coursera

Discussion Forums

Get help and discuss course material with the community.

THIS WEEK'S FORUM

Week 6

Discuss and ask questions about Week 6.

19 threads · Last post 5 hours ago

Go to forum

Forums | All Threads

Search

All Course Discussions

Ex2 Tutorial: vectorizing the gradient calculation

Tom Mosher Mentor · 2 years ago · Edited

The gradient calculation can be easily vectorized. See this two formulas from ex2.pdf pages 9 and 10.

Note: ignore the λ term in the 2nd equation if you are working on costFunction() - just do Step 1 and Step 2.

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \qquad \text{for } j = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \left(\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}\right) + \frac{\lambda}{m}\theta_j \quad \text{for } j \geq 1$$

Note that if we set θ_0 to zero (in Step 3 below), the second equation is exactly equal to the first equation. So we can ignore the "j = 0" condition entirely, and just use the second equation.

1. Recall that the hypothesis vector h is the sigmoid() of the product of X and θ (see ex2.pdf - Page 4). You probably already calculated h for the cost J calculation.
2. The left-side term is the vector product of X and (h - y), scaled by 1/m. You'll need to transpose and swap the product terms so the result is (m x n)' times (m x 1) giving you a (n x 1) result. This is the unregularized gradient. Note that the vector product also includes the required summation.
3. Then set theta(1) to 0 (if you haven't already).

4. Then calculate the regularized gradient term as theta scaled by (lambda / m).
5. The grad value is the sum of the Step 2 and Step 4 results. Since you forced theta(1) to be zero, the grad(1) term will only be the unregularized value.

==========

keywords: ex2 tutorial costfunction tutorial costfunctionreg gradient

28 Upvote · Follow 41 · Reply to Tom Mosher

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

Earliest **Top** **Most Recent**

Atheer Al Attar · 4 months ago

Tom,

Just wanted to say thanks! this great tutorial helps. The error I was doing is that I was excluding X(:,1) when calculating the gradient for J>=1, moreover I was also calculating sigmoid only for X(:,1) when I was calculating θ_0. This tutorial really helps.

2 Upvote · Hide 1 Reply

Tom Mosher Mentor · 4 months ago

That's good news.

0 Upvote

CJ

Chetan Jain · 4 months ago

Tom,

Thanks for the tutorial. I am able to pass the assignment with its help. Just curious to learn though why would you set theta(1) = 0?

one thing that prevented me from passing the assignment an hour earlier was that I was setting theta(1) =0 at the very beginning. It only has to be set for the regularization term. But why?

0 Upvote · Hide 1 Reply

Tom Mosher Mentor · 4 months ago · Edited

That is a simple way of omitting theta(1) from the regularization calculations, without having to use if-statements or any special vector addressing.

1 Upvote

Y

yuheng · 5 months ago

Hi Tom,

I'm still a bit stuck after trying out the test case for costFunctionReg(theta,X,y,3). I get an answer of 9.6832 for the cost but the grad is correct. I've checked the regularising (this is not a code) (lambda/2m x sum of theta^2) term but can't seem to find anything wrong with it.

any idea how else i can go about testing/troubleshooting it?

0 Upvote · Hide 4 Replies

Y yuheng · 5 months ago

in the case of lambda = 0 i get the right cost, so i'm suspecting that it's the lambda term that's causing the issue. have already ensured the parenthesis around 2m..

0 Upvote

 Tom Mosher Mentor · 5 months ago

Did you exclude theta(1) from the cost regularization term?

0 Upvote

Y yuheng · 5 months ago

Oh! i see.. that's where the problem is. thanks for your help!!:)

0 Upvote

 Tom Mosher Mentor · 5 months ago

Good news.

0 Upvote

 Simar Singh · 6 months ago

Hi Tom,

when i run the cost function it gives me the correct answer for J(theta) = 0.693 . but after the execution of fminunc it return me the same value of 0.693 and all the values of theta = 0 . and because of which i am getting a warning of divide by zero while executing the plotDecisionBoundary() function. Any idea why?

1 Upvote · Hide 2 Replies

 Tom Mosher Mentor · 6 months ago

Your cost function doesn't work correctly. Please check the tutorial and use the additional test cases, both from the Resources menu.

0 Upvote

 Tom Mosher Mentor · 6 months ago

Also read the FAQ thread in the Week 3 Discussion Forum area.

