

## Task 6

- Registration/Referral id : SIRSS2218
- Full name : Shrikrushnakumar Bhausahab Sondge

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

The equation for simple regression is  $y = a_1 * x + a_0$

we know that cost or error(e) =  $y^{\wedge} - y$

for n data points:

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

$$f(a) = \frac{1}{n} \sum_{i=1}^n (y^{\wedge} - (a_1 * x + a_0))^2$$

$\alpha$  = learning rate or the size of the step we take towards finding the optimal fit line

$\frac{df(a)}{da_0}$  partial derivative of  $f(a)$  w. r. t  $a_0$  will give the value of parameter  $a_0$

$$a_0 = \frac{2}{n} \sum_{i=1}^n (y^{\wedge} - (a_1 * x + a_0))$$

$\frac{df(a)}{da_1}$  partial derivative of  $f(a)$  w. r. t  $a_1$  will give the value of parameter  $a_1$

$$a_1 = \frac{2}{n} \sum_{i=1}^n x(y^{\wedge} - (a_1 * x + a_0))$$

New  $a_0 = a_0 - a_0 * \alpha$

New  $a_1 = a_1 - a_1 * \alpha$

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

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

$$\text{New } a_0 = a_0 - [+ve \text{ gradient}] * \alpha$$

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

$$\text{New } a_0 = a_0 - [-ve \text{ gradient}] * \alpha$$

$$\text{New } a_0 = a_0 + [\text{gradient}] * \alpha$$

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

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?**

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.