

# The math behind Logistic Regression

Khushwant Rai · [Follow](#)

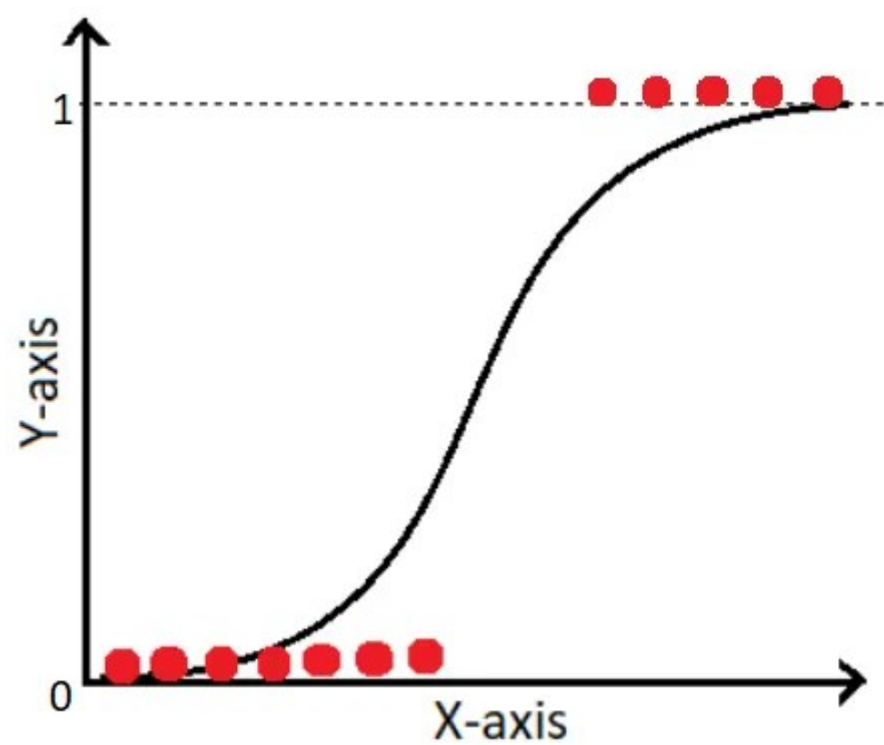
Published in Analytics Vidhya · 4 min read · Jun 14, 2020



144



1



In my last four blogs, I talked about Linear regression, Cost Function, Gradient descent, and some of the ways to assess the performance of Linear Models. Now, in this blog, we will start learning about the classification models and the first one of them is Logistic Regression.

## What is Logistic Regression?

A statistical model typically used to model a binary dependent variable with the help of logistic function. Another name for the logistic function is a sigmoid function and is given by:

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

This function assists the logistic regression model to squeeze the values from  $(-k, k)$  to  $(0, 1)$ . Logistic regression is majorly used for binary classification tasks; however, it can be used for multiclass classification

## why are we calling a classification model the Logistic ‘Regression’?

The reason behind this is that just like Linear Regression, logistic regression starts from a linear equation. However, this equation consists of log-odds which is further passed through a sigmoid function which squeezes the output of the linear equation to a probability between 0 and 1. And, we can decide a decision boundary and use this probability to conduct classification task. For example, let's assume we are predicting whether it is going to rain tomorrow or not based on the given dataset, and if after applying the logistic model, probability comes out to be 90% then we can surely say that it is highly possible to rain tomorrow. On the other hand, if probability comes out to be 10%, we may say that it is not going to rain tomorrow, and this is how we can transform probabilities to binary.

## Maths behind Logistic Regression

We could start by assuming  $p(x)$  be the linear function. However, the problem is that  $p$  is the probability that should vary from 0 to 1 whereas  $p(x)$  is an unbounded linear equation. To address this problem, let us assume,  $\log p(x)$  be a linear function of  $x$  and further, to bound it between a range of  $(0,1)$ , we will use logit transformation. Therefore, we will consider  $\log p(x)/(1-p(x))$ . Next, we will make this function to be linear:

$$\log \frac{p(x)}{1-p(x)} = \alpha_0 + \alpha \cdot x$$

After solving for  $p(x)$ :

$$p(x) = \frac{e^{\alpha_0 + \alpha x}}{e^{\alpha_0 + \alpha x} + 1}$$

To make the logistic regression a linear classifier, we could choose a certain threshold, e.g. 0.5. Now, the misclassification rate can be minimized if we predict  $y=1$  when  $p \geq 0.5$  and  $y=0$  when  $p < 0.5$ . Here, 1 and 0 are the classes.

