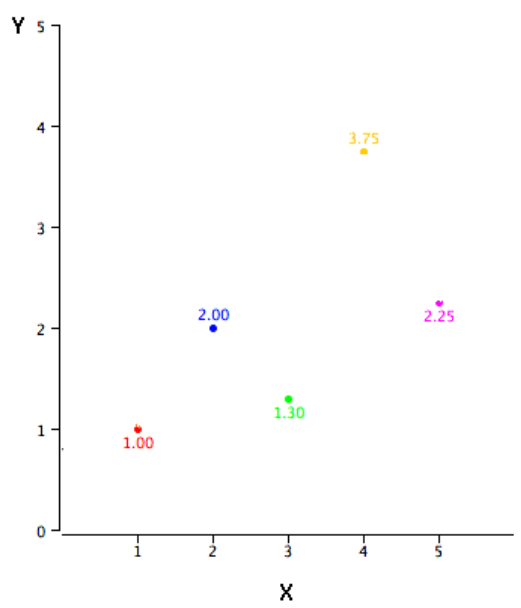
Simple linear regression is **a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line**. Both variables should be quantitative.

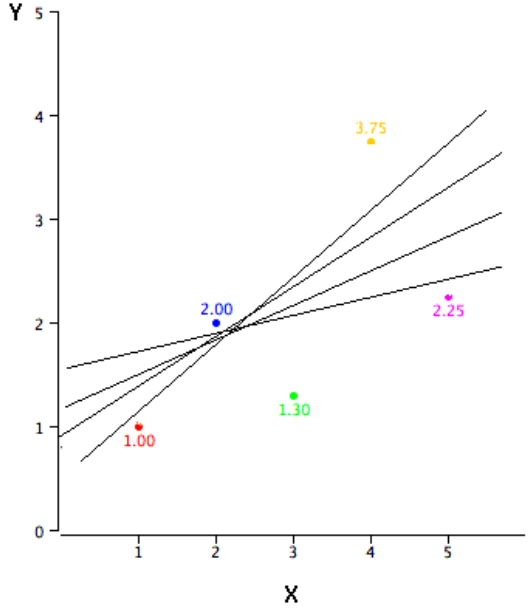
The example data in Table

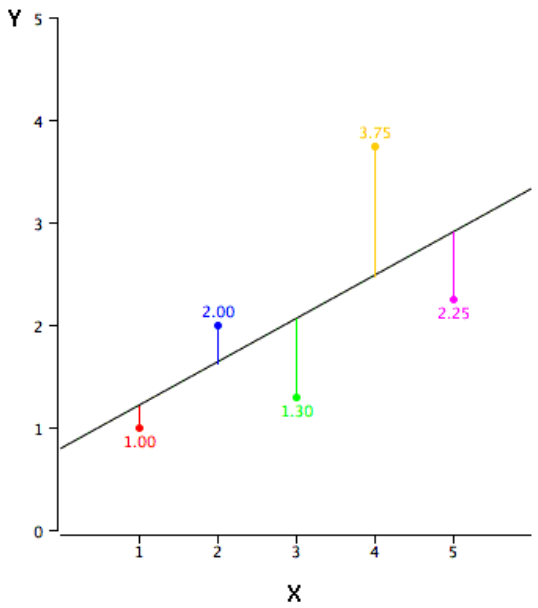
X (Independent Variable)	Y (Dependent Variable)
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

Simple Linear Regression

Linear regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a *regression line*







Simple Linear Regression

X	Y	x^2	XY	Y'
1	1	1	1	?
2	2	4	4	?
3	1. 3	9	3.9	?
4	3. 75	16	15	?
5	2. 25	25	11.25	?
15	10.3	55	35.15	

The formula for a regression line is

Y' = bX + a

where Y' is the predicted score, b is the slope of the line, and A is the Y intercept.

