Machine Learning Course
Lecture : Linear Regression

Linear Regression

Linear Regression is the most basic algorithm in Machine Learning. It is a regression algorithm which means that it is useful when we are required to predict continuous values, that is, the output variable 'y' is continuous in nature.

A few examples of the regression problem can be the following-

1. "What is the market value of the house?"
2. "Stock price prediction"
3. "Sales of a shop"
4. "Predicting height of a person"

Terms to be used here:

1. **Features** - These are the independent variables in any dataset represented by x_1 , x_2 , x_3, x_4,… x_n for 'n' features.

2. **Target / Output Variable** - This is the dependent variable whose value depends the independent variable by a relation (given below) and is represented by 'y'.

3. **Function or Hypothesis of Linear Regression** is represented by -
$$y = m_1.x_1 + m_2.x_2 + m_3.x_3 + … + m_n.x_n + b$$
Note: Hypothesis is a function that tries to fit the data.

4. **Intercept** - Here b is the intercept of the line. We usually include this 'b' in the equation of 'm' and take 'x' values for that 'm' to be 1. So modified form of above equation is as follows:
$$y = mx$$
$$\text{Where } mx = m_1.x_1 + m_2.x_2 + m_3.x_3 + … + m_n.x_n + m_{n+1}.x_{n+1}$$
$$\text{Here } m_{n+1} \text{ is b and } x_{n+1} = 1$$

5. **Training Data** - This data contains a set of dependent variables that is 'x' and a set of output variable, 'y'. This data is given to the machine for it to learn or get trained on some function (here the function is the equation given above) so that in future on giving some new values of 'x' **(called testing data)**, our machine is able to predict values of 'y' based on that function.

Linear Regression

Linear regression assumes linear relation between x and y.

The hypothesis function for linear regression is $y = m_1.x_1 + m_2.x_2 + m_3.x_3 + ... + m_n.x_n + b$ where m_1, m_2, m_3 are called the parameters and b is the intercept of the line. This equation shows that the output variable y is linearly dependent on the features x_1, x_2, x_3. The more you are dependent on a particular feature, more will be the value of corresponding m for that feature. We can find out which feature is more important or which feature is more affecting the result by varying the values of m one at a time and see if it is affecting the result, that is , the value of y.

So, here in order to predict the values of y for given features values (x values) we use this equation. But what we are missing here is the values of parameters (m_1, m_2, m_3, ... and b). So, we will be using our training data (where the values of x and y are already given) to find out values of parameters and later on predict the value of y for a set of new values of x.

Linear Regression with one variable or feature:

To understand Linear Regression better and visualize the graphs, let us assume that there is only one feature in the dataset, that is, x. So the equation goes as follows:

$$Y = mX + b$$

Let's say if we scatter the points (x,y) from our training data, then what linear regression tries to do is, it tries to find a line with (given m and b) such that error of each data point (x,yactual) is minimum when compared with x, ypredicted. Error here means combined error.
This line is also called *line of best fit* .

Look at the graph given below :

Sales vs MRP   y = 14.81x + 88.333

Here it is difficult to find out the line of best fit just by looking at the different lines.

So our algorithm finds out the m and b for the line of best fit by calculating the combined error function and minimizing it. There can be three ways of calculating error function:

1. **Sum of residuals (Σ(Yactual - Ypredict))** – it might result in cancelling out of positive and negative errors.
2. **Sum of the absolute value of residuals (Σ | Yactual - Ypredict |)** – absolute value would prevent cancellation of errors
3. **Sum of square of residuals (Σ (Yactual - Ypredict)^2)** – it's the method mostly used in practice since here we penalize higher error value much more as compared to a smaller one, so that there is a significant difference between making big errors and small errors, which makes it easy to differentiate and select the best fit line.

Note : YPredict here is the values of Y predicted by our machine for some m or b.

Further, it is possible to solve classification problem using Linear Regression as well. We will see how this happens in coming sessions. But mostly for classification problems, classifications algorithms give better results.