Linear Regression :-

⇒ it is a regression Problem.

⇒ output is infinite set. (predicting sal, price etc)

⇒ to apply this there should be linear relationship between data.

⇒ model is nothing but finding a line where it will give minimal Loss (Error)

⇒ Ex:-

heigh, weight, ~~age~~ age.

we need to predict age based on height

if we plot graph on data set then we should predict best fit line to that dataset

$target = w_1 x_1 + w_2 x_2 + c$

y

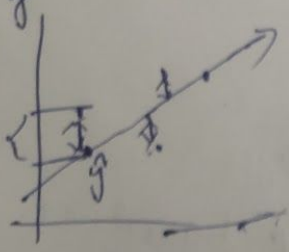
x = height

x_1 = weight

y = age.

⇒ our final goal is to get minimal loss. we will get error because of incorrect weight. so we need to find optimal weights here weight specify relation b/w x and y.

Error



let take

$y = 3x + 5$

if $x = 2$

$y = 11$ if $y = 13$

Error = 2

with this to reduce error we need to find best weights.

To find optimal weights we use gradient descent.

⇒ In gradient descent, we initially take random weights and we will predict values (o/p's) for each datapoint and calculate error.

~~for~~

⇒ if the error is not near to zero or low we will update the weights and calculate the error again

⇒ we repeat this process until weights are not getting update.

how we update weights?

Ex: ~~y = mx + c~~ $y = w_1 x + c$

$$w_{1\,new} = w_{1\,old} - \alpha \frac{\partial L}{\partial w_1}$$

α = Learning rate.

L = loss

$$L = \sum_{i=1}^{n} (y_i - (mx_i + c))^2$$

actual value — predicted value.

why we are squaring?

if one data point Error is 5.
another data point Error is -5.
while summation they will get cancelled.

Intution behind formula gradient descent :-

In calculus:

$$Eq^n = (x^2 - 2x + 2)$$

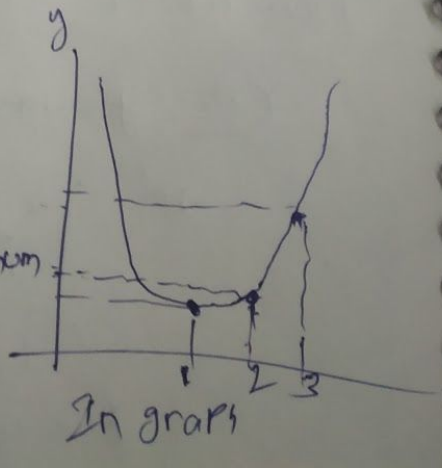
if we want to find x where we will get min for that Eq^n.

we will do $\frac{dE}{dx} = 0$.

$$2x - 2 = 0$$

$$x = 1$$

at $x = 1$ we will get minimum



In graph

let's $x=0$ $\quad 0^2-2(0)+2 = 2$ (not min)

So, if we consider our loss function on y axis and ~~independent variables~~ weights on x axis we need to find point x where we will get min Error.

~~height = w, (age) + C~~.

to find min x for $y = x^2 - 2x + 2$

initially we take random x

let $\alpha = 0.1$

$x = 3$.

$$f'(x) = \frac{dy}{dx} = 2x - 2$$

$f'(3) = 4$ not near to 0.

min.

we need to update x

$$x_{new} = x_{old} - \alpha \frac{dy}{dx}(x=3)$$

$$= 3 - 0.1(4)$$

$$= 3 - 0.4$$

$$= 2.6$$

at $x = 2.6$

$f'(2.6) = 2(2.6) - 2$

$= 5.2 - 2$

$= 3.2$ not near to 0.

update

$x_{new} = x_{old} - \alpha \frac{dy}{dx}$

$= 2.6 - 0.1(3.2)$

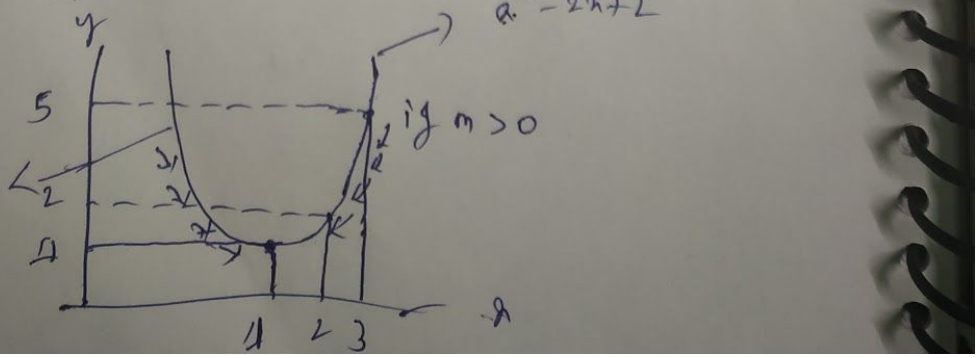
$= 2.28$.

~~as keep on~~

As keep on doing x value is moving to wards 1.

~~As~~ so when m value is +ve it will move left side

when m is negitive it will move right side.



Rule is at minimum ~~de~~ slope is zero

in Eq at $x=1$ $\frac{dy}{dx} = 2x-2$ is 0.

In similar way it will update weights and finds optimal weights where will get minimal Loss.

Why Learning rate α?

if there is no α in ~~the~~ Eqn there is no order in decreament. and it will take more iterations to reach minima.

it will specify jump size.