



## Discussion Forums

## Week 4

## SUBFORUMS

## All

Assignment: Multi-class Classification  
and Neural Networks

## ← Week 4



## ex3: tutorial for predictOneVsAll()

Tom Mosher · Mentor · a year ago · Edited

The code you add to predictOneVsAll.m can be as little as two lines:

1. one line to calculate the sigmoid() of the product of X and all\_theta. X is (m x n), and all\_theta is (num\_labels x n), so you'll need a transposition to get a result of (m x num\_labels)
2. one line to return the classifier which has the max value. The size will be (m x 1). Use the "help max" command in your workspace to learn how the max() function returns two values.

Note that your function must return the predictions as a column vector - size (m x 1). If you return a row vector, the script will not compute the accuracy correctly.

👍 30 Upvote · Follow 25 · Reply to Tom Mosher

🔒 This thread is closed. You cannot add any more responses.

Earliest Top Most Recent

PW

Peter Wright · 7 months ago

Hi all,

My OnevsAll and IrCostFunction has been marked correctly, but my prediction code does not return the correct values.

Sigmoid(X\*all\_theta') returns a 16x4 matrix.

I then proceed to use p= max(prediction,[],2).

Any ideas?

Thanks

👍 0 Upvote · Hide 2 Replies



Tom Mosher · Mentor · 7 months ago · Edited



The max() function returns two values.

- The first is the max values
- The second is the indexes of the max values.

You want the second one as your predictions. So you can use  $[v \ p] = \text{max}(\dots)$ .

👍 11 Upvote



Tom Mosher · Mentor · 7 months ago



*(note: this thread is closed to additional replies, because the Forum links to this thread no longer work correctly. To follow-up on this topic, please start a new thread in the Week 4 forum area)*

👍 0 Upvote



Hitarth Bhatt · 8 months ago



Try find out the maximum  $h(x)$  row wise. once you are done, copy that result into the  $p$  vector. That is your result.

👍 0 Upvote · Reply

RR

Rohan Raj · 8 months ago



I am getting a training set accuracy of 10 even though the grader accepted my submission. I used a for-loop for the implementation, iterating it from 1 to the number of training examples. For each row I found out the column number for the maximum value and updated 'p' with this value of the column number. Where am I going wrong?

👍 0 Upvote · Reply



Hitarth Bhatt · 8 months ago



I think when i write  $a = \text{sigmoid}(X * \text{all\_theta})$ , it results into a  $m \times \text{num\_labels}$  matrix which generally means it has the values of  $h(x)$  for 10 labels. Now the function of prediction is to find the maximum of all 10 classifiers.

Well when i write  $a = \text{sigmoid}(X * \text{all\_theta})$ ;  $p = \text{max}(a, [], 2)$ ; it should result into  $m \times 1$  column vector.

It stills gives me a zero accuracy that means it went wrong somewhere. Can you please relate to it?

👍 0 Upvote · Hide 6 Replies



Tom Mosher · Mentor · 8 months ago · Edited



When you use max() on a matrix, you have to provide for both return values.

- The first return value is the maximums - you don't care about those.
- The second return value is the index where the maximum was found. Those are your predictions.

👍 1 Upvote



Hitarth Bhatt · 8 months ago



Well  $\text{max}(a, [], 2)$  should return a matrix of length  $m \times 1$ . As it finds maximum row wise.

👍 0 Upvote



Tom Mosher · Mentor · 8 months ago





How are you saving the two return values?

👍 1 Upvote



Tom Mosher · Mentor · 8 months ago

That is...

- the vector of maximum values, and
- the vector of the indexes for the maximum values

👍 1 Upvote



Hitarth Bhatt · 8 months ago

I guess when i write  $[x,y] = \max(a,[],2)$  the maximum values or the indices end up in vector y.

👍 0 Upvote



Tom Mosher · Mentor · 8 months ago

The maximum values would be in 'x'.

The indexes of the maximum values would be in 'y'. Those are the predictions, so you can use 'p' instead of 'y' there.

👍 1 Upvote

JR

Jonathan Roman · 9 months ago

I'm still confused what the "DIM" 2 is specifying.

If we wanted to sum columns, would it be  $\max([],A,2)$ ?

👍 0 Upvote · Hide 1 Reply



Tom Mosher · Mentor · 9 months ago

The  $\max()$  function doesn't perform a sum.

