Linear Regression

Linear Regression is a Machine Learning algorithm that falls under Supervised Learning method and used to determine the value of the output/dependent variable based on the predictor/independent variable/s. Here as the name suggests, the relationship between the dependent and independent variables is assumed to be linear. ¶

Simple Linear Regression: Only one predictor variable is used to predict the values of dependent variable.

Equation of the line: y = c + mX. Where

y : dependent variable

X: predictor variable

m: slope of the line defining relationship between X and y, also called co-efficient of X

c: intercept

The aim of linear regression is to find the best fit line for given X and y variables such that we get optimal values for c and m in the above equation.

Best Fit Line:

We know that scatter plot is used to see how two numeric variables are related to each other and often if there is a linear relationship, we try to fit a line. But we cannot call any line as the Best Fit Line. Our data contains a set of values for dependent variable (denoted by y) and independent variable/s (denoted by X) and a best fit line is the one for which the Residual Sum of Squares of error terms is minimum. What are error terms? Let’s find out. So there will be a ‘y’ value given by our line (we call it y_pred) and a ‘y’ value which is already present in our data (we call it y_true). The difference between y_true and y_pred gives us the error term (e). There will be an error term associated with each X and y value. This error term may be positive or negative. So we take square or all the error terms and sum it. This is called Residual Sum of Squares of error terms. So the aim of linear regression is to find the best fit line for given X and y variables such that the Residual Sum of Squares of errors is minimum. In mathematical terms this RSS is our cost function which we need to minimize. There are several minimization techniques, but the most used is Gradient Descent method

```
Residual Sum of Squares
```

$$RSS = \sum_{i=1}^{n} \left(y_{pred} - y_{avg}\right)2$$

Explained Sum of Squares

$$ESS = \sum_{i=1}^{n} \left(y_{pred} - y_{avg}\right)2$$

Total Sum of Squares

$$TSS = \sum_{i=1}^{n} \left(y_{true} - y_{avg}\right)2$$

Coefficient of Determination (R-Squared)

For the regression line as shown in the figure, the coefficient of determination is measure which tells how much variance in the dependent variable is explained by the independent variable. In short R-squared tells us how good is our model fit, for the given data. The value of R-squared ranges between 0 to 1. A value close to 1 generally means the model is a good fit or more particularly a good amount of variance in the dependent variable is explained by the independent variable/s. So intuitively we can write RSS = ESS/TSS. Therefore R-Squared = (TSS-RSS)/TSS.

R-squared = 1- (RSS/TSS).

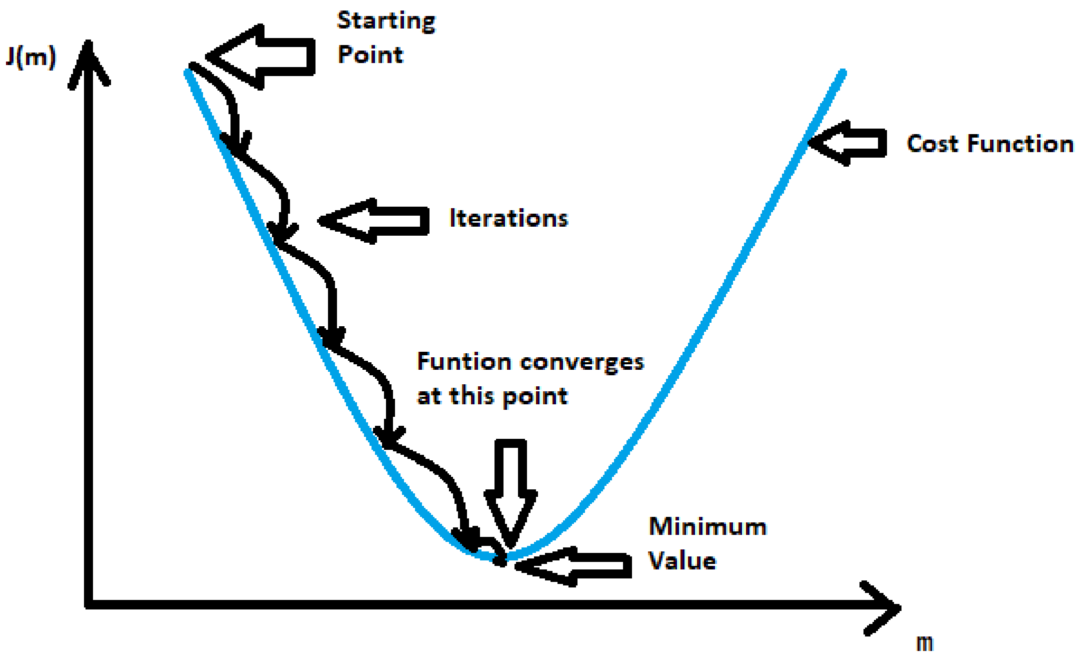
This is an important equation when it comes to linear regression.

In R2 square,it increases ,even if the independent variable is insignificant or significant If the independent features is significant then only Adjust R2 increases

$$Adjusted\ R - squared = 1 - \frac{\left(1 - R_{squared}\right)(N - 1)}{N - P - 1}$$

Where N is number of sample points in our data and P is the number of predictor variables.

Gradient Descent



J(m) is our cost function and we can see that we want to reach a point where the function converges (reaches minimum value).

In other words we want to reach a point where the value of our function is minimum. So we do this in steps, also called iterations.

Learning rate is the rate at which we move in the direction of minimization. Larger the learning rate higher is the chance that we miss the minimum value point. Hence it is wise to keep the learning rate small. This may make the process slow but it is worth it.

m1 = m0 - (learning rate). (dJ/dm)

Where : m1 is the next point given by the iteration m0 : starting point of current iteration We know that our cost function is

RSS = (y_pred-y_avg)^2

Equate y_pred by ‘c+mx’. We get a function where there are 2 unknowns m and c. We take partial derivatives of these functions, once w.r.t to m and once w.r.t c and equate them to 0. We get two equations and 2 unknowns. Solving these two equations we get m and c values. Thus in Gradient Descent, when our function reaches a point where it converges, RSS becomes minimum, we get optimal values of slope(m) and intercept (c). Hence we can say that the regression line y = mx+c is the best fit line.

ASSUMPTION CHECK

1. Linearity Check By using Scatter plot/Paiplot.
2. No Multicollinearity(to check the direction and strength of variables) By using, Correlation Matrix (or), Variance Inflation Factor(VIF) we can check it.
3. No AutoRegression output variable should not vary with time
4. Homoscedasticity For the test of Homoscedasticity, all the input features should be in a similar scale. For that we can apply, any one of the following Data Transformation technique. Standard Scaler - Mean 0, STD = 1 MinMax Scaler - 0 to 1.
5. Test for Normality Histograms/distribution plot QQ plot

Advantages

Linear regression performs exceptionally well for linearly separable data

Easy to implement and train the model

It can handle overfitting using dimensionlity reduction techniques and cross validation and regularization

Disadvantages

Sometimes Lot of Feature Engineering Is required

If the independent features are correlated it may affect performance

It is often quite prone to noise and overfitting

```
In [ ]:
```