

# Data Science Duniya

Learn Data Science, Machine Learning and Artificial Intelligence

[HOME](#)
[MACHINE LEARNING](#)
[NLP](#)
[DBMS](#)
[WRITE FOR US](#)
[ABOUT](#)
[STARTUP IDEAS](#)
[WEBINARS](#)

## WHAT IS LOGISTIC REGRESSION?

June 17, 2019 Ashutosh Tripathi Logistic Regression, Machine Learning, Regression 2 comments

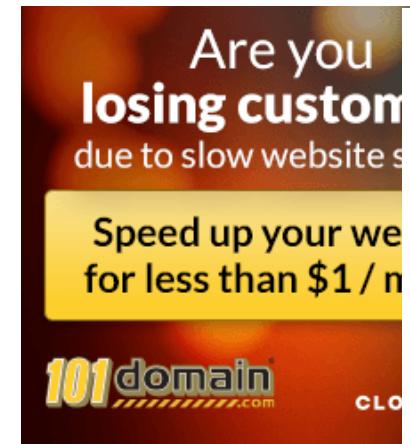
Logistic regression is the most widely used machine learning algorithm for classification problems. In its original form it is used for binary classification problem which has only two classes to predict. However with little extension and some human brain, logistic regression can easily be used for multi class classification problem. In this post I will be explaining about binary classification. I will also explain about the reason behind maximizing log likelihood function.

To understand logistic regression, it is required to have good understanding of [linear regression](#) concepts and it's cost function that is nothing but the minimization of sum of squared errors. I have explained this in detail in my earlier post and I would recommend you to refresh [linear regression](#) before going deep into logistic regression. Assuming you have good understanding of linear regression let's start deep diving to logistic

### BLOG STATS

- 180,947 hits

Advertisements





[SUBSCRIBE TO BLOG VIA EMAIL](#)

Enter your email

Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

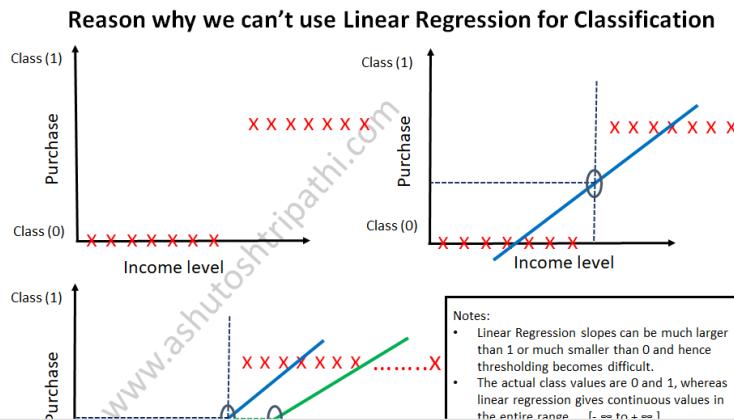
[Close and accept](#)

good foundation for understanding the logistic regression.

## Why can't we use linear regression for classification problems?

Linear regression produces continuous values between  $[-\infty \text{ to } +\infty]$  as output for the prediction problem. So if we have a threshold defined so that we can say that above the threshold it belongs to one class and below the threshold it is another class and in this way we can intuitively say that we can use linear regression to solve a classification problem. However story does not ends here. Question arises, how to set the threshold and what about adding new records won't change the threshold? Threshold we calculate by looking into the best fit line and by adding new record sets, best fit line changes which further changes the threshold value also. Hence we can not confidently say that a particular record belongs to which class as we don't have a certain threshold defined. And this is the main reason we can not directly use linear regression for classification problems.

Below image describes this whole concept with an example.



Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

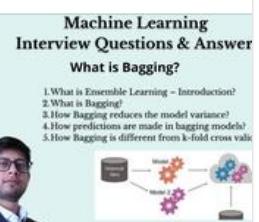
notifications of new posts by email.

Join 5,960 other followers

Enter your email

Subscribe

## FACEBOOK



YOUTUBE.COM

## What is Bagging in Machine Learning?

In this video I have explained what is Bagging in machine learning.

2 Comment



What is the Significance of AUC Curve? How to Interpret ROC AUC Curve?

