

## Logistic Regression

Logistic regression is a method for fitting a regression curve,  $y = B_0 + B_1x_1 + B_2x_2...$  when  $y$  is a categorical variable.

It is a classification algorithm which comes under nonlinear regression.

We use it to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables.

### Performance of Logistic Regression Model

To test the performance of this model, we must consider few metrics. Irrespective of a tool SAS/R/Python) you would work on, always look for:

#### a. AIC (Akaike Information Criteria)

In logistic regression, AIC is the analogous metric of adjusted  $R^2$ . Thus, we always prefer the model with smallest AIC value.

#### b. Null Deviance and Residual Deviance

Null Deviance—

In this deviance, it shows that the response predicted by a model with nothing but an intercept.

Residual Deviance—

It indicates the response predicted by a model of adding independent variables.

#### c. Confusion Matrix

It is a type of matrix in which we represent a tabular representation of Actual vs Predicted values. Also, this helps us to find the accuracy of the model and avoid overfitting.

Here are some examples of binary classification problems:

**Spam Detection** : Predicting if an email is Spam or not

**Credit Card Fraud** : Predicting if a given credit card transaction is fraud or not

**Health** : Predicting if a given mass of tissue is benign or malignant

**Marketing** : Predicting if a given user will buy an insurance product or not

**Banking** : Predicting if a customer will default on a loan.