

## Linear Regression

Before we go into the specifics of linear regression, you might be wondering why we're looking at this method in the first place.

Isn't it a statistical technique?

Machine learning, more particularly predictive modelling, is primarily concerned with minimising a model's error or producing the most accurate predictions possible at the price of explainability.

## Linear Regression Representation

- Because of the simplicity of its depiction, linear regression is an appealing model. The representation is a linear equation that combines a specific set of input values (x) and yields the anticipated output for that set of input values (y). As a result, both the input (x) and output values are numeric.
- The linear equation provides one scale factor for every input parameter or column, which is denoted by the capital Greek letter Beta (B). One more coefficient is added, which gives the line an additional degree of freedom (e.g., going up and down on a two-dimensional plot) and is frequently referred to as the intercept or the bias coefficient.
- For instance, in a simple linear regression issue (a single x and a single y), the model would have the following form:  $Y = mx + c$
- If we have more than one input (x) in higher dimensions, the line is termed a plane or a hyper-plane. As a result, the representation takes the form of the equation and the precise values used for the coefficients .

## Performance of Regression

\*The performance of the regression model can be evaluated by using various metrics like MAE, MAPE, RMSE, R-squared etc.

### Mean Absolute Error (MAE)

By using MAE, we calculate the average absolute difference between the actual values and the predicted values.

### Mean Absolute Percentage Error (MAPE)

MAPE is defined as the average of the absolute deviation of the predicted value from the actual value. It is the average of the ratio of the absolute difference between actual & predicted values, and actual value.

### Root Mean Square Error (RMSE)

RMSE calculates the square root average of the sum of the squared difference between the actual and the predicted values.

### R-squared values

R-square value depicts the percentage of the variation in the dependent variable explained by the independent variable in the model.

### RSS

Residual sum of squares: It is the measure of the difference between the expected and the actual output. A small RSS indicates a tight fit of the model to the data. It is also defined as follows:

### TSS

Total sum of squares: It is the sum of errors of the data points from the mean of the response variable.

R2 value ranges from 0 to 1. Higher the R-square value better the model. The value of R2 increases if we add more variables to the model irrespective of the variable contributing to the model or not. This is the disadvantage of using R2.

**DONE BY**

***Name - Hari Jagadeesh Iyer***

***Email - harijagsiyer@gmail.com***

In [ ]: