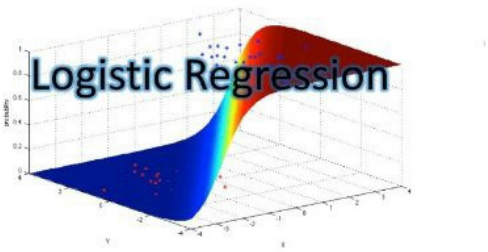
Chapter 2.0 : Logistic Regression with Math.

Madhu Sanjeevi (Mady) in Deep Math Machine learning.ai

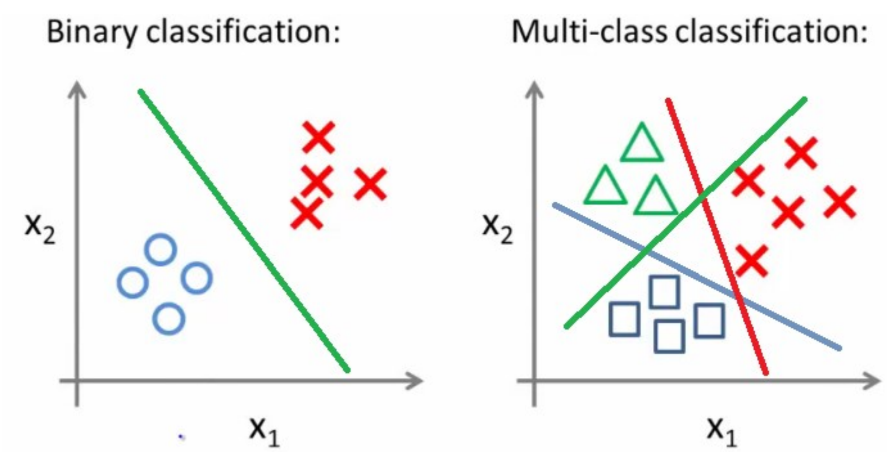
Sep 26, 2017 · 5 min read



In the previous story we talked about **Linear Regression** for solving regression problems in machine learning , This story we will talk about Logistic Regression for classification problems.

You may be wondering why the name says regression if it is a classification algorithm, well,It uses the regression inside to be the classification algorithm.

Classification : Separates the data from one to another.



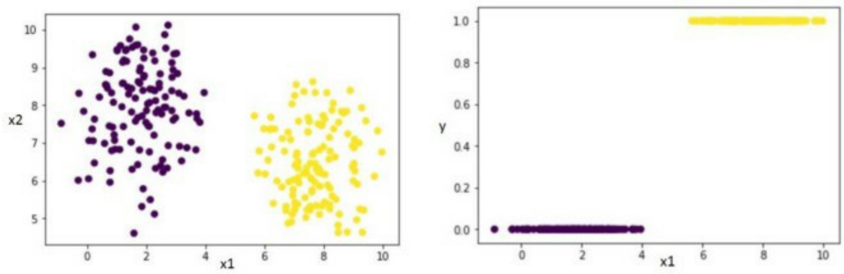
One Vs All method (Right)

This story we talk about **binary classification** (0 or 1) Here target variable is either 0 or 1

Goal is to find that green straight line (which separates the data at best)

so we use regression for drawing the line , makes sense right?

Lets take a random dataset and see how it works,

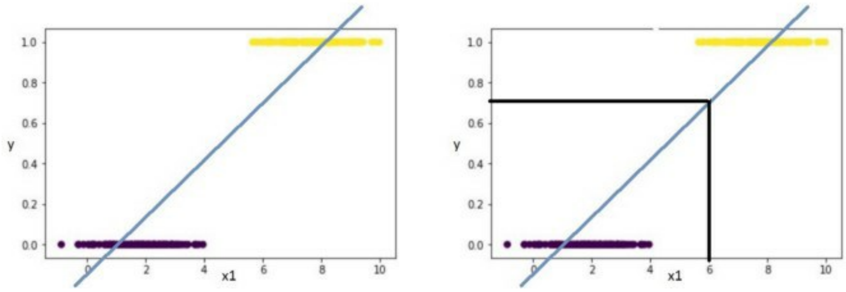


(x1,x2) plot(Left), (x1,y) plot(Right)

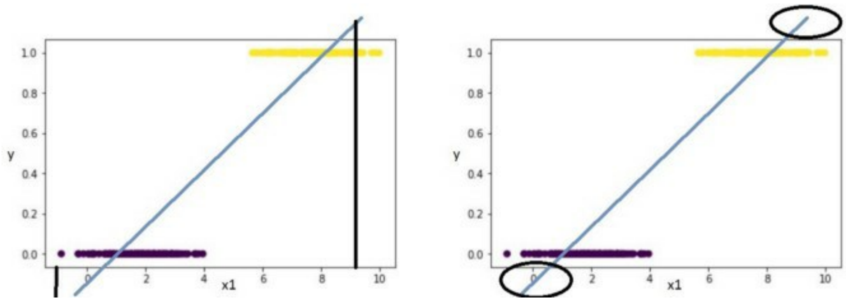
if we observe the *right* picture we have our independent variable (X) and dependent variable(y) so this is the graph we should consider for the classification problem

Given X or (Set of x values) we need to predict whether it's 0 or 1 (Yes/No).

