Supervised learning is ML approach thats defined by its use of labeled datasets

Supervised ML algo uses labels

labels (both input & output)

inputs are called as independent features

outputs are called as dependent features

outputs are also called as target Variables

Supervised ML Algorithms

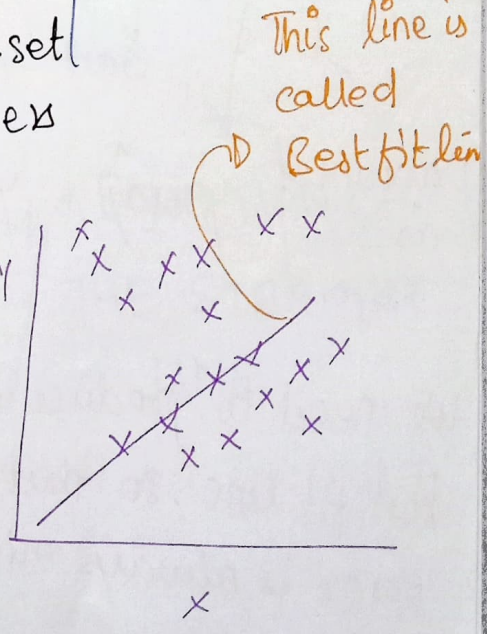
↙ Regression (continous data) output

↓ classification output Binary (or) multiclass (or) Categorical

In multiple linear regression we have more than one independent features

In supervised machine learning Algo, the model learns from the existing dataset and needs to produce accurate answers when new inputs are given to it

It predicts the output for new input based on the labels it has



Regression is one of the supervised Algorithm

Best fit line or line of best fit is line that produces the accurate result for any new input

x & y are plotted on to the graphs

input outputs

(x x / x x) on the graphs are co-ordinates of (x, y) label

If the best fit line is not correct then our y' values will wrong and large error

FRONT

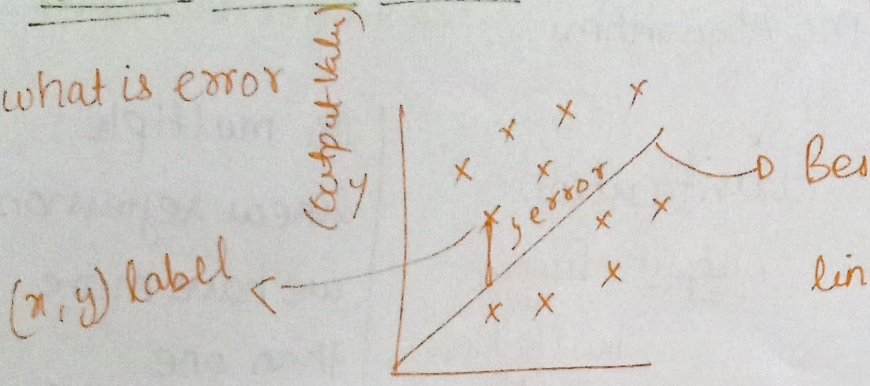
Aim of linear regression is to find the line of best fit

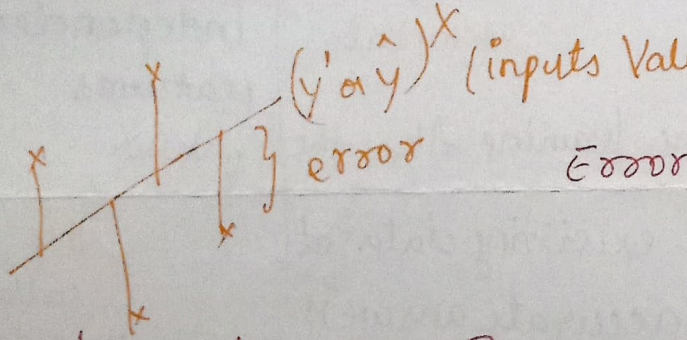
linear regression is the process of finding a line that best fits the data points available on the plot so that we can use it to predict output values for new given inputs

