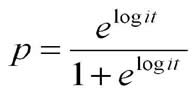
# Logistic Regression:

Allows to establish a relationship between binary outcome and a group of predictor variables of type continuous and/or categorical. Determines the percentage of variance in the dependent variable explained by independent variable.

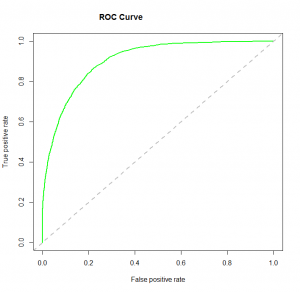
1. Logistic regression is a binary classification model. It’s a linear model which cuts the plane.
2. The impact of predictor (independent) variables is usually explained in terms of odds ratios.
3. A logistic regression model allows us to establish a relationship between a binary outcome variable and a group of predictor variables.
4. The dependent variable in a logistic regression is the log of odds ratio Ln (p/1-p)



1. We measure the accuracy of logistic model using:

* R – Squared value: Is the amount of variance explained by the model. It should be almost 1.0 or 100.
* Sum of Squared error: should be less. Almost zero.
* Roc curve (Receiver Operating Characteristic)

Roc : evaluating the trade-offs between true positive rate (sensitivity) and false positive rate(1- specificity). The area under curve (AUC)Higher the area under curve, better the prediction power of the model. It should be almost 1.

[](https://www.analyticsvidhya.com/wp-content/uploads/2015/11/logit_roc.png)

1. **AIC (Akaic Information Criteria) and BIC (Bayesian Information Criteria): Gives** info about R square value and error term. It is used to compare 2 or more models. Bigger the value better the model is.
2. **Confusion matrix:** build a model on train and test datasets to measure Accuracy, Precision, Recall and F1 score.

**Accuracy**: how often classifier is correct.

**Precision**: when it is predicted ‘YES’ how often it is correct.

**Recall (true positive rate,sensitivity):** when it is actually ‘Yes’ how many times does it predicted ‘yes’

**F1 score**: weighted harmonic mean of precision and recall.

**Support**: number of occurrences of each label.