**LINEAR REGRESSION**

Linear regression is a supervised machine learning algorithm ,it is kind of predictive analysis which uses relationship between the continues variable

**condition:** y should be continous

Let us understand it by an example , lets consider a scenerio of no of hours studied and marks obtained , so our goal is to design a model that will predict the marks on providing the data of no of hours studied,

Here no of hours studied will be our independednt variable

And maarks obtained is depenedent

So how we will dtermine it

**y=b0+b1x+e**

Where bo and b1 are the significant values choosen such that the error should be minimum

E is the error

There are various method to calculate the error one of them is ordinary least square method

**e=y^-y**

Y^ is the predicted value and y is actual value

If we dohnt square the error are positive nd negative will get cancel out

**Exploring the coffeients b1:**

If vaues of b1 >0 there is a +ve relationship btw the input and output variable

If values b1<0 there is -ve relationship btw the input an output variable

**Assumptions:**

In Linear regression we assume that varianton in error /residual are of constant variance

Hence if it is not we called this problem as heterocedasity, heterosecadastity is the unequal scatter of data

It is a problem because optimizing the best fit line using OLS method using this will be unaccurate hence

So how this problem can be solve?

**1.By increasig the weight of the variables**

**2.Transformation**

**3.Redefiining the variables**

**#while building the model**

There are variables thing u need to consider while model building

**R(Correlation analysis) :** corrlation analysis is the degree of relationship between the variables .

The correlation analysis between independent variables should not be considered

As the it is assumed that if there i the relationship between the independent variables or they effect the model will wont be ableto perform or predict as much better it can

It value lie between -1 to -1

**Multiple R^2 values**(Correlation analysis):

it is basically the variation between the predicited value and actual value it should be high A higher coefficient is an indicator of a better [goodness of fit](https://www.statisticshowto.datasciencecentral.com/goodness-of-fit-test/) for the observations.The CoD can be negative, although this usually means that your model is a poor fit for your data. It can also become negative if you didn’t set an intercept.

It is cosider that the value sof R sq value should be greter than 0.6

One more concpt comes upp with linear regression is **gradient descent**

It is basiclally a optimization technique