**Task-6**

**Registration Id: SIRSS2138**

**Name: Akhilkumar K S**

|  |
| --- |

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

Linear regression is a basic and commonly used type of predictive analysis. A linear regression model predicts values by establishing a linear relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula :

y = mx + c,

where y = estimated dependent variable score,

c = constant,

m = regression coefficient, and

x = score on the independent variable.

By minimizing the error ( difference between actual and predicted value ) ,we find the best fit line in the simple linear regression model. For this purpose we use Cost Function which in turn estimates the error and updates the coefficients to minimise the error.For linear regression we use Mean Squared Error ,MSE which is mean squared difference between actual and predicted value.

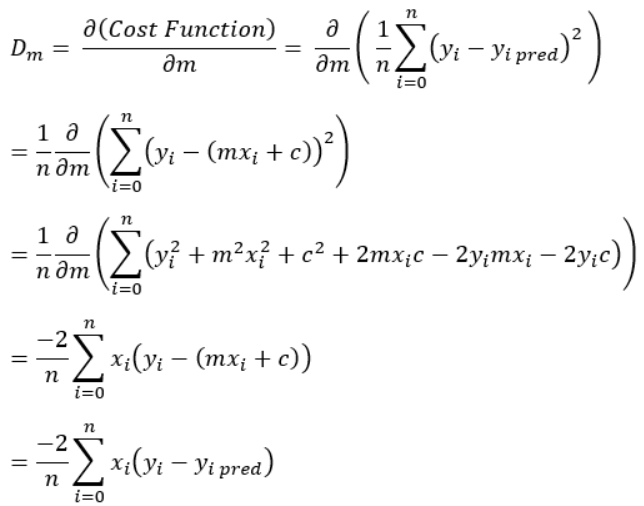
Cost Function (MSE) = (

On replacing with m + c

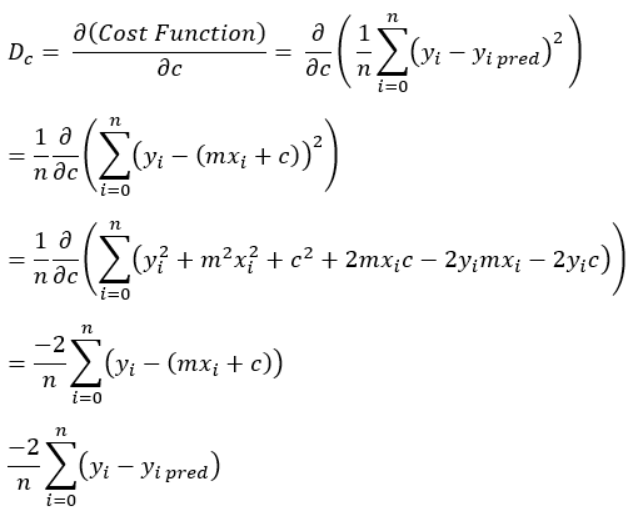
Cost Function (MSE) = (

Gradient is found out by taking partial derivative of cost function with respect to bias(c) and slope

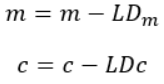
Partial derivative with respect to coefficient or slope (m).



Partial derivative with respect to bias (c)



So these parameters are updated using gradients as follows:



This is repeated until the cost function is very small(approximately 0).

This Gradient Descent gives optimum values to m and c to get a best-fit line.

**2)What does the sign of gradient say about the relationship between the parameters and cost function?**

Gradient descent is an operation to find the local minimum of cost function. So while finding local minima the size of steps taken are related to the sign of gradient.We should take steps which are proportional to negative of gradient at the current point.

For eg;

From earlier derivation we got equation to update coefficient as m = m - L

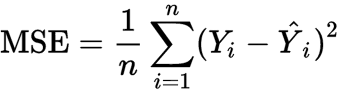
Where is the gradient and L is the learning rate

So when gradient is positive , steep size will increase as follows m=m-L(+)

When gradient is negative , step size will decrease as follows m=m-L(-) = m + L

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

The Mean Squared Error (MSE) is one of the simplest and commonly used cost function.The MSE is given by



The MSE tells us about how close our regression line is close to actual values.It does this by finding the distance between line and other points and squaring them. Squaring helps to remove negative signs.The MSe is great for ensuring that our trained model has no outlier predictions with huge errors , since the MSE puts larger weight on these errors due to squaring part of function.

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

The learning rate is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum if loss function.So this value can affect the rate at which model adapts to the problem and also speed at which model learns.

* For a perfectly selected learning rate , the model will learn to best approximate the function given available resources in a given number of training epochs.
* Larger learning rate allows models to learn faster ,but the model does not arrive at an optimal set of coefficients.
* A smaller learning rate allows the model to learn a more optimal or even globally optimal set of coefficients, but may take longer time to train.
* A learning rate that is too large will result in coefficient updates that will be too large and the performance of the model will oscillate over training epochs. Oscillating performance is said to be caused by weights that diverge.
* A learning rate that is too small may never converge or may get stuck on a suboptimal solution.