Since Logistic regression predicts probabilities, we can fit it using likelihood. Therefore, for each training data point  $x$ , the predicted class is  $y$ . Probability of  $y$  is either  $p$  if  $y=1$  or  $1-p$  if  $y=0$ . Now, the likelihood can be written as:

$$L(\alpha_0, \alpha) = \prod_{i=1}^n p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

The multiplication can be transformed into a sum by taking the log:

$$\begin{aligned} l(\alpha_0, \alpha) &= \sum_{i=0}^n y_i \log p(x_i) + (1 - y_i) \log 1 - p(x_i) \\ &= \sum_{i=0}^n \log 1 - p(x_i) + \sum_{i=0}^n y_i \log \frac{p(x_i)}{1 - p(x_i)} \end{aligned}$$

Further, after putting the value of  $p(x)$ :

$$l(\alpha_0, \alpha) = \sum_{i=0}^n -\log 1 + e^{\alpha_0 + \alpha} + \sum_{i=0}^n y_i (\alpha_0 + \alpha \cdot x_i)$$

The next step is to take a maximum of the above likelihood function because in the case of logistic regression gradient ascent is implemented (opposite of gradient descent).

### Maximum Likelihood Estimation (MLE)

A method of estimating the parameters of probability distribution by maximizing a likelihood function, in order to increase the probability of occurring the observed data. We can find MLE by differentiating the above equation with respect to different parameters and setting it to be zero. For example, the derivative with respect to one of the component of parameter alpha i.e.  $\alpha_j$  is given by:

$$\frac{\partial l}{\partial \alpha_j} = \sum_{i=0}^n (y_i - p(x_i; \alpha_0, \alpha)) x_{ij}$$

Hopefully, this post has helped you to comprehend the basic understanding of maths behind logistic regression.

### References

Wood SN. Generalized additive models: an introduction with R. CRC press; 2017 May 18.

#### Logistic regression

In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event...

en.wikipedia.org

### Maximum likelihood estimation

In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a probability...

en.wikipedia.org

### Sigmoid function

A sigmoid function is a mathematical function having a characteristic "S"-shaped curve or sigmoid curve. A common...

en.wikipedia.org

Machine Learning

Logistic Regression

Data Science

Statistics

Data



## Published in Analytics Vidhya

73K Followers · Last published 1 day ago

Follow

Analytics Vidhya is a community of Generative AI and Data Science professionals. We are building the next-gen data science ecosystem <https://www.analyticsvidhya.com>



## Written by Khushwant Rai

58 Followers · 4 Following

Follow

MESc Student at Western University | Machine Learning & NLP Enthusiast.

## Responses (1)



Veres-Lakos Zsombor

What are your thoughts?



Daksh Bhatnagar he/him

Aug 20, 2022

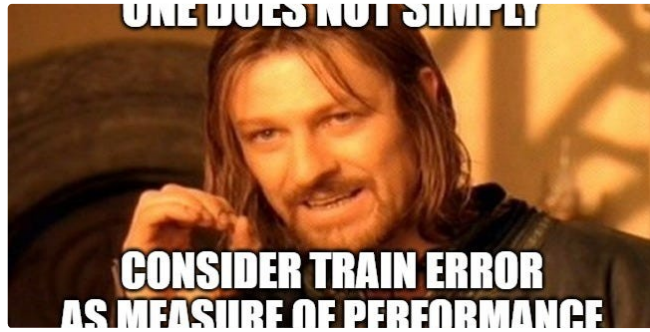


Hey, Would have been nicer if you explained how the formula is working and introduce the notations in the post as well.for example what's A which is a power of e and why is a subscript 0.



[Reply](#)

## More from Khushwant Rai and Analytics Vidhya



 In Analytics Vidhya by Khushwant Rai

### Assessing the Performance (Types and Sources of Error) in Machine...

train error; test error; overfitting; underfitting; bias; variance; noises

Feb 10, 2020  11  1  




 In Analytics Vidhya by Kia Eisinga

### How to create a Python library

Ever wanted to create a Python library, albeit for your team at work or for some open...

