coursera

Week 3

SUBFORUMS
All
Assignment: Logistic Regression

Ex2 Tutorial: vectorizing the Cost function

Tom Mosher   Mentor   Week 3 · 3 years ago · Edited

The regularized cost calculation can be vectorized easily. Here is the cost equation from ex2.pdf, page 9.

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

1. The hypothesis is a vector, formed from the sigmoid() of the products of X and θ. See the equation on ex2.pdf - Page 4. Be sure your sigmoid() function passes the submit grader before going any further.

2. First focus on the circled portions of the cost equation. Each of these is a vector of size (m x 1). In the steps below we'll distribute the summation operation, as shown in purple, so we end up with two scalars (for the 'red' and 'blue' calculations).

3. The red-circled term is the sum of -y multiplied by the natural log of h. Note that the natural log function is log(). Don't use log10(). Since we want the sum of the products,

we can use a vector multiplication. The size of each argument is (m x 1), and we want the vector product to be a scalar, so use a transposition so that (1 x m) times (m x 1) gives a result of (1 x 1), a scalar.

4. The blue-circled term uses the same method, except that the two vectors are (1 - y) and the natural log of (1 - h).

5. Subtract the right-side term from the left-side term

6. Scale the result by 1/m. This is the unregularized cost.

7. Now we have only the regularization term remaining. We want the regularization to exclude the bias feature, so we can set theta(1) to zero. Since we already calculated h, and theta is a local variable, we can modify theta(1) without causing any problems.

8. Now we need to calculate the sum of the squares of theta. Since we've set theta(1) to zero, we can square the entire theta vector. If we vector-multiply theta by itself, we will calculate the sum automatically. So use the same method we used in Steps 3 and 4 to multiply theta by itself with a transposition.

9. Now scale the cost regularization term by (lambda / (2 * m)). Be sure you use enough sets of parenthesis to get the correct result. **Special Note for those whose cost value is too high:** 1/(2*m) and (1/2*m) give drastically different results.

10. Now add your unregularized and regularized cost terms together.

===============

keywords: ex2 tutorial costfunction costfunctionreg

⇧ 65 Upvotes　　💬 Reply　　　Follow this discussion

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

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

Tom Mosher　Mentor　· 3 years ago

Notice:

This thread is closed to further discussion (due to a technical glitch in the Forum software, new replies are being made invisible).

If you are waiting for a reply on this thread, please start a new thread and repeat your question.

⇧ 0 Upvotes　　💬 Reply

SS　Samatha Shetty · 3 years ago

I got the ex2 right for all but the sigmoid method. My cost function which uses sigmoid passed, but I got a 0 for sigmoid when grading ex2. Please advise.

⇧ 1 Upvote   ▢ Hide 2 Replies

Tom Mosher   Mentor · 3 years ago

You can get a one-line program for sigmoid() if you use all element-wise operators.

- The exp() function is element-wise.
- The addition operator is element-wise.
- Use the element-wise division operator ./

⇧ 0 Upvotes

SS   **Samatha Shetty** · 3 years ago

Thanks Tom. My own for loops calculated only the first column right,leaving others to 0. This was a good repeat of matrix operations as well.

⇧ 0 Upvotes

Kumuda Gururao · 3 years ago

Hi Tom,

I am getting cost as 0.693 as expected and after fminunc calling cost function get the cost as 0.203.

But I am not getting the graph correctly. Instead of '+' and 'o', I am getting st.lines,also Y axis shows very less marks from 0.25 to very low values not 30 to 90 like x axis.

It also suggests error in files given to us such as

Error: File: plotData.m Line: 45 Column: 1

Illegal use of reserved keyword "end".

Error in plotDecisionBoundary (line 12)

plotData(X(:,2:3), y);

Error in ex2 (line 95)

plotDecisionBoundary(theta, X, y);

I've given the screen shot below.

Do let me know what should I do to get the graph correctly.