<http://ashutoshtripathi.co>

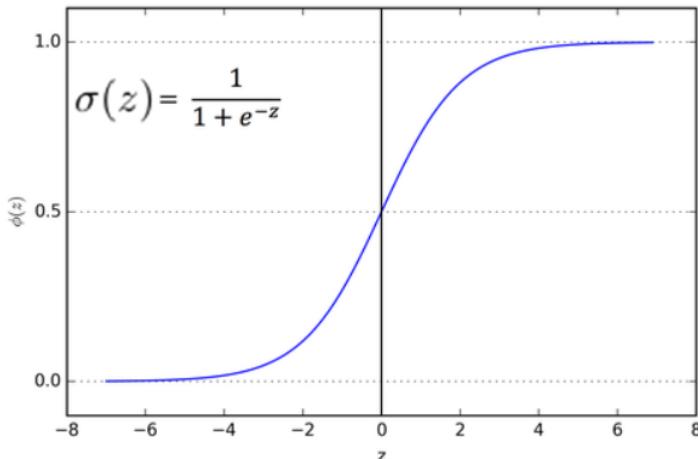
Close and accept

If we extend the concept of linear regression and limit the range of continuous values output  $[-\infty \text{ to } +\infty]$  to  $[0 \text{ to } 1]$  and have function which calculates the probability  $[0 \text{ to } 1]$  of belonging to a particular class then our job will be done. And fortunately **Sigmoid** or **Logistic** function do the job for us. Hence we also say that logistic regression is the transformation of linear regression using Sigmoid function.

## Sigmoid Function

A **Sigmoid function** is a mathematical function having a characteristic "S"-shaped curve or **Sigmoid curve**. Often, Sigmoid function refers to the special case of the [logistic function](#) shown in the first figure and defined by the formula (source: Wikipedia):

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



So Sigmoid function gives us the probability of being into the class 1 or class 0. So generally we take the threshold as .5 and say that if  $p > .5$  then it belongs to class 1 and if  $p < .5$  then it belongs to class 0. However this is not the fixed threshold. This vary based on the business problem. And what threshold value should be,



Implementation of Atomicity and Durability using Shadow Copy



Serial Schedules, Concurrent Schedules and Conflict Operations



Parts of Speech  
Tagging and Dependency  
Parsing using spaCy | NLP | Part 3



Concurrent Execution in Transaction | DBMS



How to Perform Sentence Segmentation or Sentence Tokenization using spaCy | NLP Series | Part 5

# How Logistic Regression works:

As I have already written above that logistic regression uses Sigmoid function to transform linear regression into the logit function. Logit is nothing but log of Odds. And then using log of Odds it calculate the required probability. So let's understand first what is the log of Odds.

## Log of Odds:

Odds ratio is obtained by the probability of an event occurring divided by the probability that it will not occur. and taking the log of Odds ratio will give the log of Odds. So what is the significance log of Odds here.

Logistic function or Sigmoid function can be transformed into an Odds ratio:

$$\text{Odds Ratio } S = \frac{p}{1-p}$$

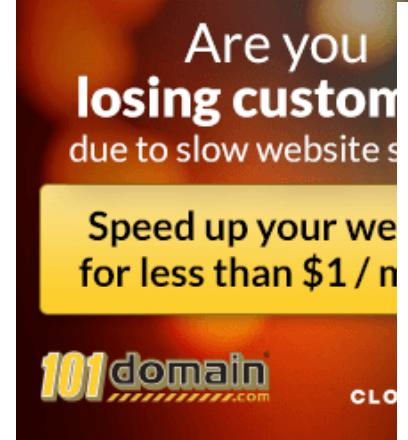
Let's do some examples to understand probability and odds:

- odds  $s = p/q$ ,  $p$  is prob of winning,  $q$  is prob of losing that is  $1-p$ . then if  $s$  is given then prob of winning  $p = \text{numerator}/(\text{numerator} + \text{denominator})$  and prob of losing  $q = \text{denominator}/(\text{numerator} + \text{denominator})$ .

Now let's solve some examples.

- If the probability of winning is  $5/10$  then what are the odds of winning?  $p = 5/10, \Rightarrow q = 1-p \Rightarrow q = 5/10$  hence  $s = p/q \Rightarrow s = 1:1$
- If the odds of winning are  $13:2$ , what is the probability of winning? prob of winning  $p =$

Advertisements



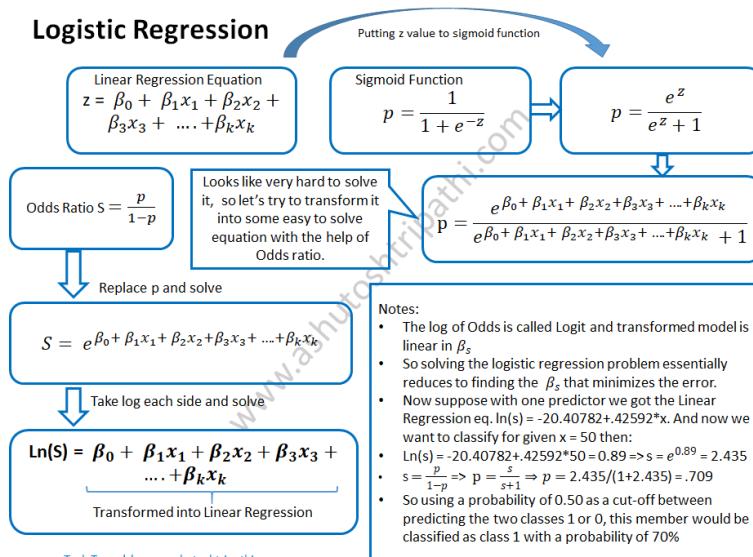
Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

[Close and accept](#)

- + denominator) =>  $q = 8/(3+8) => q = 8/11$
- If the probability of losing  $q$  is  $6/8$  then what are the odds of winning?  $s = p/q, (1-q)/q => s = 2/6$  or  $1/3$ .

