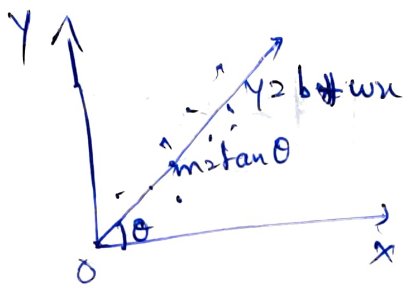
i) Deriving Gradients and to update the params in cost function optimization for simple linear regression.

sol:- For simple Linear Regression,



∴ Assume, cost function of M.S.E of gradients,

$$J(b, w) = \frac{1}{m} \Sigma (y - \bar{y})^2$$

$$= \frac{1}{m} \Sigma (y - (b + wx))^2 \quad [\because \bar{y} = b + wx]$$

Applying partial derivation with respect to 'w'.

$$\frac{d}{dw}(J) = \frac{1}{m} \Sigma \frac{d}{dw}(y - (b + wx))^2$$

derivation formulas, $f(g(x)) = \frac{d(f(x))}{d(g(x))} \times \frac{dy}{dx}$

$\therefore \frac{d}{dx}(x^n) = n x^{n-1}$

$\therefore \frac{d}{dx}(cx) = c$

assume,

$\therefore$ $f(x) = y$ & $g(x) = (b + wx)$

$\therefore \frac{d}{dw}(J) = \frac{1}{m} \sum_{1} 2(y - (b + wx)) * (-x)$

diff w.r.to 'b'

$\Rightarrow \frac{d(J)}{db} = \frac{1}{m} \sum_{1} 2[y - (b + wx)] * (-1)$

updated ~~der~~ gradients/ parameters,

i) $w = w - (\alpha)\left(\frac{dJ}{dw}\right)$

ii) $b = b - (\alpha)\left(\frac{dJ}{db}\right)$ //

ii) what does the sign of Gradient say about the relationship between the parameters and cost function.

Sol) The sign of Gradients will tell about the Inversely proportional (or) Directly proportional between parameters and cost function.

Ex:- If Gradient descent is -ve than, there will be ~~Di~~ Inverse Relation i.e.,

~~the parameters~~
the cost function reduces than parameters ~~increase~~ are increased in indirect proportional.

similarly, when cost function ~~decreases~~ increases then parameters have direct relationship.

iii) why M.S.E took as cost function in Linear regression?

Sol:- while dealing with Linear Regression, we can have multiple lines for different values of slopes and intercepts. But the main question that arises is which of those lines actually represents the right relationship between the X and Y and in order to find that we can use the M.S.E as the parameter. For Linear Regression, this MSE is nothing but the cost Function.

M.S.E is the sum of the squared differences between the prediction and true value. And the output is a single number representing the cost. So the line with minimum cost function or MSE represents the relation between X & Y in the possible manner. And once we have the slope and intercept of the line which gives the least error, we can use that line to predict Y.

iv) what is the effect of Learning Rate on optimization, discuss all the cases.

Sol:- we represent Learning rate as 'α'.

→ 'α' ~~sho~~ states the overfitting and underfitting of the model

→ It is Less time of execution if the value of 'α' could be enhanced perfectly but the computational can be confusionary in case of overfitting/stepping

Cases:-

(i) 'α' is too small (updating parameters are very small)

→ Than it will be time taking and computational steps will be more

→ This is the case of understepping

(ii) 'α' is too large (updates are large)

→ Than it will be time less and computational steps can be confusionary by moving the values.

→ This is the case of overstepping.