**All about Logistic regression in one article**

[Gaurav Chauhan](https://medium.com/@gauravc2708?source=post_page-----b0af09cdb8ad--------------------------------)

[Oct 10, 2018·5 min read](https://towardsdatascience.com/logistic-regression-b0af09cdb8ad?source=post_page-----b0af09cdb8ad--------------------------------)



Behind every great leader there was an even greater logistician.

Unlike other algorithms, Logistic Regression is easily misguided by young developers. Maybe because people still think that it is a regression machine learning algorithm.

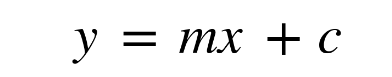
Logistic regression is a statistical machine learning algorithm that classifies the data by considering outcome variables on extreme ends and tries makes a logarithmic line that distinguishes between them.

# Explain like I am five

Logistic regression is a brother of linear regression but unlike its name, logistic regression is a classification algorithm.

Let’s brush up with first linear regression:

**formula:**



where,

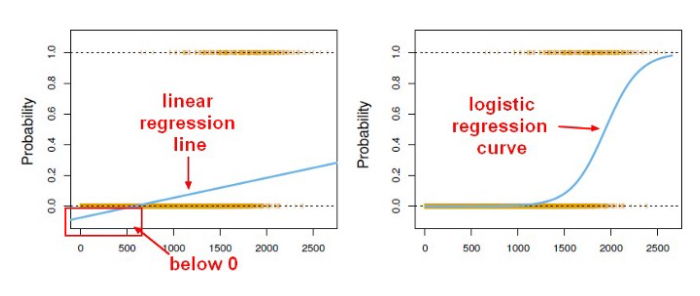
* y = value that has to be predicted
* m = slope of the line
* x = input data
* c = y intercept

With this values, we can predict y values such as.

* here the blue points are the x values (the input data).
* now using the input data, we can calculate slope and y coordinate such that our predicted line (red line)should cover most of the points.
* now using this line we can predict any values of y given its x values.

Now one thing to note from linear regression that it works with continuous data, But if we need linear regression for our classification algorithms, we need to further tweak our algorithm a bit.

First we need to define a threshold such that if our predicted value is lower than threshold then it is of class1 otherwise class2.

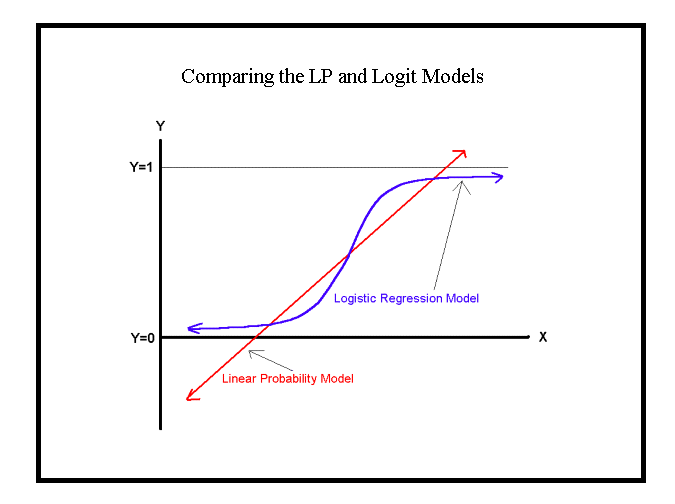


Now if you are thinking “oh that’s easy we have to define linear regression with threshold and vola it becomes classification algorithm,there is a trick in it. We have to define threshold value by ourselves, and for large datasets it will be impossible for us to calculate threshold. Moreover the threshold value once defined will be same even if our predicted values change.

For more reference go [here](https://stats.stackexchange.com/questions/22381/why-not-approach-classification-through-regression).

On the other hand, a logistic regression produces a logistic curve, which is limited to values between 0 and 1.

Logistic regression is similar to a linear regression, but the curve is constructed using the natural logarithm of the “odds” of the target variable, rather than the probability. Moreover, the predictors do not have to be normally distributed or have equal variance in each group.



# [StatQuest: Logistic Regression](https://www.youtube.com/watch?v=yIYKR4sgzI8&t=2s)

# Formula

To explain Logistic regression, i need some physical medium to express my knowledge in this digital medium.

So i have written Logistic Regression formulas in my notebook and then took pictures and posted it here.

If you want pdf version, [click here](https://drive.google.com/file/d/11Lsg-zF3yTZOp3KHBi0bYAeY4F1ayt1r/view?usp=sharing).

# Implementation in python

* Implementation of Logistic regression algorithm **from scratch** in python with explanation in each step is uploaded to my Github repository.
* Implementation of Logistic regression with **help of Scikit learn** is also added to my Github repository.

# Advantages

* One of the simplest machine learning algorithms yet provides great efficiency.
* Variance is low.
* It can also used for feature extraction
* Logistic models can be updated easily with new data using stochastic gradient descent.

# Disadvantages

* Doesn’t handle large number of categorical variables well.
* It requires transformation of non-linear features.
* They are not flexible enough to naturally capture more complex relationships.

# Applications

* Image Segmentation and Categorization
* Geographic Image Processing
* Handwriting recognition
* Spam detection

# When to use

* When we want adjusted odds ratio where we know more than one risk factors.
* When chi-square test is not significant.

Further i will add other machine learning algorithms. The main motto of this article was to give an in depth knowledge of Logistic Regression without using any hard word and explain it from scratch. Further if you want to implement Logistic Regression, start from these datasets and you can comment your predicted score with code in the comments section.

* [Iris dataset](https://archive.ics.uci.edu/ml/datasets/iris)
* [Wine dataset](https://archive.ics.uci.edu/ml/datasets/Wine)
* [Adult dataset](https://archive.ics.uci.edu/ml/datasets/Adult)