If we apply Linear regression for above data we get something like this,



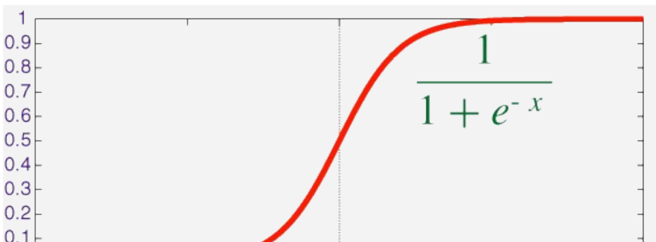
Given X value 6 we can say y is 0.7 (close to 1), that's cool but wait, What if I give negative X value or greater X value??? The output is this



We only accept the values between 0 and 1 (We don't accept other values) to make a decision (Yes/No)

> *so how do we proceed further?*

There is an awesome function called **Sigmoid** or **Logistic function** , we use to get the values between 0 and 1



This function squashes the value (any value) and gives the value between 0 and 1

How??? and what is **'e'** ???

e here is ‘exponential function’ the value is **2.71828**

this is how the value is always between 0 and 1.

$$2.71828^{+x} = \text{Positive value}$$

$$2.71828^{-x} = \frac{1}{\text{Positive value}} = \text{Value between (0 and 1)}$$

$$\boxed{\frac{1}{1 + \text{Positive value}} \quad \text{or} \quad \frac{1}{1 + \frac{1}{\text{Positive value}}}} = \frac{1}{1 + e^{-X}}$$

Sigmoid Function

So far we know that we first apply the linear equation and apply Sigmoid function for the result so we get the value which is between 0 and 1.

The hypothesis for *Linear regression* is **h(X) = θ0+θ1*X**

The hypothesis for this algorithm is

$$\frac{1}{1 + e^{-(\theta 0+\theta 1*X)}}$$

Logistic function for Logistic regression.

> *How does it work??*

1. First we calculate the **Logit function,** what the heck is that??

logit = θ0+θ1*X (hypothesis of linear regression)

2. We apply the above Sigmoid function (Logistic function) to logit.

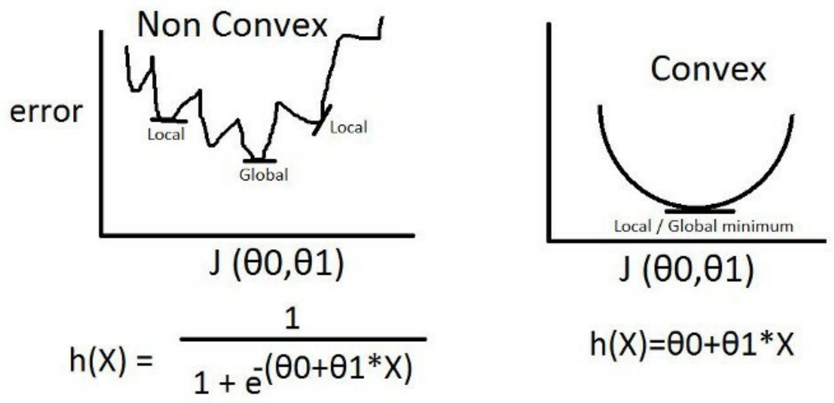
3 we calculate the error , Cost function (Maximum log-Likelihood)

Cost function for linear regression is

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Cost function

here it does not work as h(x) hypothesis gives non convex function for J(**θ0,θ1)** so we are not guaranteed that we reach best minimum.



