Regression

It is from of predictive modelling technique which investigates the relationship b/w dependent and independent variables. **Dependent variable** – which you want to predictive. (Means target). **Independent variable** – which is predictor. (Never ended factors).

“Regression shows a line or curve that passes through all the datapoints on target-predictor graph in such a way that the vertical distance between the datapoints and the regression line is minimum.”

Terminologies

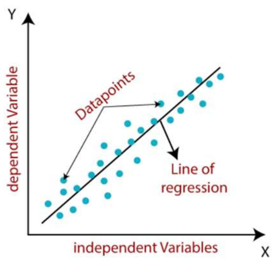
Outlier is an observation which contains either very low value or very high value in comparison to other observed values. An outlier may hamper the result, so it should be avoided.

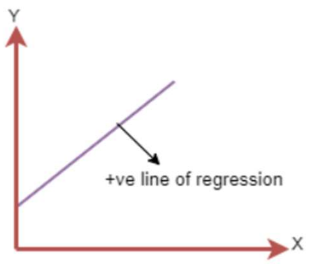
If the independent variables are highly correlated with each other than other variables, then such condition is called **Multicollinearity**.

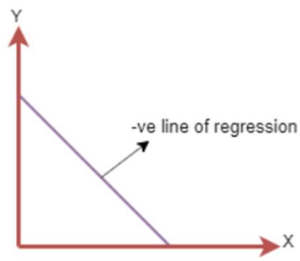
If our algorithm works well with the training dataset but not well with test dataset, then such problem is called **Overfitting**. And if our algorithm does not perform well even with training dataset, then such problem is called **underfitting**.

Linear Regression

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression.







Types of Linear Regression

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called **Simple Linear Regression**.

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called **Multiple Linear Regression**.

Linear line equation $\qquad y = mx + c$

x	Y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})$	$(y - \bar{y})$
1	3	-2	-0.6	4	1.2
2	4	-1	0.4	1	-0.4
3	2	0	-1.6	0	0
4	4	1	0.4	1	0.4
5	5	2	1.4	4	2.8

(x = 3, y = 3.6)

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} = \frac{4}{10} = 0.4$$

y = mx +c

3.6=0.4(3) + c

c = 3.6 – 1.2

c = 2.4

y = 0.4x + 2.4

x = {1, 2, 3, 4, 5}

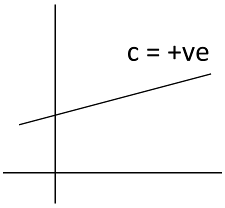
y = 0.4(1) + 2.4 = 2.8

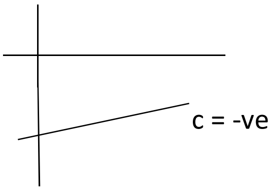
y = 0.4(2) + 2.4 = 3.2

y = 0.4(3) + 2.4 = 3.6

y = 0.4(4) + 2.4 = 4.0

y = 0.5(5) + 2.4 = 4.4





m = +ve angle $< 90'$

m = -ve angle $> 90'$

m = 0 angle = $0'$