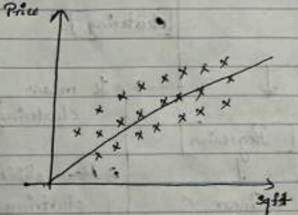
→ Linear regression:-

The main concept in the linear regression is the best-fit line such that all the points exactly fits with the line and our cost func and error will be zero.




$y = mx + c$

$m$ = slope

$c$ = intercept

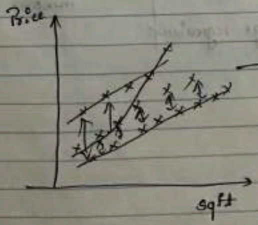
→ when ever our size (or) sq ft = 0

then $y = m(0) + c$

$[y = c]$ → constant/intercept

→ Here we can draw multiple lines for the points such that we need to get the minimum errors b/w the points

By the end when we sum all the errors the value should be minimum.



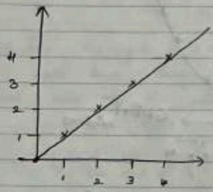

→ Summation of all these errors should be minimal

So the best fit line which gives us minimum error gives us the value of m and c(intercept)

Cost function $= \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}-y)^2$

↳ indicates the real points

m = no of points



when $x=1, y=1$
$x=2, y=2$
$x=3, y=3$
$x=4, y=4$ } So $y=x$

$\hat{y} = mx+c$

$c=0$

$(m=1)$

$\hat{y} = mx = 1(1) = 1$

$\hat{y} = mx = 1(2) = 2$

$\hat{y} = mx = 1(3) = 3$

Here the error $= 0$

Cost function $= \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}-y)^2$

$\Rightarrow \frac{1}{2m}\left((1-1)^2+(2-2)^2+(3-3)^2\right)$

when $x=1, \hat{y}=1$ | when $x=2, \hat{y}=2$

$\Rightarrow \frac{1}{2(3)}(0) = \frac{1}{6}(0) = 0$ → For this we got cost function as zero.

Similarly when we take m value as 0.5 we get

$\hat{y} = mx = 0.5(1) = 0.5$ | $\hat{y} = mx = 0.5(2) = 1$ | $\hat{y} = mx = 0.5(3) = 1.5$ → Cost fn $= \frac{1}{2m}\sum_{i=1}^{m}(\hat{y}-y)^2$

$= \frac{1}{2m}\left((0.5-1)^2+(0.5-2)^2+(0.5-3)^2\right)$

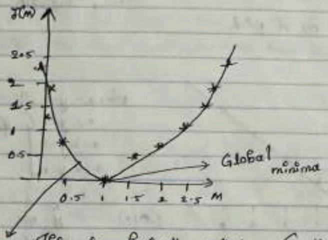
$= \frac{1}{6}\left((0.5)^2+(1.5)^2+(2.5)^2\right)$

$\Rightarrow 0.58$

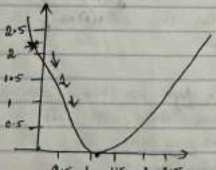
Here by using m values as 1 we got cost fn = 0
when m value as 0.5 we got cost fn = 0.58

Similarly with many m values

we can draw a graph w.r.t ~~cost~~ cost fn & M values




This is basically called as Gradient descent




Here we got a point on y axis by using some x value as our m

Now we need to minimum the point that means it should move to global minima

So for that we use convergence theorem

$$m = m - \left(\frac{\partial m}{dm}\right) * \alpha \rightarrow \text{learning rate}$$

↳ slope

- Here we have a point how can we find slope?

** ⇒ we draw a line and we can see that right side of the point moving down. So it is (-ve) slope

** ⇒ if it is moving up. it is called (+ve) slope.

$m = m - \left(\frac{\partial m}{dm}\right) * \alpha$

| when we have (-ve) slope | when we have (+ve) slope |
|---|---|
| $m = m - (-ve) * \alpha$ <br> $= m(+ve) * \alpha \rightarrow$ we always choose the small value as alpha | $\Rightarrow m = m - (+ve) * \alpha$ <br> $= m + (-ve) * \alpha$ |
| So here we gets (+ve) value it moves down through line and move to the positive axis. | Here it is already positive it moves down and we reach the global minima. |

$\Rightarrow$ By the end of all these when we reach the global minima there the slope we have is the best fit line for the point. till then we need to follow the convergence theorem.

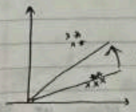
1) Gradient descent (MSE)

1) There is one global minima
2) It is differentiable.

$\Rightarrow$ MSE penalizes the errors

Disadvantage

1) It is not robust to outliers.



$\rightarrow$ moves upwards Because of these outliers why because we are subtracting and squaring the error that is $\left[\frac{1}{2m}\sum_{i=1}^{m}(y-\hat{y})^2\right]$

2. mean absolute error

$\frac{1}{m}\sum_{i=1}^{m}|y-\hat{y}|$

Here there will be no squaring so that the best fit line will be no penalized (or) doesn't get upwards towards the outliers.