0 Upvote

 Анастасия · 6 months ago · Edited by moderator

Hi, Tom!

Please help

There's something wrong with my 'costFunctionReg'. I have an error:

```
>> [j g] = costFunctionReg(theta, X, y, 0)
error: costFunctionReg: A(I) = X: X must have the same size as I
error: called from
    costFunctionReg at line 26 column 9
```

Well, I have:

grad equals **{mentor edit: code removed due to Honor Code violation}** multiply theta

I'm not sure how to make the correct code for grad(1):

just grad(1) equals {**mentor edit}** ? Or should I add something in the end?

h is sigmoid(X*theta)

0 Upvote · Hide 9 Replies

Анастасия · 6 months ago

Now It says 'Nice work!'

```
 % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  1600  100   636  100   964    320    485  0:00:01  0:00:01 --:--:--   489
==
==                                  Part Name |     Score | Feedback
==                                  --------- |     ----- | --------
==                           Sigmoid Function |   5 /   5 | Nice work!
==                       Logistic Regression Cost |  30 /  30 | Nice work!
==                   Logistic Regression Gradient |  30 /  30 | Nice work!
==                                    Predict |   0 /   5 |
==          Regularized Logistic Regression Cost |  15 /  15 | Nice work!
==      Regularized Logistic Regression Gradient |  15 /  15 | Nice work!
==                                  ------------------------------
==                                           |  95 / 100 |
==
```

But test results are wrong. They're the same as for costFunction.

```
X = [ones(3,1) magic(3)];
y = [1 0 1]';
theta = [-2 -1 1 2]';

% un-regularized
[j g] = costFunction(theta, X, y)
% or...
[j g] = costFunctionReg(theta, X, y, 0)

% results
j = 4.6832

g =
  0.31722
  0.87232
  1.64812
  2.23787

% regularized
[j g] = costFunctionReg(theta, X, y, 3)
% note: also works for ex3 lrCostFunction(theta, X, y, 3)

% results
j = 7.6832

g =
  0.31722
  -0.12768
  2.64812
  4.23787
```

```
>> X = [ones(3,1) magic(3)];
>> y = [1 0 1]';
>> theta = [-2 -1 1 2]';
>> [j g] = costFunction(theta, X, y)
j =  4.6832
g =

   0.31722
   0.87232
   1.64812
   2.23787

>> [j g] = costFunctionReg(theta, X, y, 0)
j =  4.6832
g =

   0.31722
   0.87232
   1.64812
   2.23787

>>
```

What should I do? Can't figure out where is a mistake ((

0 Upvote

Tom Mosher Mentor · 6 months ago · Edited

costFunction() and costFunctionReg() give the same results if you set lambda to zero.

0 Upvote

Анастасия · 6 months ago

So, that means everything is ok, right? Or not?

0 Upvote

Анастасия · 6 months ago

How can I get these results?

```
% regularized
[j g] = costFunctionReg(theta, X, y, 3)
% note: also works for ex3 lrCostFunction(theta, X, y, 3)

% results
j = 7.6832

g =
  0.31722
  -0.12768
  2.64812
  4.23787
```

Please, I don't know how to fix it

0 Upvote

Tom Mosher Mentor · 6 months ago

For costFunctionReg(), you can copy your code for costFunction(), and add to it the regularlized terms that use lambda, for both cost J and the gradients.

See ex2.pdf pages 9 and 10 (highlighted here):

2.3 Cost function and gradient

Now you will implement code to compute the cost function and gradient for regularized logistic regression. Complete the code in `costFunctionReg.m` to return the cost and gradient.

Recall that the regularized cost function in logistic regression is

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2.$$

Note that you should not regularize the parameter θ_0. In **Octave/MATLAB**, recall that indexing starts from 1, hence, you should not be regularizing the `theta(1)` parameter (which corresponds to θ_0) in the code. The gradient of the cost function is a vector where the j^{th} element is defined as follows:

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \qquad \text{for } j = 0$$

9

$$\frac{\partial J(\theta)}{\partial \theta_j} = \left(\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}\right) + \frac{\lambda}{m}\theta_j \qquad \text{for } j \geq 1$$

Once you are done, `ex2_reg.m` will call your `costFunctionReg` function using the initial value of θ (initialized to all zeros). You should see that the cost is about 0.693.

