

Logistic Regression

In this article we are going to discuss Logistic regression, what kind of problem does it solve, what are its limitations.

Logistic regression is an algorithm used to solve classification problem. Classification can be of text data, tabular (structured) data, image classification and some more. Some of the examples of classification problems are, Emails can be spam/non spam, online transactions can be fraud or valid, Tumor malignant or benign. Logistic regression uses labeled data for training and use a sigmoid function to return the probabilities in the end, which can be used to determine the class of a particular test input.

Types of Logistic regression:-

Binary classification (2 classes)

Multi-linear functions failsClass(more than 2 classes)

Logistic Regression

Logistic regression is a machine learning algorithm used for classification problem. It is based on concept of probability and one of the best and simple method for predictive analysis.

Logistic regression is modified version of simple linear regression and can be easily understand if having basic knowledge of Linear regression. Logistic regression uses much complex cost function (called sigmoid function) than linear regression as it produce values between 0 and 1 while linear regression can produce anything between negative infinity to positive infinity which is not possible as per the hypothesis of linear regression.

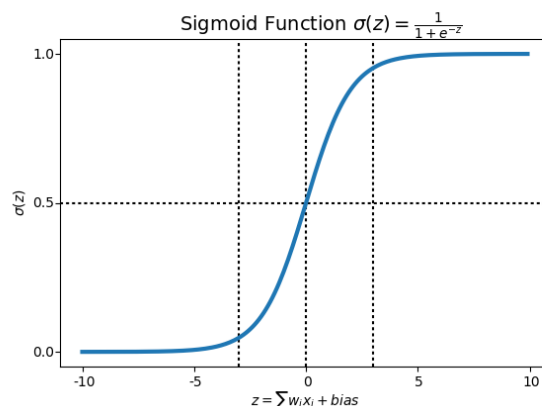
$$0 \leq \hat{y} \leq 1$$

Sigmoid Function

To map predicted value to probability, sigmoid function can be used. Sigmoid function maps any real value to the value between 0 to 1 including both. Sigmoid function is widely used in machine learning for predictive analysis.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Following is the equation and graph plot for sigmoid.



Hypothesis

For linear regression we use following equation for hypothesis.

$$\hat{y} = w^T x + b$$

Where w^T is the transpose of weight matrix and b is bias, for the simplicity we can consider it as equation of line $y=mx+c$.

For logistic regression we modify it as

$$\hat{y} = \sigma(w^T x + b)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

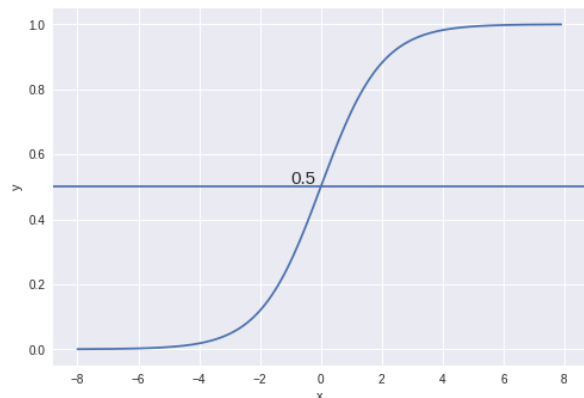
$$\hat{y} = \frac{1}{1 + e^{-(w^T x + b)}}$$

Decision boundary

We expect our classifier to predict a class based on the probability score when a particular testcase is passed through the trained model.

Let's take an example of simple binary classifier for tumor which have two classes malignant or benign(1 for malignant, 0 for benign). Now we decide a threshold value which will divide the two classes.

In this image, threshold is chosen as 0.5, so if the predicted value is 0.7 then it is considered in class 1 (malignant) and if the value is 0.3 it will fall in second class, class 0 (benign).



Cost function

The objective of cost function is to be minimized, so that model we prepared will have better accuracy and minimum total error.

In linear regression we use sum of squared error as our cost function but that cannot be used for logistic regression. If we try to use linear regression cost function for logistic regression, it will end up creating a non-convex function with many local optima (minima and maxima) and it will be very difficult to minimize cost function to get global minima.

So, the loss function for logistic regression can be defined as

$$L(\hat{y}, y) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$$

Intuition is,

$$\text{If } y=1; L(\hat{y}, y) = -\log(\hat{y})$$

We want $\log(\hat{y})$ large, want \hat{y} to large, which can't be greater than 1, because of sigmoid.

$$\text{If } y=0; L(\hat{y}, y) = -\log(1 - \hat{y})$$

We want $\log(1 - \hat{y})$ large, want \hat{y} to small, which can't be smaller than 0, because of sigmoid.

So, the cost function for logistic regression will become,

$$\text{cost}(\hat{y}, y) = \begin{cases} -\log(\hat{y}), & y = 1 \\ -\log(1 - \hat{y}), & y = 0 \end{cases}$$

It can be written as,

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m L(\hat{y}^i, y^i)$$

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^m (y^i \log(\hat{y}^i) + (1 - y^i) \log(1 - \hat{y}^i))$$

Gradient descent

Gradient descent is an algorithm which is used to minimize any function. In this case we will minimize the cost function using gradient descent.

Gradient descent is first order iterative optimization algorithm

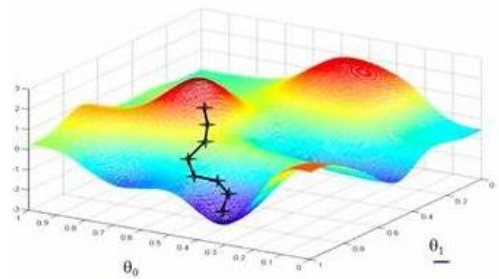
Using gradient descent we change weight and bias such a way that we move in direction so the error is minimum (global minima).

$$w = w - \alpha \frac{\partial}{\partial w} J(w)$$

$$b = b - \alpha \frac{\partial}{\partial b} J(b)$$

w is change in weight matrix, b is change in bias and α is learning rate.

An easy analogy can be, imagine a blind person on top of mountain valley, his objective is to move to the bottom. He will do it by sensing the slope of hill. This action is similar to calculating gradient descent and updating the parameters accordingly.



Advantages of Logistic Regression:

It is a widely used technique because it is very efficient, does not require too many computational resources, it's highly interpretable, it doesn't require input features to be scaled, it doesn't require any tuning, it's easy to regularize, and it outputs well-calibrated predicted probabilities.

Logistic regression does work better when you remove attributes that are unrelated to the output variable as well as attributes that are very similar (correlated) to each other. Therefore Feature Engineering plays an important role in regards to the performance of Logistic and also Linear Regression.

Because of its simplicity and the fact that it can be implemented relatively easy and quick, Logistic Regression is also a good baseline that you can use to measure the performance of other more complex Algorithms.

Disadvantages of Logistic regression:

Logistic Regression is also not one of the most powerful algorithms out there and can be easily outperformed by more complex ones. Also, we can't solve non-linear problems with logistic regression since it's decision surface is linear.

Logistic regression will not perform well with independent variables that are not correlated to the target variable and are very similar or correlated to each other.