Math behind Linear Regression

Linear Regression is the method from statistics to describe the relationships between independent variables(inputs) and dependent variable(output). The data points of these variables are plotted in graph and best fit straight line is drawn through these data points. The best fit straight line is then used to predict the result for new data points (unknown value) of independent variable.

But the main part is “How to draw the best fit straight line?”

The equation of a straight line is: $y=mx + c$

Where, y= output value , x= input value

m = slope of a line

c =intercept (where line crosses y – axis)

How to find the value of ‘m’ and ‘c’? and If we calculate these values how to know the line drawn is best fit?

To find the best fit straight line, we need to find the error between our actual values and predicted values. i.e

$$\text{error} = y_i - \hat{y}_i$$

There are thousand of such points and we need to calculate error for all those points. To measure total error, we use ‘Mean Squared Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

When do we use Mean Squared Error and Mean Absolute Error?

Mean Squared Error is generally used when outliers are exponentially rare, and Mean Absolute Error is used when outliers are high in number.

So now, how do we calculate the values of ‘m’ and ‘c’ then?

In Machine Learning, Gradient Descent is the most popular method to adjust the weights, so we will use this method to find the values for ‘m’ and ‘c’.

First, the values are randomly assigned , the predicted value is computed for all the points and then error is calculated. Then the gradient is calculated (It is the direction of steepest increase of error). It is calculated by doing the partial derivative of MSE with our parameters ‘m’ and ‘c’.

Taking partial derivative of MSE wrt m,

$$\frac{\partial MSE}{\partial m} = -\frac{2}{N}\sum_{i=1}^{N} x_i \left(y_i - (mx_i + c)\right)$$

Taking partial derivative of MSE wrt c,

$$\frac{\partial MSE}{\partial c} = -\frac{2}{N}\sum_{i=1}^{N} \left(y_i - (mx_i + c)\right)$$

Since the gradients are calculated , now we need to update the weights by gradient descent rule given as:

$$m := m - \alpha \cdot \frac{\partial MSE}{\partial m}$$

$$c := c - \alpha \cdot \frac{\partial MSE}{\partial c}$$

where, α = learning rate which controls the step size.

The values of m and c are calculated. Now, we need to calculate the error and repeat the process until the error is as small as possible or until the error stops changing significantly.

At this point, we have calculated the optimal m and c value, and now we can draw the best fit straight line and can predict the output value for unknown input data(known as test data).

So, The equation of line is: y=mx + c

So,This is the math behind Linear regression and Using this linear regression we can predict the output values for test data in machine learning.