We take **log(hypothesis)** to calculate the cost function

Logistic regression cost function

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$= -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log (1-h_\theta(x^{(i)}))]$$

$$P(y=1 \mid x;\theta) = h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

Taken from Prof. Andrew Ng.'s Coursera ML course

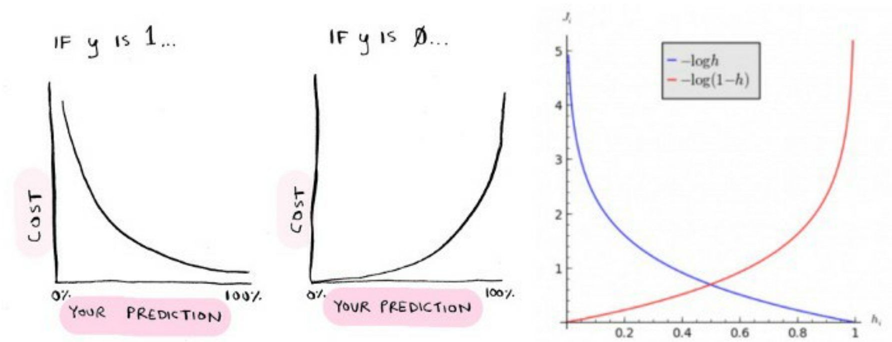
If it does not make sense , let me make it sense to you

usually error is what?? (predicted — actual)**2 right??

```
so if predicted  = 1 and actual= 1
error = 0
so if predicted  = 1 and actual= 0
error = 1
so if predicted  = 0 and actual= 1
error = 1
so if predicted  = 0 and actual= 0
error = 0


Note: predicted can be 0.5 and so on... also
So every time we get the error between 0 and 1 which is not
useful.
```

just take a look at this picture and observe something..



0%-100% = 0–1

From Left picture

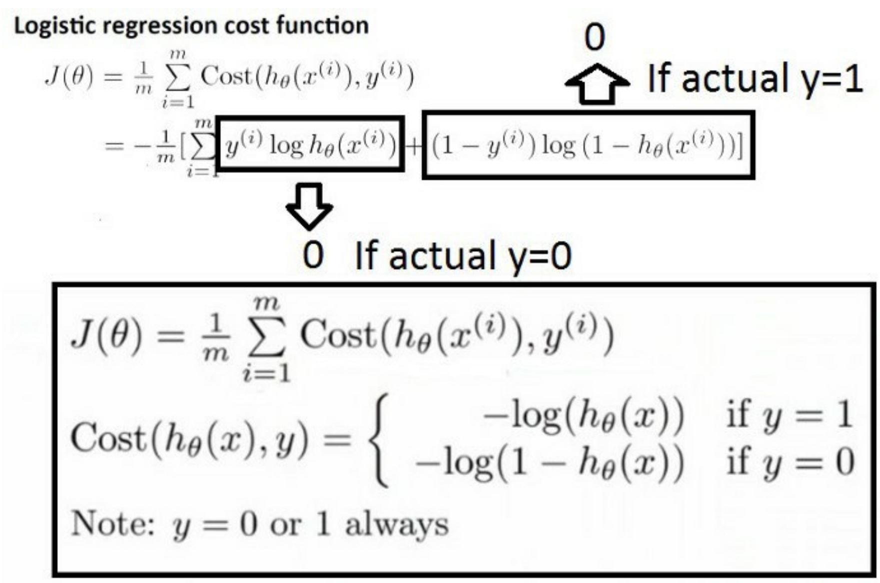
If actual y =1 and predicted =0 the cost goes to infinity and If actual y =1 and predicted =1 the cost goes to minimum.

If actual y =0 and predicted =1 the cost goes to infinity and If actual y =0 and predicted =0 the cost goes to minimum.

From Right picture

if we apply ***log*** to ***hypothesis*** (***predicted***) we get some values (cost) which is useful to estimate the overall error.

Here is the final picture.



that's it. based on the actual **y** values we calculate different functions.

4. Next step is to apply **Gradient descent** to change the **θ** values in our hypothesis (I already covered check this **link**).

That's it We are done!

we got the Logistic regression ready, we can now predict new data with the model we just built.

Predicting new data, remember?? we give new X values we get the predicted y values how does it work ??

$$1 + e^{-(\theta 0+\theta 1*X)}$$

Bam!!!!

we get the probability score(s).

So That's it for this story , In the next story I will code this algorithm from scratch and also using Tensorflow and scikitlearn.

See ya!

Machine Learning · Logistic Regression · Supervised Learning · Classification · Artificial Intelligence

WRITTEN BY

Madhu Sanjeevi (Mady)

Follow

Writes about Technology (AI, ML, DL) | Writes about Human Mind and Computer Mind. interested in ||Programming || Science || Psychology || NeuroScience || Math

Deep Math Machine learning.ai

Following

This is all about machine learning and deep learning (Topics cover Math,Theory and Programming)

Also tagged Logistic Regression

Would I embarrass myself if I went on Netflix's Nailed It?

Marguerite Siboni in The Startup

Jul 7 · 5 min read ★

51

Related reads

Introduction to Logistic Regression

A

Ayush Pant in Towards Data Science

Jan 22 · 5 min read

424