Before we need to Know

ERROR = Actual value - estimated value

what is error





Error = Actual value - estimated value

pronounced as [y' or ŷ = y hat]

y [Actual output from the data we have]

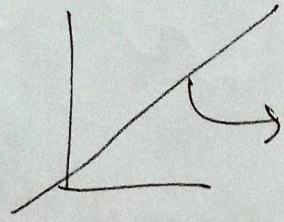
y' or ŷ [predicted data that best fit line]

We need to produce the best fit line, so that the error is always minimum

We need to calculate various line that is nearest to all the points and then find the line that produces min Error

The line which produces the min Error is then concluded as best fit line and it will give outputs to the new input

Aim for linear regression: produce this straight line
(or)
best fit line



→ this is the straight line or in linear regression called line of best fit

but in general we need to know the Math of the Straight line

So → What is the equation of Straight line

The equation (general equation) for any straight line is $y = mx + c$

$y = mx + c$ → is y-intercept (where the line crosses the y-axis)

is Gradient (steepness) of the line

$$Y = mx + c$$

} EQUATION OF THE STRAIGHT LINE

Y is the data point that has to predicted (OUTPUT)

m is the slope

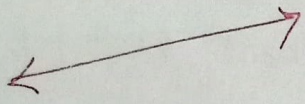
X is the data point that we have (INPUT)

C is the intercept

FRONT

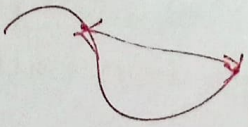
Math of Straight line

Straight line



B/w any two points of Straight line distance and displacement are Same

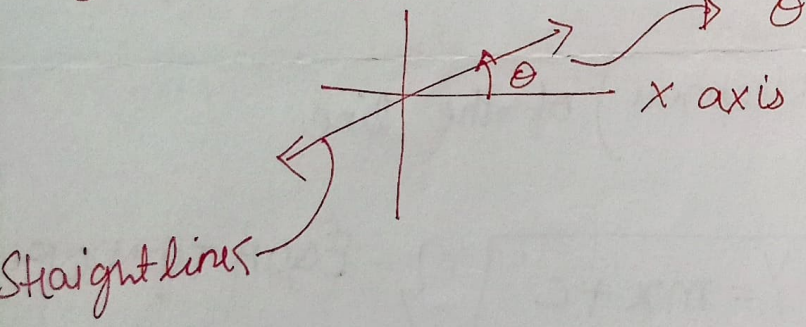
Curve line



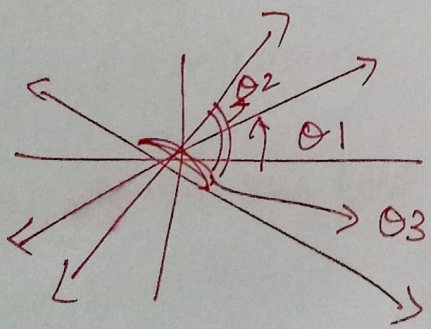
B/w any two points of curve line distance & the displacement is not same

Inclination Important

It is the angle made with some axis



θ is the angle created when the degrees is measured from some axis to the Straight line