👍 1 Upvote

JA

javaid aslam · 9 months ago

1) The representation of all\_theta here is different--each row represents 401 parameters corresponding to the 401 features.

2) The help doc in Octave is not very well written. Here  $\max(A, [], 2)$  means maximum of a row of the matrix A, with the second dimension (2), i.e., all the columns of for a given row. [] represents a matrix as opposed to a vector.

3) The index tells us which element among the 10 is maximum.

👍 1 Upvote · Reply

JR

Jonathan Roman · 9 months ago · Edited

Got this to work and accepted successfully, but baffled by what is going on. If you can elaborate, it would help my understanding.

1) By reading the first bullet, Z would be  $X (m \times n)$  times all\_theta' ( $n \times \text{num\_labels}$ ), which i thought made sense for matrix multiplication rules ( $n = n$ ), however, I did not get the test case answers even with the same code-- I get only 1 and 3 (which I can make into a column vector.) I had to swap the order and transpose. In one of the prior exercises, we said  $Z = X * \text{theta}$ . Why not here?

2) In addition, try as I might reading help max, I cannot understand the significance of  $\max(A, [], 2)$ . I got it to work with out it, but cannot figure out [] and 2 represent.

3) Also, I'm not sure why the max "Index" is being used.

👍 1 Upvote · Reply



Karthikeyan Ramakrishnan · 9 months ago

Hi, Tried two line of codes as suggested,

$h_{\theta}x$  = sigmoid fn of product of  $x$  and  $\theta$

$[val, p] = \max$  of  $h_{\theta}x$  as syntax for max function.

but while submitting one vs all is only running.. not predict one vs all function. Am i missing anything very basic? please guide

👍 0 Upvote · Hide 3 Replies



Tom Mosher Mentor · 9 months ago

The submit script automatically runs all of your functions.

👍 0 Upvote



Karthikeyan Ramakrishnan · 9 months ago

But predict one vs all is not running at all, am not getting the expected output as described. In addition to above lines do I need to run any other scripts?

👍 0 Upvote



Tom Mosher Mentor · 9 months ago

You run the exercise scripts to test your functions.

You run the submit script for grading.

👍 0 Upvote

LL

lennie leong · a year ago

Hi mentors. I have gotten full marks for first 2 assignments (lrCostFunction.m and oneVsAll.m)

Currently working on predictOneVsAll.m, i am stuck.

I am working on the 2 lines of codes as per Tom's tutorial. currently, my assignment is returning me: a column vector with various numbers within a column vector thats 16x1.

But i am not getting the marks for it. May I know what is going wrong?

I had also used the the test case for predictOneVsAll

(<https://www.coursera.org/learn/machine-learning/discussions/5g8LaZTCEeW0dw6k4EUmPw>) , it returned 2 answers:

"ans: 1.0000; 0.2689; 0.8808; 0.9991" and "ans: 0; 0 ; 0 ; 0". Instead of getting an answer that's [1; 2; 2; 1]

Mentors, can kindly help me see what went wrong here?

👍 0 Upvote · Hide 1 Reply

LL

lennie leong · a year ago

I had fixed it already. I just needed to store the numbers into a dummy

and index. I thank you mentors.

👍 0 Upvote



Lee Dong Lin · a year ago



The cost values that I obtained are different from each time I run ex3, is this normal?

👍 0 Upvote · Hide 1 Reply



Tom Mosher Mentor · a year ago · Edited



You are posting in the thread for the tutorial on the predictOneVsAll() function. There is no cost calculation required.

If you have a different question, please start a new thread, and post some examples of what you are discussing.

👍 0 Upvote



Sunil Skanda · a year ago



Can you tell me why we use index value again ? How does it give the best match ?

👍 0 Upvote · Hide 1 Reply



Tom Mosher Mentor · a year ago



It's just a handy coincidence that the index numbers are the same as the value of the digits we're classifying. We use the classifier with index 1 to identify the digit '1', the classifier with index 2 for the digit '2', etc. The only exception is that we use the index 10 to represent the digit '0', since the index value 0 isn't allowed in MATLAB/Octave.

In general, we would normally have to convert the index values into what they represent, maybe using a lookup table. But we can skip that step in this exercise.

👍 3 Upvote



Sunil Skanda · a year ago



I got 95% prediction accuracy.