Jan 26, 2020  2.8K  33  

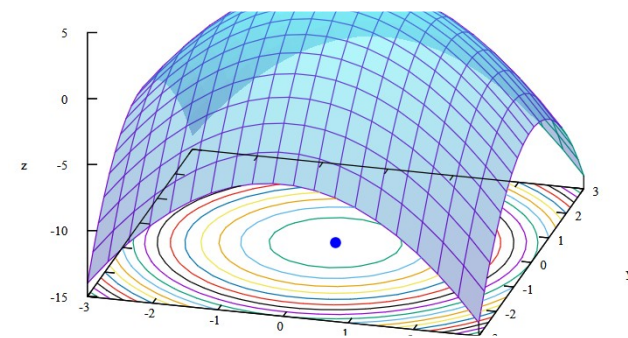


 In Analytics Vidhya by Harikrishnan N B

### Confusion Matrix, Accuracy, Precision, Recall, F1 Score

Binary Classification Metric

Dec 10, 2019  1.2K  6  



 In Analytics Vidhya by Khushwant Rai

### Cost Function, Learning rate, and Gradient Descent in Machine...

Cost Function

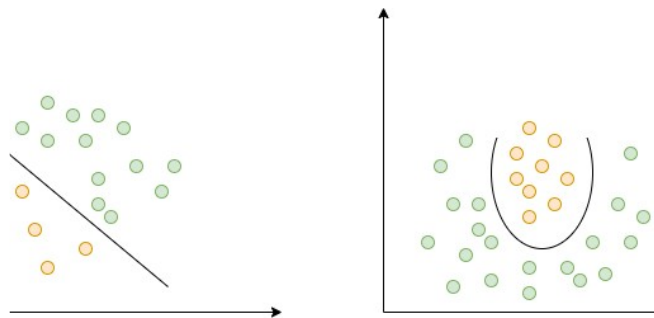
Feb 8, 2020  168  


[See all from Khushwant Rai](#)

[See all from Analytics Vidhya](#)

## Recommended from Medium




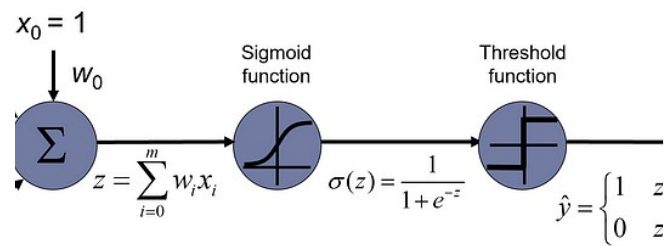



 Okeshakarunaratne

## Understanding Decision Boundaries in Machine Learning

When training a machine learning model for classification tasks, one of the most...




Sep 27, 2024  59



 In TDS Archive by Dr. Roi Yehoshua

## Mastering Logistic Regression

From theory to implementation in Python

 May 20, 2023  527  3



### Lists



#### Predictive Modeling w/ Python

20 stories · 1848 saves



#### Natural Language Processing

1969 stories · 1614 saves



#### Practical Guides to Machine Learning


10 stories · 2217 saves



#### data science and AI


40 stories · 338 saves



 CodexRushi

## Gradient Descent: Explanation with Python Code

Gradient descent is one of the algorithms considered to form the basis of machine...

 Oct 24, 2024



Probability of Success

$$P(y=1 | X) \text{ or } P(X) \text{ or } P = \frac{1}{1 + e^{-(B_0 + B_1 * X)}}$$



where,  
X: independent variable

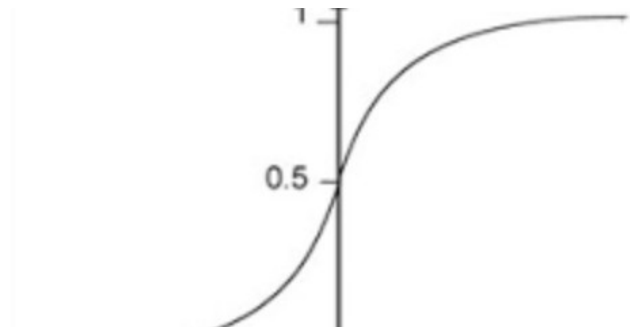
$P(y=1 | X)$ : probability of getting a “1” classification given the set of observed X values.  
Also denoted as  $P(X)$  or  $P$

 Vishwa S. Bhadoria

## Logistic Regression Explained

A Complete Guide for Data Science Beginners [2024]

 Nov 14, 2024  11




 Ajit Kumar Singh

## My Data Scientist — 2 Interview Experience at Zepto

Hey everyone! 🙌

Jan 31  
See more recommendations



 Rishabh Singh

## Logistic Regression

Logistic Regression is one of the most fundamental algorithms in Machine Learnin...

Oct 7, 2024

👏 143

💬 1