0 Upvote

Tom Mosher Mentor · 6 months ago · Edited

See also the tutorial for the regularized cost portion of this exercise, from the Resources menu. That is in addition to this thread, which covers the regularized gradient.

0 Upvote

Анастасия · 6 months ago

Yes, that's exactly what I did. I added the regularlized terms. And I looked through all the tutorials and all the discussions before I decided to ask you. Still I don't know where is a mistake.

I guess it's with gradient.

Should I write 2 lines of code: one - for 'grad' and another for 'grad(1)'? Could I use (lambda divided by m) multiply [0; theta(2:end)] for j>=1? Or just put 'theta' instead of [0; theta(2:end)] ? because I set theta(1) to zero before that?

I'm so sorry, Tom, for disturbing you. Just don't want to leave it as it is. I want to understand.

Thank you.

0 Upvote

Tom Mosher Mentor · 6 months ago

Step 3 in the tutorial covers this.

If you use Step 3 - setting theta(1) = 0 - then you do not need the "j = 0" equation at all.

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)} \qquad \text{for } j = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \left(\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}\right) + \frac{\lambda}{m}\theta_j \qquad \text{for } j \geq 1$$

0 Upvote

Анастасия · 6 months ago

Hi, Tom!

I've just decided to get back to this task after a while.

Everything was right, and was just so inattentive doing test cases. Now I see.

I'm sorry.

Thank you !

1 Upvote

S

Srinivas · 6 months ago

Though I am getting result, when I submit "Logistic regression Cost", "Logistic regression Gradient" are not being considered for submission - I don't see any score. What could be the reason?

Also I get 2 separate graphs - 1) scattered scores 2) cross line. What's the way to merge them to single graph. I see "hold off" statement in ex2.m

0 Upvote · Hide 2 Replies

Tom Mosher Mentor · 6 months ago

1) Be sure that you have modified the "plotData.m" function correctly.

2) An error in your plotting function will not cause your code to fail the submit grader. So there is also an error in your code.

Use the additional test cases and tutorials from the Resources menu.

0 Upvote

S

Srinivas · 6 months ago

Thanks Tom. I realized I wrote plot code in ex2 instead of plotData.m. I am able to see graph correctly. I still have to figure out the error issue.

0 Upvote

WS

wooyoung Son · 7 months ago

theta(1) = 0 is just brilliant!

I was wondering around to find this one single line

thank you!

1 Upvote · Reply

yiyushi · 7 months ago

Hi Tom, I hav already got 85/100 except the regularized gradient problem, I first copy the hypothesis formula and set theta(1)=0, then calculate a= lamdba/m*theta, then add a to the previous term, i don't know where the error is, I set theta(1)=0 after unregularized term for both costfunction and gradient, then do additional calculation , finally add them together, why is still wrong? Please help me out.

0 Upvote · Hide 3 Replies

Tom Mosher Mentor · 7 months ago

Your method sounds good. I do not know what the problem is.

0 Upvote

Thomas · 5 months ago

I'm having the same issue as yiyushi was. Something is clearly wrong with my gradient function b/c it fails even when lambda = 0. Even though I copied it directly from my Costfunc (which passes the test cases).

0 Upvote

Tom Mosher Mentor · 5 months ago

Cost and Gradient should be returned by the same function.

There is no separate "gradient function".

0 Upvote

Gabriel Castillo · 8 months ago

Awesome post! thank you very much! I had such a busy week and was stuck in here!! you're a life saver.

0 Upvote · Reply

AM

ahmed montasser · 8 months ago

Hi Tom,

why i'm getting wrong answer when adding the sum() function exactly where it's in the formula ? I got correct results only when i removed it from regularized gradient formula ? what is the reason for that ?

0 Upvote · Hide 2 Replies

Tom Mosher Mentor · 8 months ago · Edited

Matrix-vector multiplication automatically includes calculating the sum. Try these commands in your console:

```
Q = reshape(1:10,5,2)        % creates a matrix for testing
v = [1;2]                    % creates a vector for testing
Q*v                          % the product of a matrix and a
    vector
sum(Q*v)                     % the sum of that vector
```

If you want a vector result, then adding that last sum() function ruins it.

