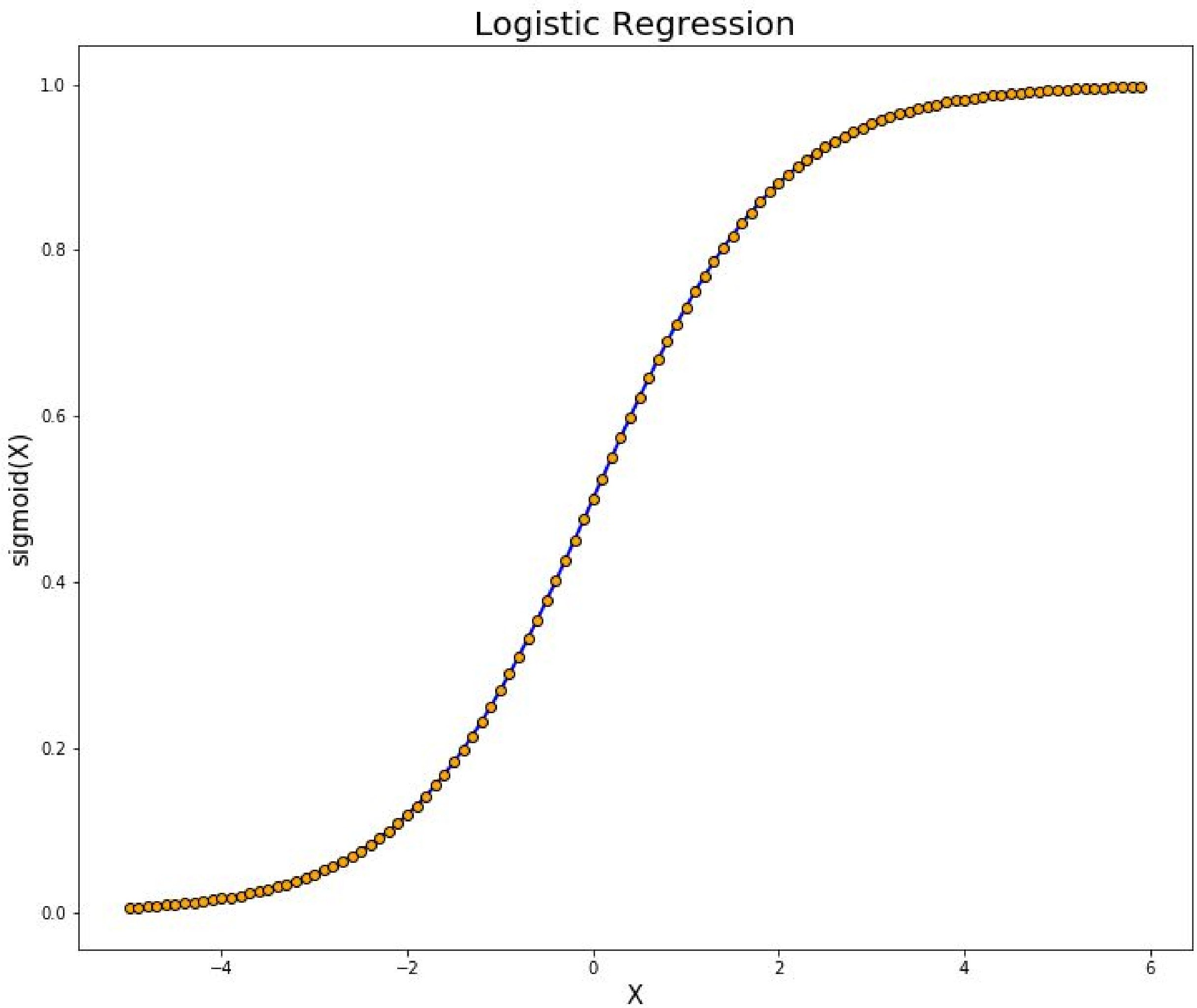
# Logistic Regression




logistic function credits[1]


By:

- Lakshay Goyal: lgoyal@wisc.edu - 9081607617
- Yashishvin Pothuri: pothuri@wisc.edu - 9081569171
- Yatharth Bindal: ybindal@wisc.edu - 9080248553


