



Hands on Lab on Logistic Regression

AGENDA OF THE VIDEO

- What is Logistic Regression?
- Sigmoid (Logistic Regression) Function
- Cost Function... (Optimization Function)
- How does the algorithm work?
- Pseudocode for Logistic Regression algorithm
- Loan Default Toy Dataset
- How to plot the data in Matplotlib?
- Coding the algorithm from scratch
- Logistic Regression Class in sklearn
- Task For you.....
- One more Task with Real World Dataset

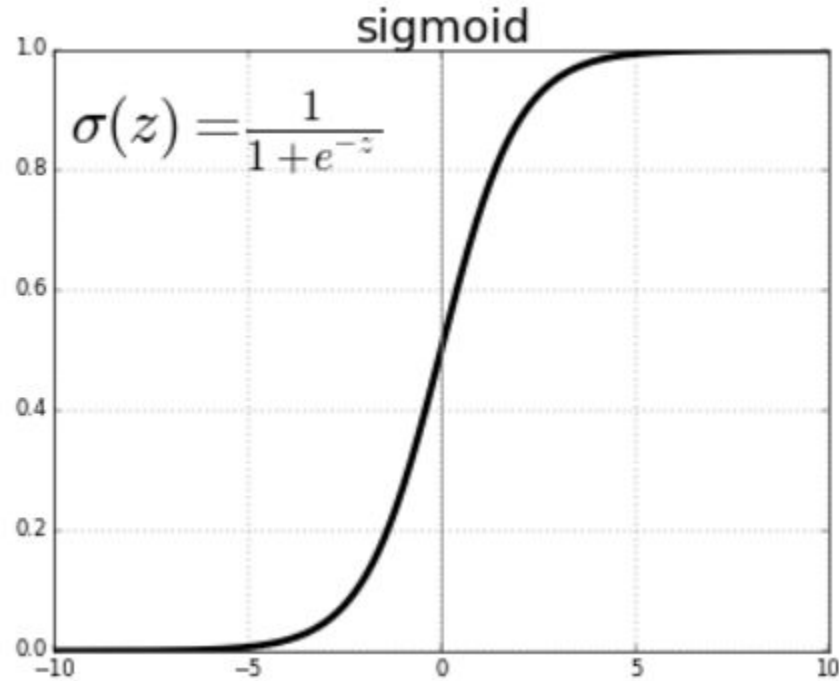
What is Logistic Regression?

Logistic Regression is an algorithm which is used to solve Binary Classification Related Problems.

It is called Regression for some historical reasons but it is used to solve Classification Related tasks.

It is called logistic because it uses a function called Logistic (or Sigmoid) which maps any real number to the range 0 to 1.

Logistic or Sigmoid Function



Linear vs Logistic Function

In Linear Regression , we compute the linear combinations of input and weights.

Then, that is what will be our Prediction. It will be a real valued output (Continuous).

But, in the Logistic Regression, we take that combination and pass that through the Logistic Function. Then, the resulting value will be ranging from 0 to 1.

Cost Function or Optimization Function

Cost = average of $[- \text{Actual} * \log(\text{Predicted}) - (1 - \text{Actual}) * \log(1 - \text{Predicted})]$

Why does this cost function?

It has to do with convex functions. Earlier, in the Linear Regression Functions, the Predicted values were just computed from the linear combinations of weights and features but here we have to take the sigmoid of it which makes it non-convex for which the minimum is not so easy to compute.

How does the algorithm work?

Step.1 Define Sigmoid Function

Step.2 Define Prediction Function (Using Sigmoid Function).

Step.3 Define Cost Function

Step.4 Initialize random weights, `n_iter` and `learning_rate`

Step.5 For `n_iter` times, do the following:

1. Get Prediction based on the Input and the `current_weight`.
2. Update the weight by using simultaneous update. (It is very important).

Pseudocode for Logistic Regression Algorithm

```
def sigmoid(z):  
    return ...  
def prediction_function(X, theta):  
    return ....  
def cost_function(prediction, actual_values):  
    return ....  
def calculate_update(theta, X, y):  
    get predictions  
    calculate update  
    return update  
# Define Learning_rate  
# Define n_iter  
# get X and y  
# initialize theta = some random  
  
for iter in n_iter:  
    compute update  
    calculate loss  
    theta -= update * learning_rate  
  
# Finally, after the loop ends, hopefully, we would have got the right values  
# of theta
```


Loan Default Toy Dataset

Income (in 1000 dollars)	Total Assests Value (in 100 dollars)	Deafult or Not
5	300	Defaulter
10	150	Defaulter
10	350	Defaulter
10	420	Defaulter
15	220	Defaulter
17	430	Defaulter
30	600	Non Default
30	710	Non Default
35	310	Non Default
35	450	Non Default
37	520	Non Default
40	630	Non Default
42	305	Non Default
44	400	Non Default

How to plot the data in Matplotlib?

Coding the algorithm from Scratch in Python

Let us now see Logistic Regression Class in Sklearn in action

Task For you...

Build a Logistic Regression Model to predict whether an Individual (say A) is going to Purchase a Product with a given Price and Rating or not.

Price	Rating	Purchase or Not
5	2.1	No
10	2.3	No
10	3.5	No
10	2.3	No
15	1.2	No
17	1.7	No
30	4.1	Yes
30	4.2	Yes
35	4.8	Yes
35	3.4	Yes
37	5	Yes
40	4.5	Yes
42	4.7	Yes
44	4.9	Yes

Task For you... One more Task

Go to Kaggle.com which is a website for hosting Data Science Competitions. On this website, you can get real world dataset for various problems.

So, what i want you to do is go to search bar and search for “Titanic Survival Prediction” and from this page, download the dataset.

Now, using the sklearn library, fit a Logistic Regression Model on this dataset to predict whether a particular person has survived or not.

Thank You