0 Upvote

AM

ahmed montasser · 8 months ago

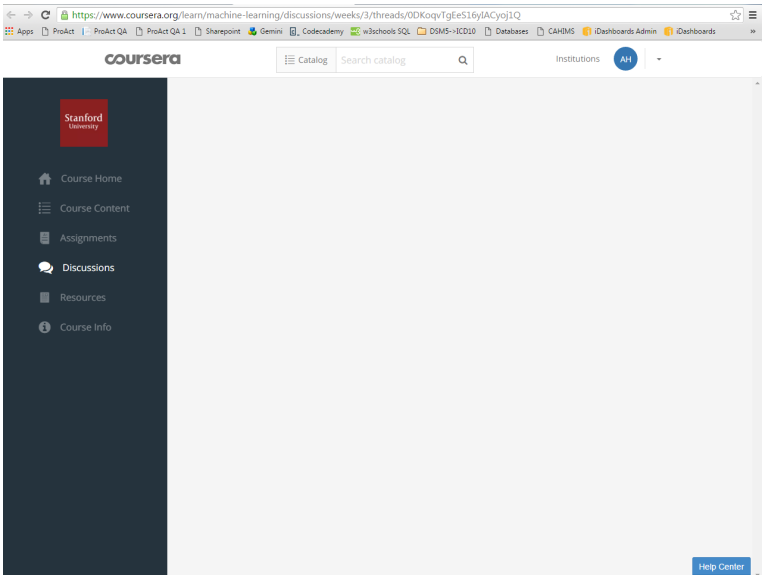
thank you, i think i should have a linear algebra review

0 Upvote

AH

Anne Houser · 9 months ago · Edited

Does anyone know why the "costFunction() cost tutorial" is missing? When clicking on that hyperlink, I am directed to this page:



However, the "costFunction() gradient tutorial" hyperlink works fine.

0 Upvote · Hide 5 Replies

Tom Mosher Mentor · 9 months ago · Edited

Yesterday, Coursera updated the software that runs the Discussion Forum, and the update broke access to many of the Forum areas. Since the tutorials were developed at different times, the actual posts are stored in many different Forum areas. Some are still working - some are not.

Hopefully they will fix it soon.

1 Upvote

AH

Anne Houser · 9 months ago

Thanks for the update!

0 Upvote

AH

Anne Houser · 9 months ago

I am trying to submit my work for Logistic Regression Cost and Logistic Regression Gradient but it is not grading it. I have gotten the correct cost from page 5 of the associated pdf, but when I go to submit my work, only the Sigmoid Function is graded. I would think this means I have the wrong code, but since my answer is the same as in the pdf, I am unsure of why I am not getting graded for this exercise.

0 Upvote

Tom Mosher Mentor · 9 months ago

The cost value 0.693 is just for an extremely simple test, where all theta values are zero.

Multiplying by zero can hide any number of problems. And the submit grader uses a totally different test case, it's not even looking for the 0.693 value.

Sadly, the cost function test cases are not available right now, due to Coursera breaking the Forum software during yesterday's update.

My best advice is to pay close attention to the placement of the parenthesis, and be sure you're using the cost function equation from the PDF file - the one in the video lecture is wrong.

0 Upvote

AH

Anne Houser · 9 months ago

Thank you! I had thought about the forum breaking being a reason why it was not grading the assignment but thought that couldn't be right. I will be sure to update my notes from the lecture video with correct equations, thanks again!

0 Upvote

VM

Vedant Mehta · 9 months ago

Hi!

I am having trouble finishing the costFunctionReg file. I simply used the code previously written in the costFunction file and add the regularization term. I am getting the answer as written in the pdf for cost function, but still the grader is giving me 0 points.

I get the following output when I run the ex2_reg.m file.

```
Cost at initial theta (zeros): 0.693147

Program paused. Press enter to continue.

Local minimum possible.

fminunc stopped because the size of the current step is less than
the default value of the step size tolerance.

<stopping criteria details>

Train Accuracy: 51.694915
```

Thanks in advance!

0 Upvote · Hide 2 Replies

Tom Mosher Mentor · 9 months ago

The answer given in the PDF file means very little. That's not the test case that the submit grader uses.

The "fminunc stopped..." message is not a problem by itself. The current version of fminunc() from MATLAB provides you that message, so you know what condition was met before it returned. That's not an issue.

But your low training accuracy, that means there is a problem somewhere else.