But, " [dummy, p]=max (..)" - this line confuses me a bit. We take the max value and store it in p right? How does it predict ?

what do we mean by dummy and what does it do?

Thanks.

👍 0 Upvote · Hide 1 Reply



Tom Mosher Mentor · a year ago



No, we store the max value in "dummy", and the index of the max value in p. The variable name "dummy" isn't important, you can use any name you like. We totally ignore the values we store there. The index values is what we want.

👍 0 Upvote

RP

Ronni Pedersen · a year ago



Hi @Tom Mosher

I have stumbled over a small curiosity that I hope you can help me with.

My first version of predictOneVsAll is working perfectly fine despite not using the

sigmoid function at all.

It has two lines:

- One line that calculates the multiplication of the theta values and the input values
- One line that returns the classifier with the max value

This is really similar to what you suggest in the start of this thread, save the sigmoid function. Is this working because the sigmoid function is not really needed, as we are just looking for the number with the largest value (instead of the binary 0 or 1 that we are using later on in the neural network where the sigmoid function are indeed needed)?

👍 1 Upvote · Hide 1 Reply



Tom Mosher · Mentor · a year ago

Good question.

Since the sigmoid function is monotonic increasing, you can find the correct classification either with it or without it, at least for the purposes of this exercise. You'll get the same result.

Since we're doing logistic classification, the method formally expects the outputs to in the range for logical values - between 0 and 1. Whether to include the sigmoid function depends on what you may use the results for in some future application.

👍 0 Upvote

JJ · a year ago

I'm testing my code using the small test case provided and I've misunderstood something about how `h_theta_X` is calculated:

```
1 all_theta = [1 -6 3; -2 4 -3];
2 X = [1 7; 4 5; 7 8; 1 4];
3 predictOneVsAll(all_theta, X)
```

I hand calculate `h_theta_X` for the second input set (1, 4, 5) by first multiplying the theta matrix with the input set to get:

```
1 [1 - 24 - 15; -2 + 16 -15]
```

both of which are less than 0.5 so the sigmoid returns 0 for both - which means that `h_theta_X` is [0, 0] means that it wasn't able to classify it.

I'm kind of unsure of where I'm going wrong here.

👍 0 Upvote · Hide 3 Replies



Tom Mosher · Mentor · a year ago

The `sigmoid()` function doesn't have a 0.5 threshold. It just returns the value.

In `predictOneVsAll()` we want to return the classifier with the highest value, regardless of whether it is greater or less than 0.5.

The 0.5 threshold is only used in logistic regression where you have two classes. It is not part of the sigmoid function itself.

👍 0 Upvote

J

· a year ago

Thanks Tom. That makes sense. My classifier is working now.

👍 0 Upvote

Tom Mosher · Mentor · a year ago



Good news!

👍 0 Upvote



Elad Wasserstein · a year ago

Hello,

I just want to make sure why I'm interested in using the MAX:

normally if i had one class prediction i would have compare the predicted value against 0.5

$p > 0.5 \Rightarrow$  it is true else it is false.

However in this exercise we have multiple results greater than 0.5 therefore we are looking for the closest prediction to 1 (implemented by max).

with this implementation even if all results predict lower than 0.5 we will get some result.

Is the explanation correct?

thanks ahead!

👍 0 Upvote · Hide 4 Replies



Tom Mosher · Mentor · a year ago

Mostly right. You do not care whether any of the predictions are  $> 0.5$  You simply use the one with the maximum value.

👍 1 Upvote

NX

nika xue · 9 months ago

Same confusion. Since we are predicting the accuracy, we should judge whether the number returned by sigmoid is  $\geq 0.5$  or not, which is similar as what predict.m in ex2 doing. If yes, the classifier should be put to p, otherwise it means the record doesn't belong to any classifier and we can put 0 instead. After that return p to pred.

Without using condition, my accuracy is 94.92% but with it, it is just 89.64%. I think the latter number makes more sense.

👍 0 Upvote



Tom Mosher · Mentor · 9 months ago

You're not predicting the accuracy. You are predicting the confidence that a given example is a member of a specific class.

When you have two classes and one logistic output, you split them using a fixed threshold.

When you have multiple classes and multiple outputs, you do not use a fixed threshold. You select the class that gives you the highest value.

👍 1 Upvote

nika xue · 9 months ago



NX

Got it, thanks!

👍 0 Upvote



