Linear regression refers to the machine learning algorithm used to divide the data linearly.

That is in this algorithm machine is trained to find the equation of the line such that most of the points comes under the line

Linear regression can be written in the hypothesis form:



Here θ is also known as the bias

 

Can be rewritten as



In general for data of ‘n’ parameters this equation can be rewritten as



In the above mathematical equation the notation  represents the random value which in neural network term it is known as weight or the importance or impact that parameter to change the output in staring it the random value x defines the value of the features that is the value of the parameter we want to give as input in order to obtain y as output which represents the value of feature we want as our output

Now the task of choosing the right parameters come. So it is important to choose the parameter () such that h(x) is equal to or approaches to y such that



In order to make the value of h(x) is equal to or approaches to y one need to find the error or difference between the value obtained and actual value such that we adjust these value to decrease the error. One such method is known as least square Mean. Represented as:

 Also known as Cost function

On finding error now it is required by us to use it in such a way that our error minimizes that is we get the h(x) approaches to y. In order to obtain such condition we need to minimize the difference between x and y; directly pointing to the condition where our model becomes perfect and gives zero cost.

We can obtain such condition by changing the value of using the value ofthus minimizing the error. In order to do so we have a mathematical approaches known as differentiation.



Where 

 = Learning rate

But before understanding the approach let us discuss about learning rate () one of the component required for understanding this mathematical approach. Now, the learning rate simply refers to a real value which defines the rate at which we want our model to be trained. It is important to note that large learning rate can miss some values and small one can even take lots of time so it is important to choose the correct leaning rate in order to make our model perfect and accurate. But for now let’s assume this learning rate as an random fixed value which further can be changed if one wants to.

Using the values obtained from above, we simply differentiate the value of error  with respect to the change in its weight () and multiplying the change with a small value known as learning rate () and at last incrementing the weight (). This process continues till we reach our desired condition.

This whole process of updating the value of weight () with respect to all other components is known as Gradient Descent, which in general can be defined as an optimization technique, used to get such value of a variable, that our loss becomes least.

In general

1. we have taken an random value of  at first to obtain the h(x) value
2. then we have used the obtained value i.e. value of h(x) to calculate the value of  which tells the the error or how wrong obtained value is from actual value
3. Now we used the value to change the value of  by using the formula 
4. This will change the value of
5. Now we will use this value again and repeat the process till we get value of h(x) approaches to y.

Random ;  refers to parameter