Try to debug your cost function using the additional test cases in this thread:

https://www.coursera.org/learn/machine-learning/discussions/iyd75Nz_EeWBhgpcuSIffw

2 Upvote

VM

Vedant Mehta · 9 months ago

Got the answer. I understood my mistake. Thanks.

0 Upvote

VM

Vedant Mehta · 9 months ago

Hi!

I solved the costFunction for cost but somehow my gradient is incorrect since the grader is giving me zero points for the gradient part. Can someone please tell me the answer of the gradient at initial theta?

Thanks in advance.

0 Upvote · Hide 8 Replies

Tom Mosher Mentor · 9 months ago

```
Cost at initial theta (zeros): 0.693147
Gradient at initial theta (zeros):
 -0.100000
 -12.009217
 -11.262842
```

0 Upvote

VM

Vedant Mehta · 9 months ago

Thank you for a prompt reply.

My answer comes out to be -23.372059 and it generates only one value. Am I doing something wrong?

0 Upvote

Tom Mosher Mentor · 9 months ago

If your solution is not a vector, then you're doing something wrong.

From your number compared to my numbers, it is pretty easy to see what the problem might be. I recommend you think about it.

0 Upvote

Tom Mosher Mentor · 9 months ago

I also recommend you read the tutorial:

https://www.coursera.org/learn/machine-learning/discussions/iyd75Nz_EeWBhgpcuSIffw

0 Upvote

VM

Vedant Mehta · 9 months ago

Got it. Thank you so much for your help.

0 Upvote

GB

Graham Benson · 5 months ago

I am getting the same gradient values as above except they are positive not negative. Any suggestions why this might be happening.

Thank you.

0 Upvote

Tom Mosher Mentor · 5 months ago

Use the equations from the exercise PDF file - not the ones from the video lectures. Check if you have correct parenthesis grouping.

0 Upvote

GB

Graham Benson · 4 months ago

My error was because I was implementing gradient descent update rather than just calculating the derivative part via the following formula;

The vectorized version;

$$\nabla J(\theta) = \frac{1}{m} \cdot X^T \cdot \left(g(X \cdot \theta) - \vec{y}\right)$$

In the test cases my cost function is correct and the gradient calculation is correct for the unregularized logistic regression model. For the regularized version my gradient calculation is incorrect. I have followed your advice above (point 4.) and simply added (lambda/m)*theta to the unregularized value (theta(1) has been set to 0). Could you please advise where I am going wrong.

1 Upvote

DL

David Liu · 10 months ago · Edited by moderator

I find the suggested instructions problematic. You should NOT set $theta_1$ to 0 as suggested in some of the tutorial text here. It all depends on how you are using your "theta" matrix in your code. In some places you do still need that first element of "theta", e.g. you calculate $\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$.

What you should do instead is to use $theta(2:end,:)$. e.g.

```
1  (mentor edit: code removed)
```

or

use the following to replace the first element of theta to 0

```
1  (mentor edit: code removed)
```

0 Upvote · Hide 2 Replies

Tom Mosher Mentor · 10 months ago

The tutorial tells you to compute the hypothesis first, then you are free to set theta(1) = 0, since you no longer need the original theta.

