**Task-6**

**Name: Trivedula Dhinakar swamy**

**Reg ID: SIRSS1193**

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

The gradients used to update the parameters in cost function optimization for simple linear regression

The equation for simple regression is y= 𝑎1 \* 𝑥 + 𝑎0 we know that cost or error(e) = 𝑦^ − 𝑦 for n data points:

𝑓2

𝑛

𝑓2

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

𝑑𝑓(𝑎)

𝑝𝑎𝑟𝑡𝑖𝑎𝑙 𝑑𝑒𝑟𝑖𝑣𝑎𝑡𝑖𝑣𝑒 𝑜𝑓 𝑓(𝑎) 𝑤. 𝑟. 𝑡 𝑎0 𝑤𝑖𝑙𝑙 𝑔𝑖𝑣𝑒 𝑡ℎ𝑒 𝑣𝑎𝑙𝑢𝑒 𝑜𝑓 𝑝𝑎𝑟𝑎𝑚𝑒𝑡𝑒𝑟 𝑎0 𝑑𝑎0

𝑎

𝑛

𝑑𝑓(𝑎)

𝑝𝑎𝑟𝑡𝑖𝑎𝑙 𝑑𝑒𝑟𝑖𝑣𝑎𝑡𝑖𝑣𝑒 𝑜𝑓 𝑓(𝑎) 𝑤. 𝑟. 𝑡 𝑎1 𝑤𝑖𝑙𝑙 𝑔𝑖𝑣𝑒 𝑡ℎ𝑒 𝑣𝑎𝑙𝑢𝑒 𝑜𝑓 𝑝𝑎𝑟𝑎𝑚𝑒𝑡𝑒𝑟 𝑎1 𝑑𝑎1

𝑎

𝑛

New 𝑎0= 𝑎0 − 𝑎0 ∗ α

New 𝑎1= 𝑎1 − 𝑎1 ∗ α

**2. 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:

New 𝑎0= 𝑎0 − [+𝑣𝑒 𝑔𝑟𝑎𝑑𝑖𝑒𝑛𝑡] ∗ α when the sign is negative then the step will increase as seen below:

New 𝑎0= 𝑎0 − [−𝑣𝑒 𝑔𝑟𝑎𝑑𝑖𝑒𝑛𝑡] ∗ α

New 𝑎0= 𝑎0 + [ 𝑔𝑟𝑎𝑑𝑖𝑒𝑛𝑡] ∗ α

Every ML model has its own set of parameters that given an input, are modified according to that input until we achieve a testing accuracy to our liking. These parameters were introduced with the sole purpose of being changeable so that the model can adapt to different inputs. So by calculating the change of the cost function with respect to the model parameters, you are checking how well your model is doing. You can use this to further change your parameters with the ultimate scope of minimizing this cost function.

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

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.

This MSE is nothing but the Cost Function. Because Mean Squared Error 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 the minimum cost function or MSE represents the relationship between X and Y in the best 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.

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

One of the main challenges when training deep learning is to balance the quality of the final solution with the training time it needs to get there. Learning rate is the most important hyper-parameter to optimize this balance.You can think of small and large learning rates as having different personalities:

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