Slope then What is Slope

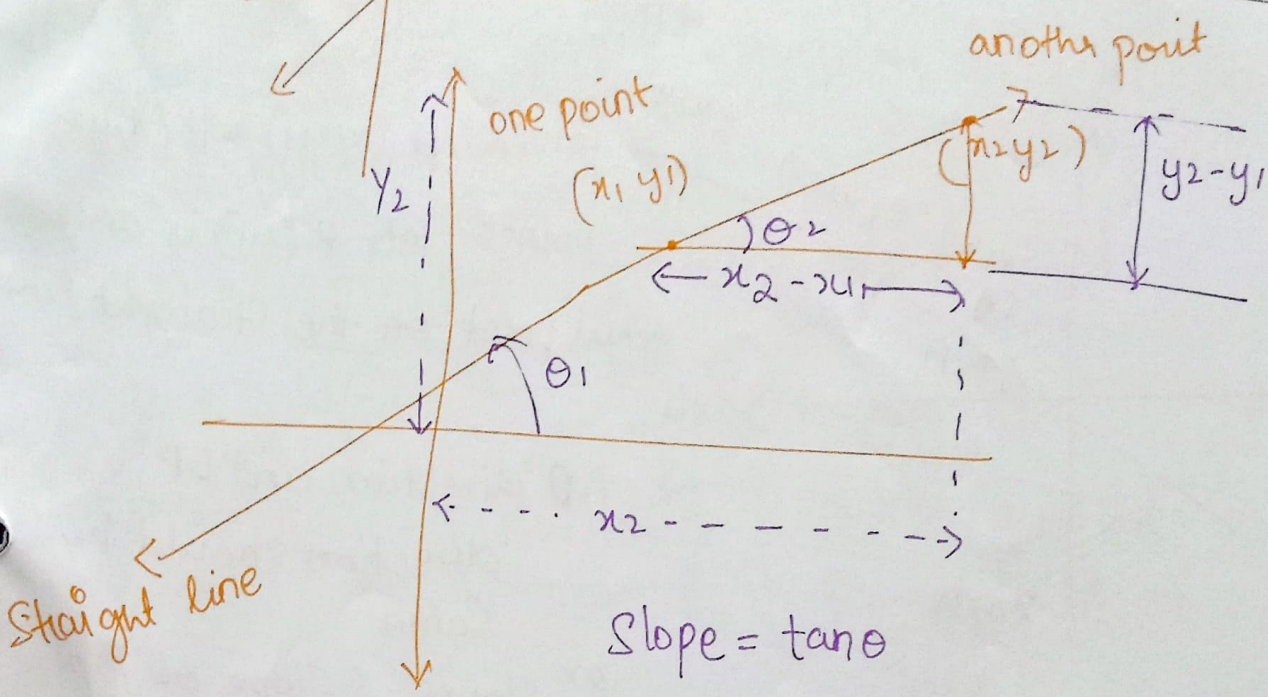
Inclination and Slope are Connected

Slope $\tan\theta$

Straight line

θ is the inclination

Slope = $\frac{y_2 - y_1}{x_2 - x_1}$



θ_1 and θ_2 will be same by corresponding rule

Slope = $\tan\theta$

$= \frac{\text{opposite}}{\text{adjacent}}$

$= \frac{\text{opposite} \Rightarrow y_2 - y_1}{\text{adjacent} \Rightarrow x_2 - x_1}$

$$\text{Slope} = \frac{y_2 - y_1}{x_2 - x_1}$$

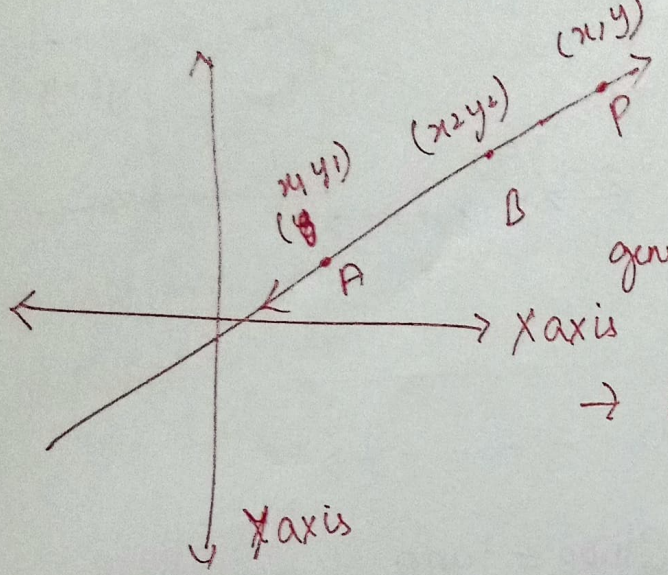
} This is the formula of the slope

In slop we always take $\tan\theta$, Not sin, cos this is because we need to find the angle, and that is purbl cos we have $(y_2 - y_1)$ (opposite) and adjacent $(x_2 - x_1)$ this two are required to calculate the slope