[1] https://towardsdatascience.com/logistic-regression-explained-9ee73cede081: opened on March 30,2020

# Learning Objectives:

- Introduction to Logistic Regression
- Logistic or Sigmoid Function
- Why is Logistic Regression better than Linear Regression?
- Decision Boundary for Logistic Regression
- Cost Function for Logistic Regression

# Introduction to Logistic Regression

## Definition:

Logistic regression is a statistical model that finds out continuous or probabilistic values and then makes a categorical decision based on whether the value is beyond a certain threshold. Logistic Regression has a wide array of applications in medical predictions and is also widely used in other fields such as engineering, marketing and economics.

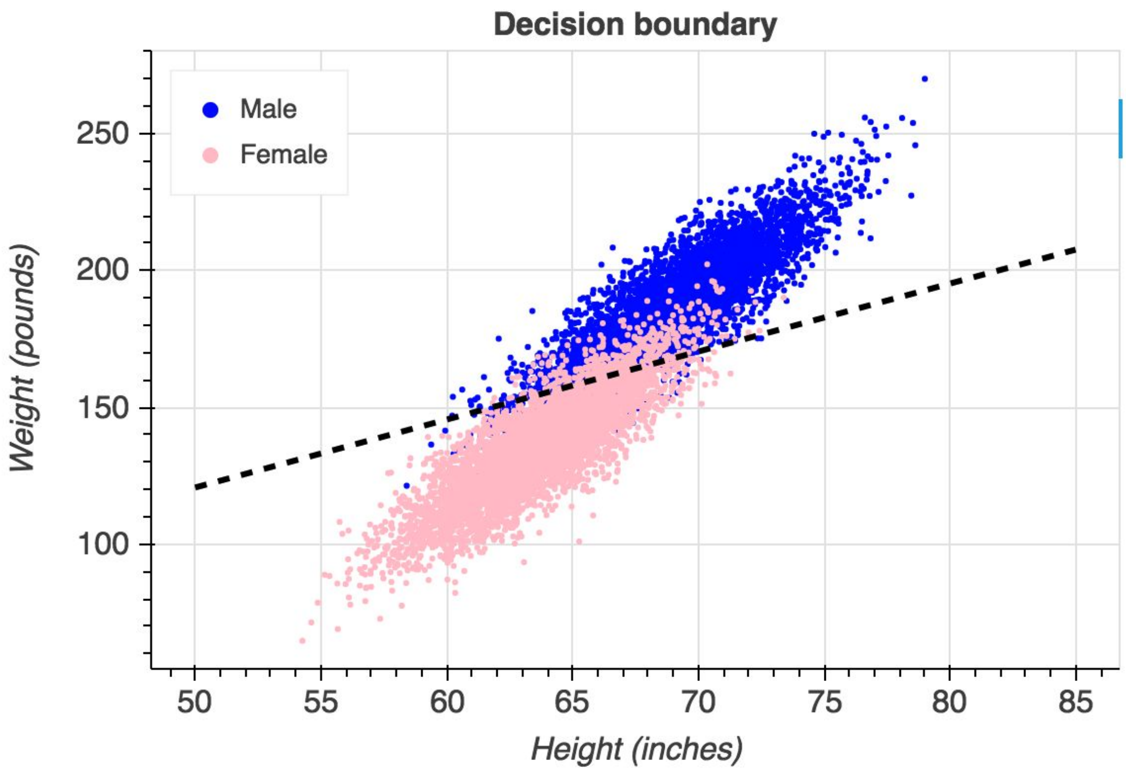
There are three types of Logistic Regression:

- Binomial
    - Where the observed outcome can be of two possible types.
    - For example: yes/no, pass/fail, male/female etc.
- Multinomial
    - Where observed outcomes have three or more possible types.
    - For example: cat/ dog/ sheep/ bird etc.
- Ordinal
    - Where the observed outcomes are ordered
    - For example: low/ medium/ high etc.

For this activity, we will focus on Binomial Logistic Regression.

Let's see a data set where we are given the height and weight of individuals, based on which we have to classify their sex as male or female. We have two possible outcomes in this situation.

