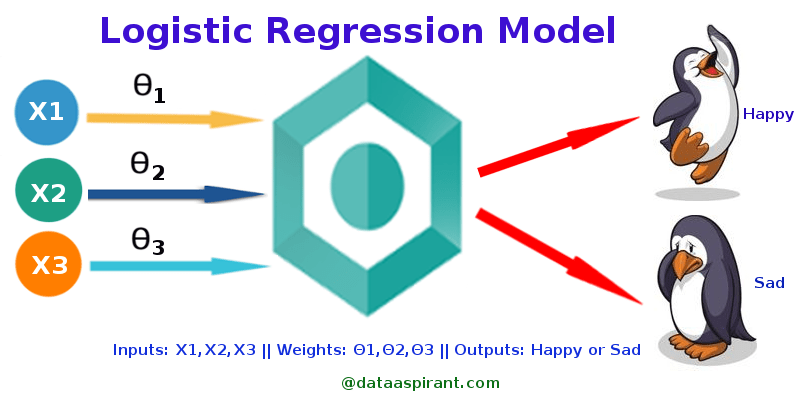
Logistic Regression python



Classification techniques are an essential part of machine learning and data mining applications.

Approximately 70% of problems in Data Science are classification problems.

There are lots of classification problems that are available, but the logistics regression is common and is a useful regression method for solving the binary classification problem.

Another category of classification is Multinomial classification, which handles the issues where multiple classes are present in the target variable.

For example, IRIS dataset a very famous example of multi-class classification. Other examples are classifying article/blog/document category.

Logistic Regression can be used for various classification problems such as spam detection. Diabetes prediction, if a given customer will purchase a particular product or will they churn another competitor, whether the user will click on a given advertisement link or not, and many more examples are in the bucket.

Logistic Regression is one of the most simple and commonly used Machine Learning algorithms for two-class classification.

It is easy to implement and can be used as the baseline for any binary classification problem. Logistic regression describes and estimates the relationship between one dependent binary variable and independent variables.

### Logistic Regression

Logistic regression is a statistical method for predicting binary classes. The outcome or target variable is dichotomous in nature. Dichotomous means there are only two possible classes. For example, it can be used for cancer detection problems. It computes the probability of an event occurrence.

It is a special case of linear regression where the target variable is categorical in nature. It uses a log of odds as the dependent variable. Logistic Regression predicts the probability of occurrence of a binary event utilizing a logit function.

Properties of Logistic Regression:

\* The dependent variable in logistic regression follows Bernoulli Distribution.

\* Estimation is done through maximum likelihood.

\* No R Square, Model fitness is calculated through Concordance, KS-Statistics.

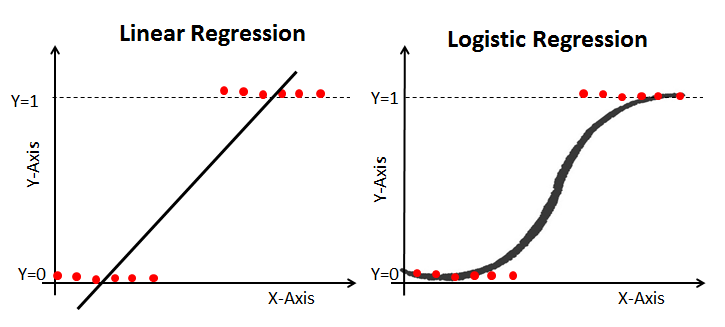
### Linear Regression Vs. Logistic Regression

Linear regression gives you a continuous output, but logistic regression provides a constant output.

An example of the continuous output is house price and stock price.

Example's of the discrete output is predicting whether a patient has cancer or not, predicting whether the customer will churn.

Linear regression is estimated using Ordinary Least Squares (OLS) while logistic regression is estimated using Maximum Likelihood Estimation (MLE) approach.



### Maximum Likelihood Estimation Vs. Oridinary Least Square Method (Least Square Method) MLE vs OLS

The MLE is a "likelihood" maximization method, while OLS is a distance-minimizing approximation method. Maximizing the likelihood function determines the parameters that are most likely to produce the observed data. From a statistical point of view, MLE sets the mean and variance as parameters in determining the specific parametric values for a given model. This set of parameters can be used for predicting the data needed in a normal distribution.

Ordinary Least squares estimates are computed by fitting a regression line on given data points that has the minimum sum of the squared deviations (least square error). Both are used to estimate the parameters of a linear regression model. MLE assumes a joint probability mass function, while OLS doesn't require any stochastic assumptions for minimizing distance.

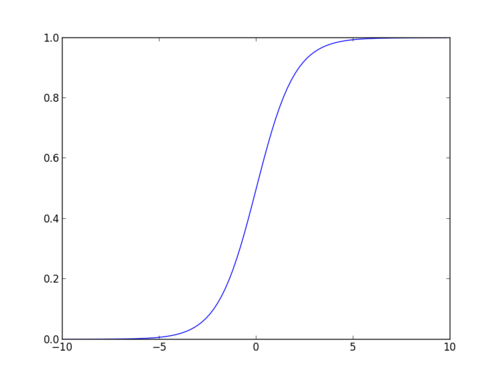
### Sigmoid Function

The sigmoid function, also called logistic function gives an ‘S’ shaped curve that can take any real-valued number and map it into a value between 0 and 1.

If the curve goes to positive infinity, y predicted will become 1, and if the curve goes to negative infinity, y predicted will become 0.

If the output of the sigmoid function is more than 0.5, we can classify the outcome as 1 or YES, and if it is less than 0.5, we can classify it as 0 or NO.

If the output is 0.75, we cannot say in terms of probability: There is a 75 percent chance that patient will suffer from cancer.



### Types of Logistic Regression

Types of Logistic Regression:

\* Binary Logistic Regression: The target variable has only two possible outcomes such as Spam or Not Spam, Cancer or No Cancer.

\* Multinomial Logistic Regression: The target variable has three or more nominal categories without ordering (order is not important ) such a Example: Predicting which food is preferred more (Veg, Non-Veg, Vegan)

\* Ordinal Logistic Regression: the target variable has three or more ordinal categories such as restaurant or product rating from 1 to 5. (order is important)

### Model building in Scikit-learn

Let's build the logistic model.