FRONT

Now we know that Slope = $\frac{y_2 - y_1}{x_2 - x_1}$

Now we need to know the TWO POINT FORM OF A STRAIGHT LINE



Given A $(x_1 y_1)$ & B (x_2, y_2) points lets P $(x y)$ is any genral point on the straight line

→ AP direction and BP direction should be same

or Slope of AB = Slope of BP

then P point (x, y) will Sure lie on the straight line

$m_{AB} = m_{AB}$

Slope of AP = Slope of BP

$\Rightarrow \frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}$

$y - y_1 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x - x_1)$

We now that Slope formula = $\frac{y_2 - y_1}{x_2 - x_1}$

$m = \frac{y_2 - y_1}{x_2 - x_1}$

$y - y_1 = \underline{m}\,(x - x_1)$

Slope

This is the equation of Straight line passing through the two points (x_1, y_1) & (x_2, y_2)

Now we know that general two point form of a straight line is $\Rightarrow y - y_1 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x - x_1)$

$y - y_1 = m(x - x_1)$ $\Rightarrow y - y_1 \quad m(x - x_1)$

$y - y_1 = mx - mx_1$

$y = mx - mx_1 + y_1$ y this is constant

$\rightarrow c$

$y = mx + c$

$\rightarrow$ $\boxed{y = mx + c}$ is the general equation of the slope

General equation of straight line

Y = prediction value (output)

m = slope $\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$

x = datapoint value (input)

c = constant

Any linear equation $Ax + By + C = 0$ we can find out the slope of it

$mx = y + c$ or $y = mx + c$

$Ax + By + C = 0$ compare

$By = -Ax - C$

$\left\{y = -\left(\frac{A}{B}\right)x - c\right\} \Rightarrow$ It is in the form of $y = mx + c$

$-\frac{A}{B}$ is the slope

FRONT

Given $3x + 4y - 5 = 0$

Find out the slope

$3x + 4y - 5 = 0$

$4y = -3x + 5$

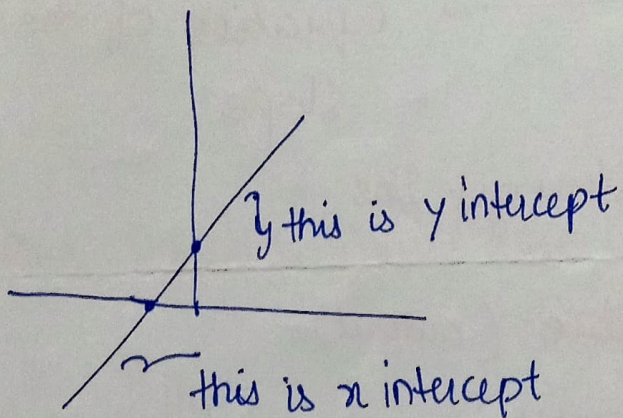
$y = \frac{-3}{4}x + \frac{5}{4}$ It is in the form of $y = mx + c$

