<u>**TASK-6(Regression)**</u>

AIRSS1237
Name : Tejasvani

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

<u>**Solution**</u> : An iterative optimization algorithm to find the minimum of a function(loss function) is known as gradient descent.

Consider the equation of a line y = mx + c the value of m is called the slope.
Let's apply gradient descent to m and c :

- Initially let m = 0 and c = 0. Let L be our learning rate. This controls how much the value of m changes with each step. L could be a small value like 0.0001 for good accuracy.
- Calculate the partial derivative of the loss function with respect to m, and plug in the current values of x, y, m and c in it to obtain the derivative value D.

$$D_m = \frac{1}{n}\sum_{i=0}^{n} 2(y_i - (mx_i + c))(-x_i)$$

$$D_m = \frac{-2}{n}\sum_{i=0}^{n} x_i(y_i - \bar{y}_i)$$

Derivative respect to m

- Dm is the value of the partial derivative with respect to m. Similarly lets find the partial derivative with respect to c, Dc :

$$D_c = \frac{-2}{n}\sum_{i=0}^{n}(y_i - \bar{y}_i)$$

Derivative respect to c

- Now update the current value of m and c using the following equation.

$$m = m - L \times D_m$$

$$c = c - L \times D_c$$

- Repeat the process until loss function minimise or ideally 0 (which means 0 error or 100% accuracy). The value of m and c that will be left becomes optimum values.

<u>Question 2</u>: What does the sign of gradient say about the relationship between the parameters and cost function?

<u>Solution</u> : A function may have one or more stationary points and a local or global minimum or maximum of the function are examples of stationary points.
The gradient can be positive or negative .The sign of the gradient tells if the target function is increasing or decreasing at that point.

- **Positive** : Function is increasing at that point.
- **Negative** : Function is decreasing at that point

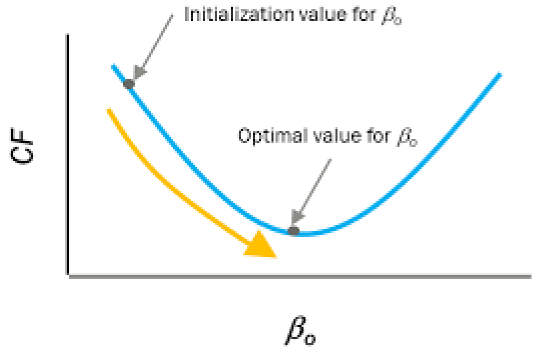
<u>Question 3</u>: Why Mean squared error is taken as the cost function for regression problems.

<u>Solution</u> : 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 &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. The lower the MSE, the better the predictions.

<u>Question 4</u>: What is the effect of learning rate on optimization, discuss all the cases.

<u>Solution</u> : Learning rate is used to scale the magnitude of parameter updates during gradient descent. The choice of the value for learning rate can impact two things:

- How fast the algorithm learns
- Whether the cost function is minimized or not.



It can be seen that for an optimal value of the learning rate, the cost function value is minimized in a few iterations (smaller time). This is represented by the blue line in the figure. If the learning rate used is lower than the optimal value, the number of iterations/epochs required to minimize the cost function is high (takes longer time). This is represented by the green line in the figure. If the learning rate is high, the cost function could saturate at a value higher than the minimum value. This is represented by the red line in the figure. If the learning rate selected is very high, the cost function could continue to increase with iterations/epochs. An optimal learning rate is not easy to find for a given problem. Though getting the right learning is always a challenge, there are some well-researched methods documented to figure out optimal learning rates. Some of these techniques are discussed in the following sections. In all these techniques the fundamental idea is to vary the learning rate dynamically instead of using a constant learning rate.