## Logistic Model

In the below info graphics I have explained complete working of logistic model.



One more thing to note here is that logistic regression uses maximum likelihood estimation (MLE) instead of least squares method of minimizing the error which is used in linear models.

## Least Squares vs Maximum likelihood Estimation

In Linear regression we minimized SSE.

$$SSE = \sum (y_i - \hat{y}_i)^2$$

Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

[Close and accept](#)

Log likelihood is a convex function and hence finding optimal parameters is easier. Optimal could be max or min and here in case of log likelihood it is max.

$$\text{Log likelihood} = \sum [y_i \ln \hat{y}_i + (1 - y_i) \ln(1 - \hat{y}_i)]$$

Now Lets understand how log likelihood function behaves for two classes 1 and 0 of target variable.

**Case 1:** when Actual target class is 1 then we would like to have predicted target  $y$  hat value as close to 1. let's understand how log likelihood function achieve this.

Putting  $y_i = 1$  will make the second part (after the +) of the equation 0 and only remaining is  $\ln(\hat{y}_i)$ . And  $\hat{y}_i$  will be between 0 to 1.  $\ln(1)$  is 0 and  $\ln(\text{less than } 1)$  will be less than 0 means negative. Hence max value of Log likelihood will be 0 and this will be only when  $\hat{y}_i$  will be as close to 1. So maximizing the log likelihood is equivalent to getting  $\hat{y}_i$  as close to 1 which means it will clearly identify predicted target as 1 which is same as actual target.

**Case 2:** when actual target class is 0 then we would like to have predicted target  $y$  hat as close to 0. let's again understand this how maximizing log likelihood in this case will produce  $\hat{y}_i$  closer to 0.

putting  $y_i = 0$  will make the first part (before + sign) of the equation 0 and only  $(1-y_i)\ln(1-\hat{y}_i)$  will remain.  $1-y_i$  will again be 1 as  $y_i$  is 0 hence after reducing further equation will remain  $\ln(1 - \hat{y}_i)$ . So now again  $1 - \hat{y}_i$  will be less than 1 as  $\hat{y}_i$  will be between 0 to 1. So maximum value of  $\ln(1 - \hat{y}_i)$  can be 0. Means  $1 - \hat{y}_i$  should be close to 1 which

Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

[Close and accept](#)

So that's all about Logistic Regression. In the next article I will be explaining a complete example logistic regression using python.

Video Explanation:



Hope it has given you the good understanding about the concept behind logistic regression.

If you are an aspiring data scientist or an experienced professional who is trying to make his career in Data Science, then you must visit [E-network](#). Where we focus on high-quality interactive [mock interview sessions](#) and help you to QuickStart your Data Science and Machine Learning journey by Preparing a learning roadmap, providing study material, suggesting Best training institutes and provide practice problems with their solutions and many more...

Feel free to [contact us](#) for more details and discussions.

Thank you!

Recommended articles on Regression:

- [Basic Statistics For Data Science Part-1](#)

Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

[Close and accept](#)

- [Feature Selection Techniques in Regression Model](#)
- [What is "stepAIC" in R?](#)
- [What is Multicollinearity in R?](#)

#### Share this:

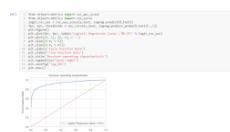


Like

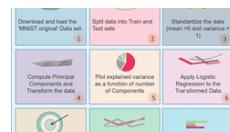


3 bloggers like this.

#### Related



[A Step by Step Guide to Logistic Regression Model Building using Python | Machine learning](#)  
September 26, 2020  
In "Logistic Regression"



[Step by Step Approach to Principal Component Analysis using Python](#)  
July 15, 2019  
In "Data Science"



[What is Linear Regression? Part:1](#)  
January 16, 2019  
In "Linear Regression"

DATA SCIENCE LOGISTIC REGRESSION LOGLIKELIHOOD ODDSRATIO

« [What is Multicollinearity?](#)

[How to Use LinkedIn to Drive](#)

Traffic to Your Blog »

#### 2 COMMENTS

Pingback: [Data Science and Machine Learning Articles | Yearly round-up 2019 – Data Science, Machine Learning & Artificial Intelligence](#)

Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

[Close and accept](#)

## LEAVE A REPLY

Enter your comment here...

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)

### RSS LINKS

- o RSS - Posts
- o RSS - Comments

### FOLLOW US



Privacy & Cookies: This site uses cookies. By continuing to use this website, you agree to their use. To find out more, including how to control cookies, see here: [Cookie Policy](#).

[Close and accept](#)