$\therefore \frac{-3}{4}$ is the slope for the above equation

$y = \left(\frac{-3}{4}\right)x + \frac{5}{4}$ ↳ slope

m = slope

Another concept Incept



The line crossing on the axis can be considered to be intercept

intercept can be either positive or either negative

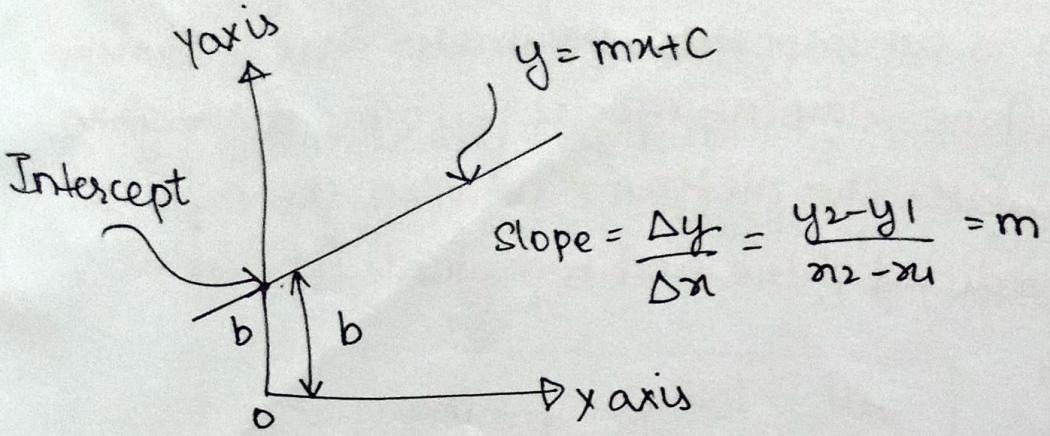
+y or -y	-x or +x
intercept	intercept

Now we know all the terms of (equation of straight)

tasks accomplished → Slope ✓

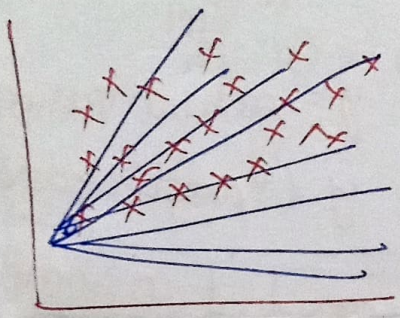
→ Intercept ✓

General equation of the line



$$\text{Slope} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = m$$

So back to machine learning from Maths



we need to change this value m and c and see what line best fits all the data point

We need to check which point (w, c) value will give the min error

The minium error is achieved then that "line" slope is declared to be "best fit line" for the given data set

This (w, b) in different terms we use as θ_1, θ_2
w, c
w, b
β_1, β_2

This Reduction of errors to find best line is done by Calculating the "COST FUNCTION"

FRONT

WHAT IS COST FUNCTION

a cost function is an important parameter that determines how well a machine learning model performs for a given dataset. It calculates the difference between the expected value and the predicted value and represents its as a single number

Cost function are of various type one among then is (MSE)

MSE => mean squared error

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y-\hat{y})^2$$

y = actual value

$\hat{y}$ = predicted value

$\frac{1}{n}$ or $\frac{1}{2n}$ in Some books

Wenced to substitute various M & C values and Calculate the $y = mx+C$ ↳ this is y' Now calculate the