0 Upvote

Tom Mosher Mentor · 10 months ago

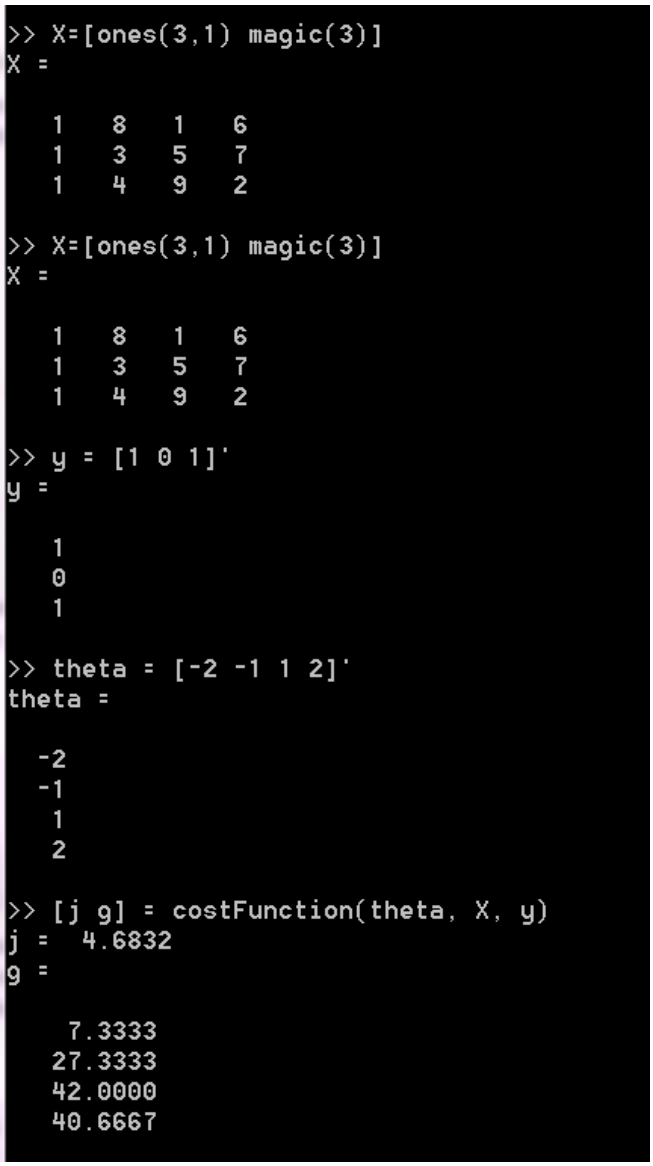
Posting your code violates the course Honor Code. I have edited your post.

0 Upvote

NM

Nihalf Mohammed · 10 months ago

Hi - for this on the unregularised case I seem to have calculated the sigmod right, and the cost right, however my grad values calculation is off? I've tried the test cases, and my results seem off by a large amount, I get the following output from the test case:



my grad calculation is (1/m) times (X transpose times (h(x)-y)) where h(x) is the sigmoid function applied to (X times theta). I'm getting the (nx1) matrix, but the values... any ideas where I'm going wrong?

0 Upvote · Hide 3 Replies

Tom Mosher Mentor · 10 months ago · Edited

I think you're using the wrong cost function - your description is of the linear regression cost function, but with the sigmoid() function added. That's not how logistic regression works.

Please look at the bottom of Page 4 of ex2.pdf, and see if that's what you've implemented.

0 Upvote

NM

Nihalf Mohammed · 10 months ago

Oh gods, embarassingly, I forgot to apply the sigmoid function in the grad calculations. Silly mistake, my bad. Adding in the sigmoid function gives me the correct answer

It works, but I am still confused by your response - I did not use the linear regression cost function, I use the cost function J = (-y times log (sigmoid(X times theta))) - ((1-y) times log(1 - sigmoid(X times theta))).

1 Upvote

Tom Mosher Mentor · 10 months ago

Sorry, I was confused. I missed that you said it was your grad calculation, not the cost function. My mistake.

I'm glad you got it working.

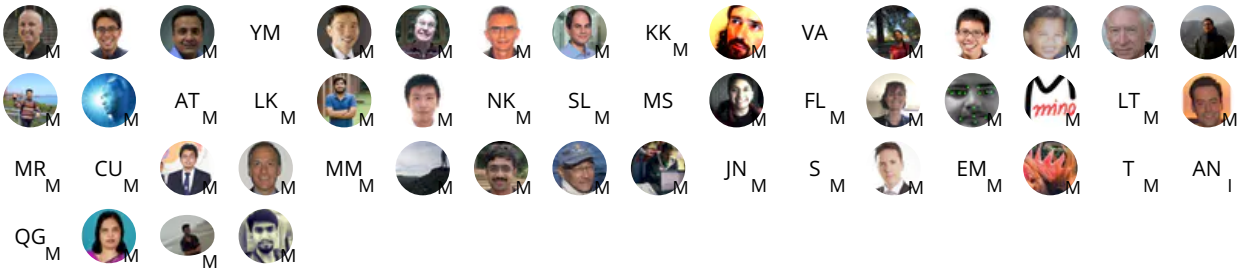
0 Upvote

1 2 3

DESCRIPTION

Welcome to the course discussion forums! Ask questions, debate ideas, and find classmates who share your goals. Browse popular threads below or other forums in the sidebar.

MODERATORS



Learn more about becoming a Mentor

Forum guidelines