We can graph these data points with a scatter plot:




The decision boundary and the data points classified after linear regression[2]

Logistic Regression is used to solve classification problems on all sorts of categorical data.

For example:

- Email: Spam/ Not Spam
- Online Transactions: Fraudulent(Yes/ No)
- Tumor: Malignant/Benign
- Sex: Male/Female etc.

---

[2] https://towardsdatascience.com/understanding-logistic-regression-step-by-step-704a78be7e0a: opened on March 30,2020

## Advantages of Logistic Regression

- It is a widely used technique because it is very efficient, does not require too many computational resources, it is highly interpretable, it doesn’t require input features to be scaled, it doesn’t require any tuning, it is 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 easily and quickly. 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 not among the most powerful algorithms out there and can be easily outperformed by more complex ones.
- Logistic Regression is also ineffective against non-linear problems as its decision surface/boundary 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.

## Ethical Aspect of Logical Regression

The classification is not perfect, when we categorize a data point as A or B there is no true answer or ethically sound answer. Keep in mind it is a prediction based on trends or data we already collected. For example, election forecasting sees the voters profile and then predicts their ballots but it is not necessary that they cast the same vote. Suppose we see the trend of some specific news cycle on voters which are not yet in the new cycle. This gives the candidates points to talk about or they can create false narrative or news cycle.

Similarly these classification algorithms have both ethical sound goals and unethical sound goals.

---

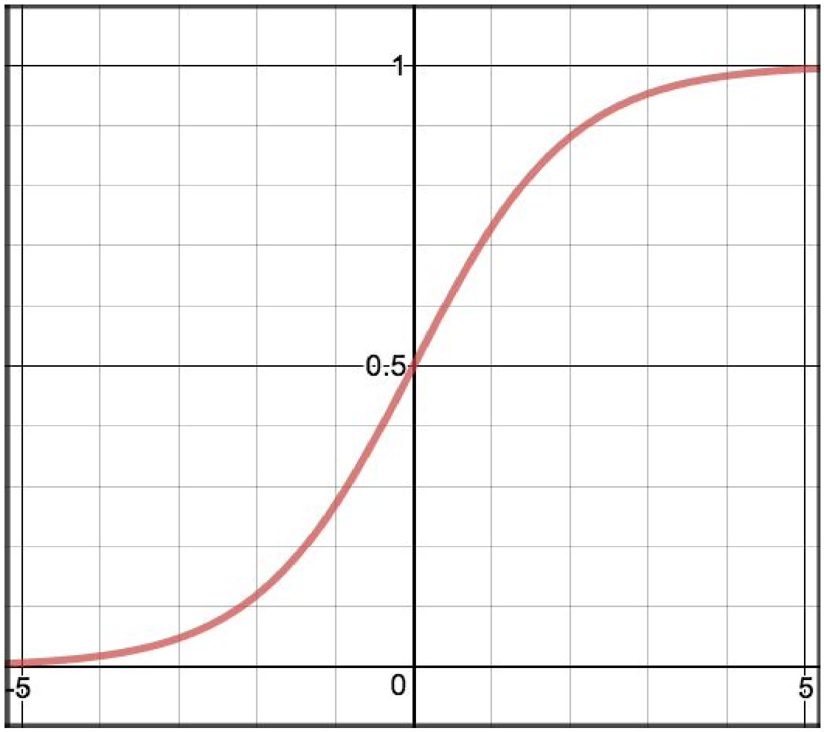
## Logistic Function or Sigmoid Function

In Logistic Regression we use the Sigmoid function in order to map predicted values to probabilities. This function maps any real value to another value between 0 and 1.

Mathematically the Sigmoid function is represented as

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

Graph:




sigmoid function credits[3]

As you can see 0 < *S (z)* <1 from the graph.

## Interpretation of Logistic Function

Suppose we are given *S(z)* = 0.9 in the above case, which means that there is a 90% probability that the individual with the specific height and weight is a male (our data assumes 1 as male and 0 as female).