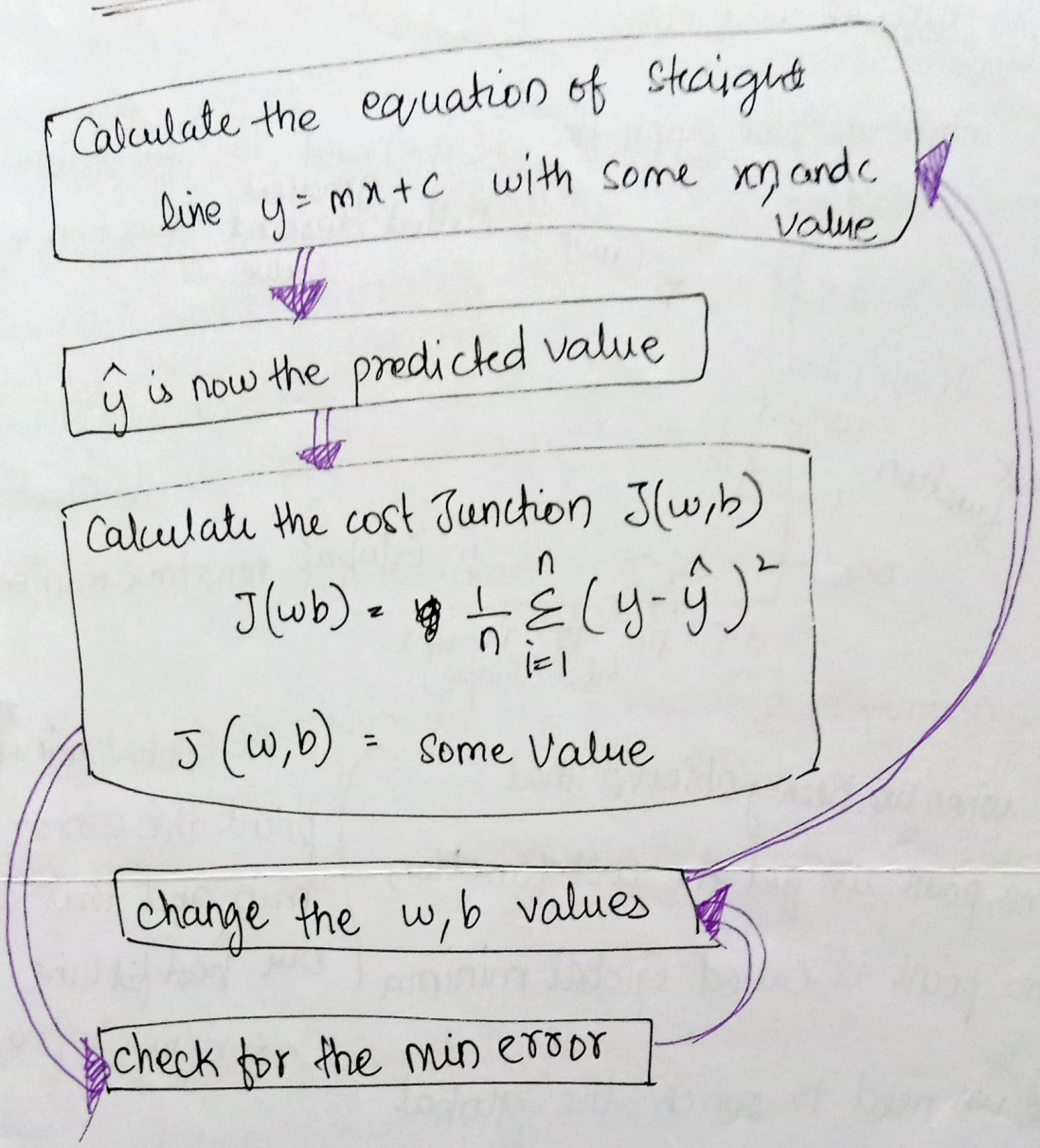
Cost function => $\frac{1}{n}\sum_{i=1}^{n}(y-y')^2$ = [x] one Value

Now again Change M & C Value and Calculate the

$y = mx+C$ ↳ this is y' Now Calculate the

Cost function = $\frac{1}{n}\sum_{i=1}^{n}(y-y')^2$ = [x] some Value

COST FUNCTION IN FLOW DIAGRAM



This is how we calculate the cost function by changing the (w,b) and geting the "best fit line"

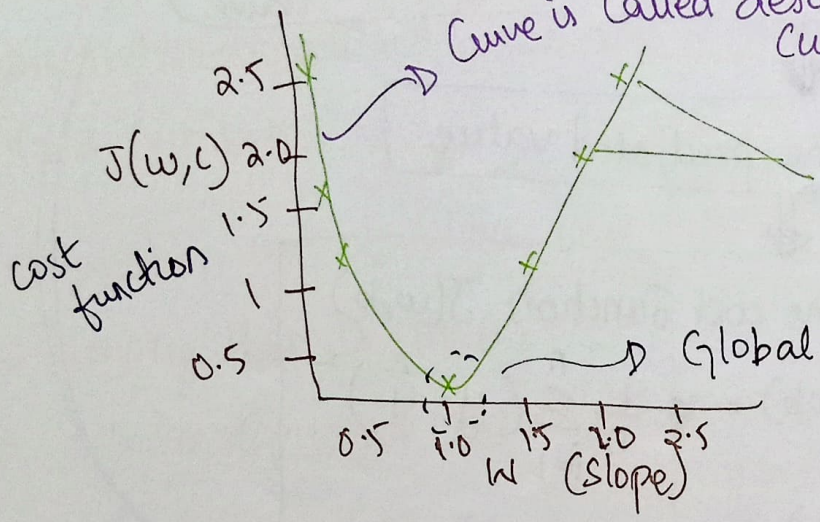
BUT

How do we know that We got the min error. What Technique do We Use for it Should be increase (w,b) or decrease (w,b) values

