Logistic Regression

Sameer Dhiman

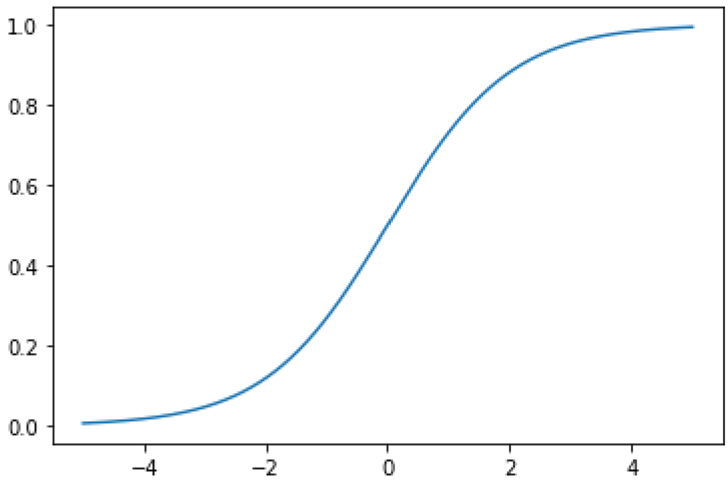
31-07-2020

1 Introduction

In the previous session we saw linear regreesion, in linear regresison we we trying to fit the equation m*x + b to the data. In Logistic Regression we are will fit the function

$$\frac{1}{1 + e^{-x}}$$

This is called the sigmoid function and it looks like this:



At x = 0 $sigmoid(x) = 0.5$, x = infinity $sigmoid(x) = 1$ and x = -infinity $sigmoid(x) = 0$

Now, let's suppose we have a fucntion $f(x)$, how will the following look when compared to $f(x)$:

1. $f(x + 4)$
2. $f(4x)$
3. $f(x/4)$
4. $f(4x + 4)$

Answers:

1. $f(x + 4)$ - shifted towards left

2. $f(4x)$ - compressed along x axis

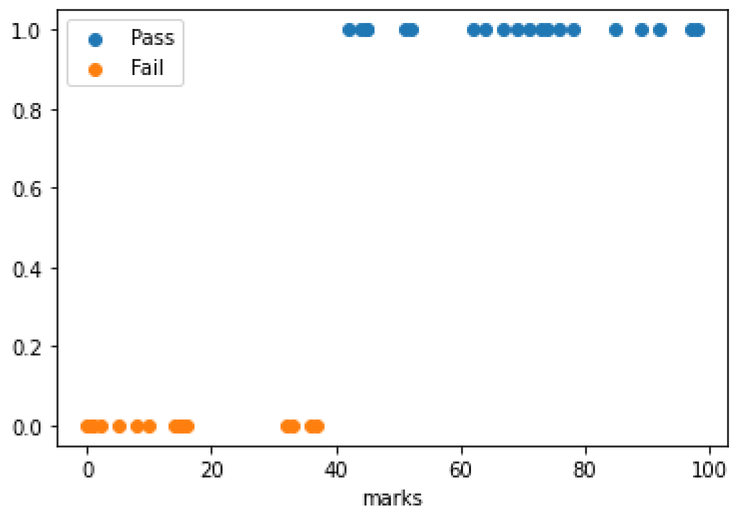
3. $f(x/4)$ - streached along x axis

4. $f(4x + 4)$ - combination of 1 and 2

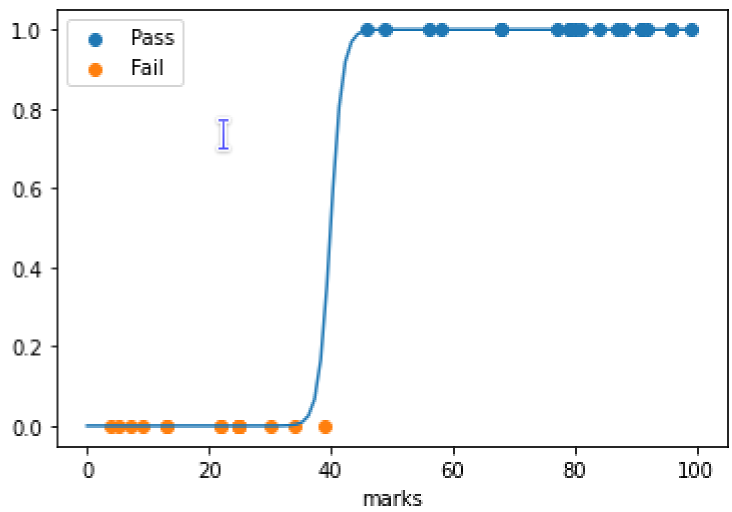
In Logistic Regression problems you will have to classify data into two categories. We will say those categories are 1 ans 0 for ex. Classifying if a student passed or failed an exam given marks of the students and passing marks are 40.

Note: Above is just a simple example and can be done without logistic Regression but it will be easy to understand how it works using a simple example.

Let's say following is the distribution of marks:



Students who passed the exam are put into category 1 and who failed are put into category 0. Now our task will be to fit the sigmoid function such that when we enter marks ≥ 40 it outputs value close 1 ie. pass and when we enter marks < 40 it outputs value close to 0 ie. fail. Let's looks at how the sigmoid function will look like when it is fit.



Here the fuction used is $sigmoid(x - 40)$. As you can see this fucntion outputs values close to 1 when it is ≥ 40 and 0 when it is < 40. We can say if $sigmoid(x-40) \geq 0.5$ then it belongs to category 1 and if $sigmoid(x-40) < 0.5$ it belongs to category 0.

So now we can simply define the logistic regression problem as finding the m and b in the function $sigmoid(mx + b)$ that best fit the data.

2 Loss function

Now lets talk about the loss function that we will use in Logistic Regression. Notice that we want to classify our data into 2 categories 1 and 0 unlike Linear Regresion where we get real values as outputs we can't directly take the MSE as the Loss function here.

So here's an idea, If correct catergory for the data point is 1 and our model predicts 1 then error should be zero but if our model predicts 0 it should be very large, similarly for 0 also. That means if our prediction is correct we get an error of 0 or rather close to 0 and if we make an incorrect prediction we get a very large error value. Makes sense right?

Lets analyse it case wise:

Case 1: Category 1 : Here correct output should be 1
if we get 0 as output we get a large error. How about we use log fucntion? If we say Loss = $-log(y_{pred})$: where y_{pred} is our prediction then when its 1 log is 0 and when it's 0 it will be infinity.

Case 2: Category 0 : Here correct output should be 0
if we get 1 as output, we get a large error. We will use log function here also.
Loss = $-log(1-y_{pred})$ then it satisfies both the conditions. See for youself.

Note: $y_{pred} = sigmoid(mx + b)$ and y_{true} takes only two values {0,1}

Now how do we combine both the cases?

$$Loss = -y_{true} * log(y_{pred}) - (1 - y_{true}) * log(1 - y_{pred})$$

See for youself that this is indeed true.

3 Gradient Descent

Now we have the Loss function that we want to minimize we can simply put it into the gradient descent equation ie.

$$m = m - \alpha * \frac{\delta(Loss)}{\delta(m)}$$

$$AND$$

$$b = b - -\alpha * \frac{\delta(Loss)}{\delta(b)}$$

On putting the values of y_pred and solving the above equation we get

$$m = m - \alpha * (y_{pred} - y_{true}) * x$$

$$AND$$

$$b = b - \alpha * (y_{pred} - y_{true})$$

This looks familiar, Doesn't it?

Note: I haven't shown any calculations here but if you want to try to derive this equation you can do so yourself.