So if we need to calculate the probability of the individual being a female:

[3] https://ml-cheatsheet.readthedocs.io/en/latest/logistic_regression.html#id21: opened on March 30,2020

P(male|( H,W)) + P(female|( H,W)) = 1 (basic probability theorem)

Where P(male|(H,W)) means probability of the individual being male with height(H) and weight(W).

So,

P(female|(H,W) = 1 - P(male|(H,W))

In this case:

P(female|(H,W)) = 0.1

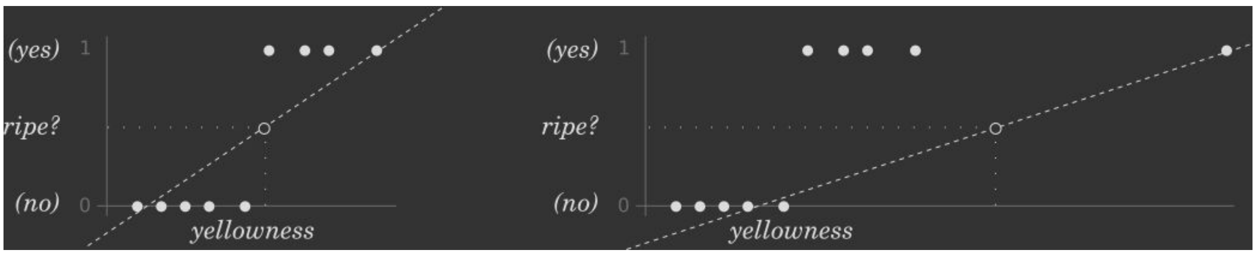
---

## Why is Logistic Regression better than Linear Regression?

Linear regression performs poorly in the classification problem. The main reason is that in classification, unlike regression, one doesn't have to choose the best line through a set of points, but rather you want to somehow separate those points.

To give us a better idea, solving the classification problem using linear regression will result in a decision boundary (as shown in the figure below) which acts as a threshold for the classification for the initial training data. However as we include more training data, the slope of the decision boundary would be altered and would increase our classification error.

Suppose we try to classify a banana as ripe or not ripe based on its color. When there are sufficiently less data sets nearby, linear regression works good to classify them as in figure 1. In order to correctly classify another data set, linear classifier penalizes the decision boundary as in figure 2.




decision boundary penalized by linear classifier[4]

Logistic Regression deals with this classification error through the logistic function which separates the point rather than finding the best curve for the given points.

---

## Decision Boundary for Logistic Regression

Let’s find a decision boundary for logistic regression.

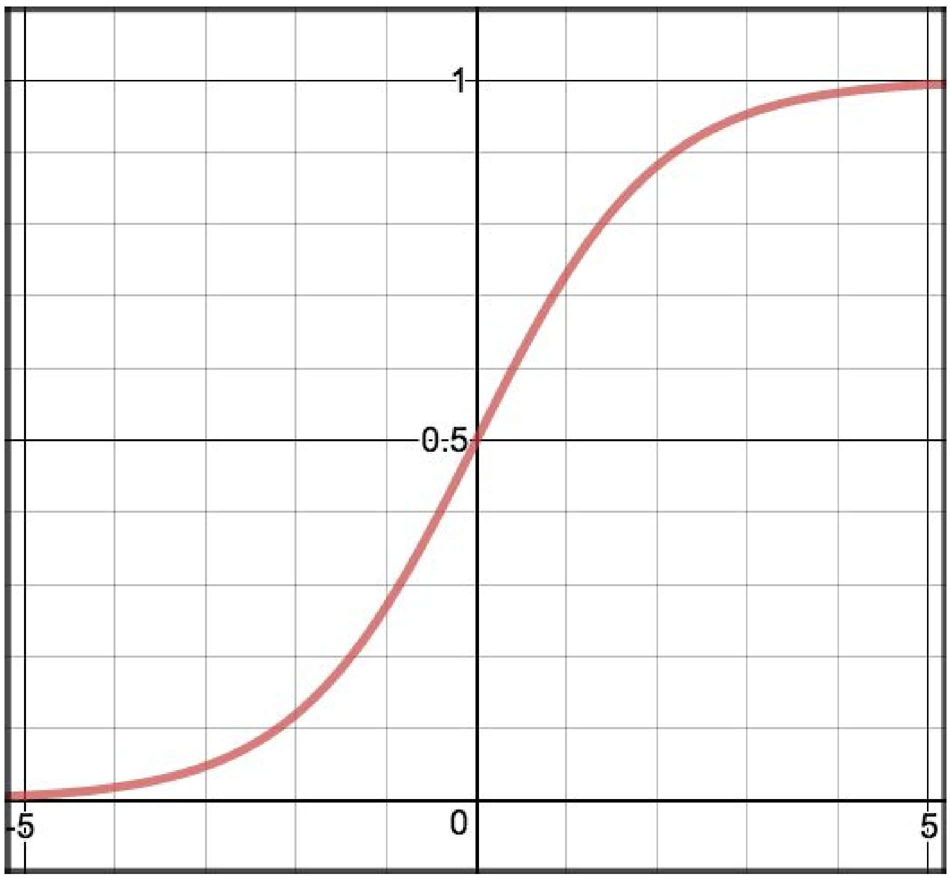
Suppose every *S(z)* >= 0.5 gets rounded off to 1.