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

$$= \frac{5 * 35.15 - 15 * 10.3}{5 * 55 - 15 * 15}$$

$$= 0.425$$

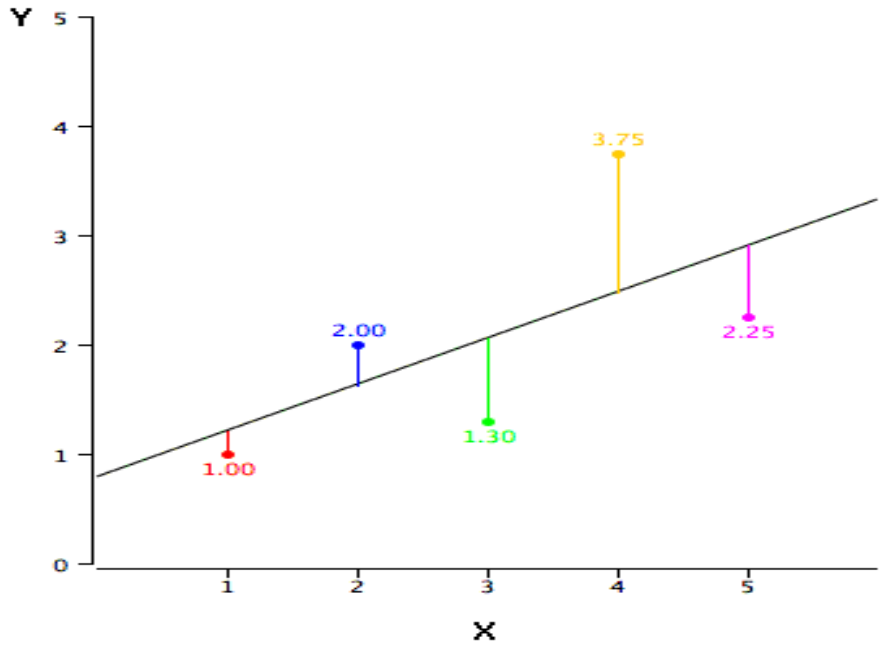
$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n(\sum x^2) - (\sum x)^2}$$

$$= \frac{10.3 * 55 - 15 * 35.15}{5 * 55 - 15 * 15}$$

$$= 0.785$$

Simple Linear Regression

X	Y	Y'
1.00	1.00	1.210
2.00	2.00	1.635
3.00	1.30	2.060
4.00	3.75	2.485
5.00	2.25	2.910



Side Note:

Y′=0.425*1+0.785=1.210

Y′=0.425*2+0.785=1.635

Y′=0.425*3+0.785=2.060

Y′=0.425*4+0.785=2.485

Y′=0.425*5+0.785=2.910

Simple Linear Regression

X	Y	Y'	Y-Y'	$(Y-Y')^2$
1.00	1.00	1.210	-0.210	0.044
2.00	2.00	1.635	0.365	0.133
3.00	1.30	2.060	-0.760	0.578
4.00	3.75	2.485	1.265	1.600
5.00	2.25	2.910	-0.660	0.436

The error of prediction for a point is the value of the point minus the predicted value (the value on the line). Table shows the predicted values (Y') and the errors of prediction (Y-Y'). For example, the first point has a Y of 1.00 and a predicted Y (called Y') of 1.21. Therefore, its error of prediction is -0.21.

Total Error = Sum of Squares Error=2.791

The sum of the squared errors of prediction shown in Table is lower than it would be for any other regression line.

Simple Linear Regression

X	Y	Y'
1.00	1.00	1.210
2.00	2.00	1.635
3.00	1.30	2.060
4.00	3.75	2.485
5.00	2.25	2.910

Side Note:

Y′=0.425*1+0.785=1.210

Y′=0.425*2+0.785=1.635

Y′=0.425*3+0.785=2.060

Y′=0.425*4+0.785=2.485

Y′=0.425*5+0.785=2.910

Y′=0.425*6+0.785=3.335