**Assignment 6**

**Reg ID:SIRSS1210**

**Name:Rayapati Naga Hema Chandrika Sravani**

**Q1. Calculate/ derive the gradients used to update the parameters in cost function**

**optimization for simple linear regression.**

What is Cost Function?

The primary set-up for learning neural networks is to define a cost function (also known as a loss function) that measures how well the network predicts outputs on the test set. The goal is to then find a set of weights and biases that minimizes the cost. One common function that is often used is the mean squared error, which measures the difference between the actual value of y and the estimated value of y (the prediction). The equation of the below regression line is hθ(x) = θ + θ1x, which has only two parameters: weight (θ1)and bias (θ0).

Hypothesis: hθ(xi)=θ0+θ1xi

Parameters:θ0,θ1

Cost Function:

J(θ0,θ1)=(1/2m)∑i=1m(hθ(xi)−yi)2

Objective:

Image

Update Rules:

Image

Image

**Derivatives:**

Image

Image

**Derivation:**

Image

Image

Image

Image

Image

**2. What does the sign of gradient say about the relationship between the**

**parameters and cost function?**

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?**

For linear regression, 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 to eliminate the negative values. We can also say that it is used to check how close predictions made by the model are to actual values. 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. The lower the MSE, the closer is prediction to actual.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:

1)A small learning rate is cautious. That is, it makes the network adjust slowly and carefully.

2)A large learning rate is impetuous. That is, it adjusts quickly but might be overshooting.