And every *S(z)* < 0.5 gets rounded off to 0.

---

[4] :https://www.internalpointers.com/post/introduction-classification-and-logistic-regression: opened on March 30,2020

Now from the graph we can see that for every *S(z)* >= 0.5, z >= 0 which means $X^T$@W >= 0 will be our decision boundary, where X is the feature matrix and W is the weight matrix.




Sigmoid function credits[5]

If we consider our sex classification problem,

Every point which lies on the right of the vertical axis will be male, and every point on the left will be female.

[5] https://ml-cheatsheet.readthedocs.io/en/latest/logistic_regression.html#id21:opened on March 30,2020

Let's take a weights vector as [1,3,-2] and we know the feature vector is [1,H,W].

So by $X^T@W >= 0$,

We get, $1 + 3H - 2W >= 0$

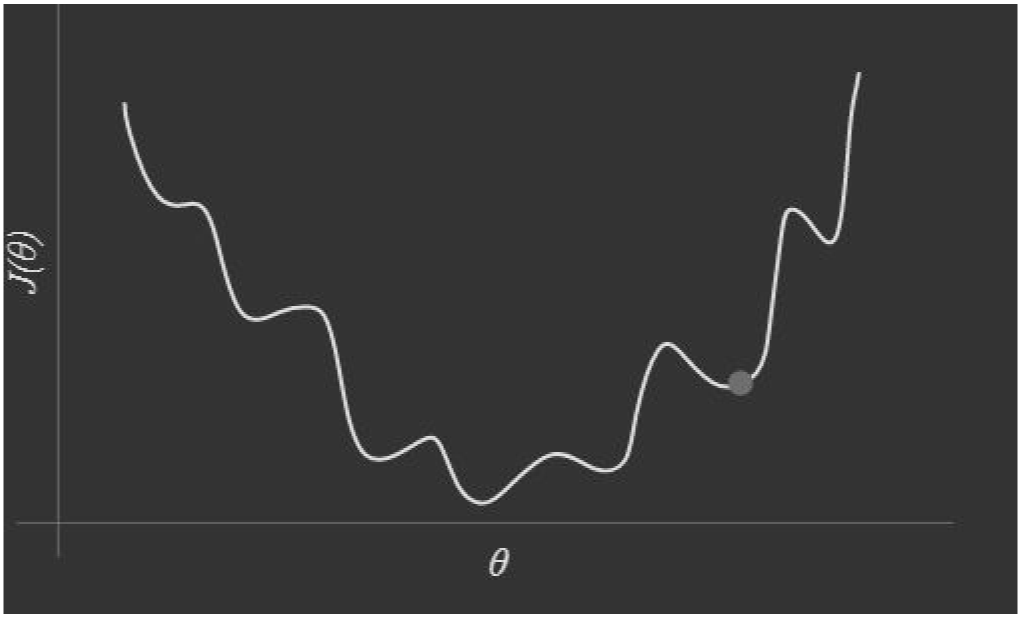
And $3H - 2W = -1$ is the decision boundary.

