Task 6

- Registration/Referral id : SIRSS2212
- Full name : Pranav Rahul Karanjawane

Q1. Calculate/ derive the gradients used to update the parameters in cost function optimization for simple linear regression.

Ans = The equation for simple regression is y= $a1 * x + a0$

we know that cost or error(e) = y ^ $- y$

for n data points:

$$f(a) = \frac{1}{n}\sum_{i=1}^{n} (y^{\wedge} - y)^2$$

$$f(a) = \frac{1}{n}\sum_{i=1}^{n} (y^{\wedge} - (a1 * x + a0))^2$$

α= learning rate or the size of the step we take towards finding the optimal fit line

$\frac{df(a)}{da0}$ *partial derivative of* $f(a)$ *w.r.t* $a0$ *will give the value of parameter* $a0$

$$a0 = \frac{2}{n}\sum_{i=1}^{n} \left(y^{\wedge} - (a1 * x + a0)\right)$$

$\frac{df(a)}{da1}$ *partial derivative of* $f(a)$ *w.r.t* $a1$ *will give the value of parameter* $a1$

$$a1 = \frac{2}{n}\sum_{i=1}^{n} x\left(y^{\wedge} - (a1 * x + a0)\right)$$

New $a0$= $a0 - a0 * α$

New $a1$= $a1 - a1 * α$

Q2. What does the sign of gradient say about the relationship between the parameters and cost function?

Ans = The cost function is a function of the parameters and when the sign is positive then the step will decrease as seen below:

New $a0 = a0 - [+ve\ gradient] * \alpha$

when the sign is negative then the step will increase as seen below:

New $a0 = a0 - [-ve\ gradient] * \alpha$

New $a0 = a0 + [\ gradient] * \alpha$

Q3. Why Mean squared error is taken as the cost function for regression problems.

Ans = MSE or Mean Squared Error is used to check how close predictions made by the model are to actual values. It calculates the error as actual - prediction and squares the difference to eliminate the negative values. The lower the MSE, the closer is prediction to actual. In Regression models, a lower MSE usually indicates a better fit.

Q4. What is the effect of learning rate on optimization, discuss all the cases?

Ans = In an ideal scenario with an optimal learning rate, the cost function value will be minimized rather quickly.

If we take a large learning rate then the cost function value will be minimized very quickly but will settle at a value that is not the lowest.

If we take a lower than optimal learning rate, then even after substantial iterations the cost function will not minimize sufficiently and will take longer time.