FRONT

We were Calculating the Cost Function J(w,c) with different w,b or w,c values

when we plot graph for J(w,c) and (w) for various cost function & w

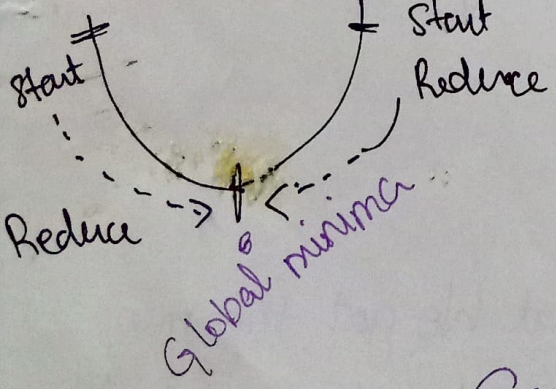


when we keenly observe that one point we get the cost function This point is called Global minima

At Global minima point the error is min and this is our "best fit line" value for (w,c)/(w,b)

Hence we need to reach the global minima point

So we need to start some point and Reduce Value and Reach Global ~~maxima~~ minima



This Global minima is found through a Algorithm Called as

CONVERGENCE ALGORITHM

(OR) GRADIENT DESCENT ALGORITHM

Performance metics: metics = measurement

Performance metics are used to monitor and measure the performance of machine learning model

Performance metics for regression are different and performance metics for classifications are different

Supervised machine learing Algorithm
↳ Regression
↳ classification

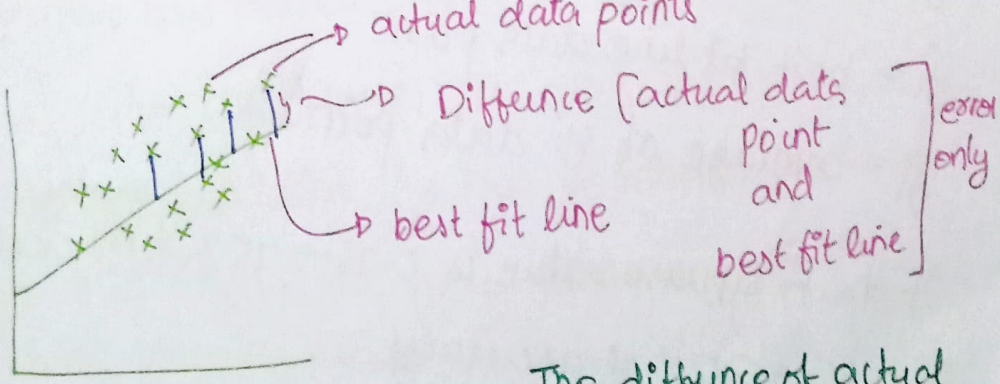
The most used performance metics for linear regression is
① R squared
② Adjusted R squared

R squared: The formula for R squared performance metic is given by

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOTAL}}$$

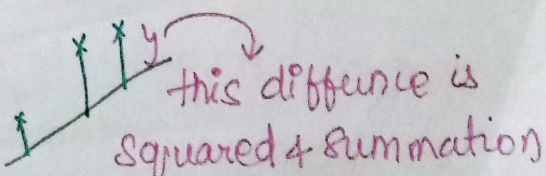
where SS_{RES} = Sum of squared residuals

SS_{TOTAL} = Sum of squared totals



data points are plotted for some dataset and best fit line is produced

The difference of actual data point & best fit line



is done and called as SS_{RES}