Where for every $3H - 2W + 1 >= 0$ the individual is male and for $3H - 2W + 1 < 0$ the individual is female.

---

## Cost Function for Logistic Regression

Least Square cost function won't work here as when we replace $X^T@W$ to $S(z)$ in $\min||X^T@W-y||^2$, the curve becomes non convex like in the figure below.




An example for a non convex convex function.
You would end up with a similar cost function for Logistic regression if the least squared cost function is used[6].

where J($\Theta$) is the cost function and $\Theta$ is the weight.

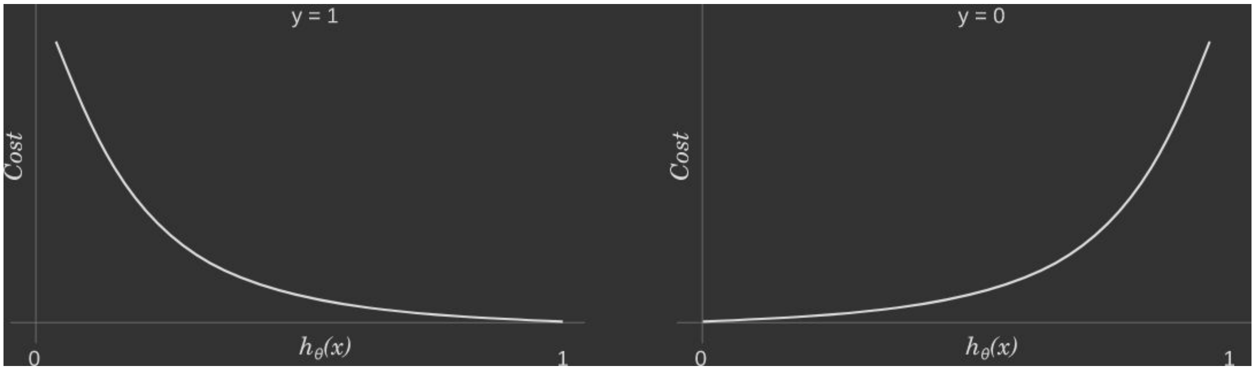
So we define a new cost function for logistic regression which is a convex function.

Cost function(S(z),y) = -log(*S(z)*) if y = 1

= -log(1 - *S(z)*) if y = 0

[6] https://www.internalpointers.com/post/cost-function-logistic-regression: opened on March 30,2020

Graph of the cost function




Cross entropy cost function or log loss[7]

Where $h_{\theta}(x)$ is *S(z)*.

We can write the cost function = -y*log(*S(z)*) - (1-y)*(log(1 -*S(z)*).

As we can see when y = 1, (1 - y) will zero out the second term and when y = 0 it will zero out the first term.

[7] https://www.internalpointers.com/post/cost-function-logistic-regression: opened on March 30,2020

## How to avoid Overfitting?

To solve the problem of overfitting we can use a regularizer in the cost function and then minimize the cost function like we did in L1(LASSO) or L2(Ridge).

To minimize we can use the gradient descent algorithm.

and

We can find the gradient of our cost function with respect to every feature and follow gradient descent algorithm.

But that’s for later.

---

Information Complete: Go to Quiz

---

# References used :


1. https://ml-cheatsheet.readthedocs.io/en/latest/logistic_regression.html#id21: introduction
2. https://www.internalpointers.com/post/introduction-classification-and-logistic-regression: linear vs logistic regression & images
3. https://towardsdatascience.com/understanding-logistic-regression-step-by-step-704a78be7e0a: images and classification example
4. https://towardsdatascience.com/understanding-logistic-regression-9b02c2aec102: difference between logistic and linear regression
5. https://www.coursera.org/learn/machine-learning/home/welcome: Coursera Course used to design the project
6. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html: Python library manual used in logistic regression
7. http://www.stat.columbia.edu/~gelman/research/published/ChanceEthics13.pdf : ethical concerns
8. https://github.com/lakshaygoyal/code_machine_learning_algorithms/blob/master/logistic_regression/logistic_regression.py: python code for logistics regression and data set