⬆ 0 Upvotes   💬 Hide 5 Replies

Tom Mosher **Mentor** · 3 years ago

It appears that you may not have pasted the code into plotData.m in the correct location within the script template.

⇧ 0 Upvotes

Tom Mosher **Mentor** · 3 years ago

Also, go to this post, and read the notes in the ex2 section about the plotDecisionBoundary() function.

https://www.coursera.org/learn/machine-learning/forum/8LDwTL2SEeSEJSlACyEKsQ/discussions/m0ZdvjSrEeWddilAC9pDDA

⇧ 0 Upvotes

Kumuda Gururao · 3 years ago

Thanks Tom. I had made a small mistake of putting +p instead of p in the prediction file. I've changed that and it worked fine.

Thx a ton! I'v submitted the results and done well.

⇧ 0 Upvotes

AA **abdulrahman aljahoosh** · 3 years ago

should i refer to sigmoid(z) with @ in cost function

⇧ 0 Upvotes

Tom Mosher **Mentor** · 3 years ago

No.

⇧ 0 Upvotes

jim · 3 years ago

Are the costFunction for 1.2.2 and 1.2.3 the same ? when i finish the costFunction for part 2 in ex2, i get 0.693 correctly. But when i try to to run part 3 (the text does not imply you have to change anything) , it doesnt give the cost 0.203.

coursera

Tom Mosher **Mentor** · 3 years ago

It is the same cost function. If your cost function works correctly (both for J and grad), it will give 0.693 for the all-zeros theta case, and will give the 0.203 as the final value of cost after fminunc() runs gradient descent and finds the final theta values.

↑ 0 Upvotes

DM **Don McClean** · 3 years ago

Is the value chosen for lambda arbitrary? I have not seen anywhere guidelines on what to use.

↑ 0 Upvotes   💬 Hide 3 Replies

Tom Mosher **Mentor** · 3 years ago

For now, lambda is provided for you. In Week 6, you will learn how to select it yourself.

↑ 0 Upvotes

DM **Don McClean** · 3 years ago

Tom

Thanks, I didn't realize it was a passed parameter

Don

↑ 0 Upvotes

Tom Mosher **Mentor** · 3 years ago

We don't get a lambda parameter until the ex2_reg.m script and costFunctionReg().

↑ 0 Upvotes

ZW **Ziqi Wang** · 3 years ago · Edited by moderator

Hi Tom, for the ex2_reg, i got the grad right but the cost really drives me mad. I got the corresponding answer for inital cost as 0.693 when i use the X*theta, but it is wrong since the theta(1) was changed to 0...However, when i change it into h =

X*theta and use h than the answer becomes 0.728...however, i think it is the correct version and i just do not know why i cost function as wrong... Here is what i wrote:

{Mentor edit: code removed due to deliberate Honor Code violation}

btw, for initial_theta = zeros..doesn't it set all theta to zero...then what is the purpose of theta(1)=0?

i will delete immediately after getting your reply.

⇧ 1 Upvote   ▢ Hide 2 Replies

ZW   Ziqi Wang · 3 years ago

And it hints that i shouldn't regularize theta(1), and i know that by using theta(1)=0...but the answer was wrong...but when i do regularize theta(1), i got the same answer as the ex2.pdf, but it is still wrong....so...

⇧ 0 Upvotes

Tom Mosher   Mentor   · 3 years ago

I will delete your code right now, because you have violated the course Honor Code by posting it. And you seem to know that is a wrong thing to do.

⇧ 0 Upvotes

Vikram Bahl · 3 years ago

Hi Tom,

In the regularized cost function, when I did, as was suggested in Step 8

```
1   thetasquare = theta * theta';
```

It did not grade my submission.

Instead it graded me correctly when I implemented it as

```
1   thetasquare = sum(theta.^2);
```

I understand that vector multiplication with itself(transposed) automatically sums up and therefore these two should be the same.

What am I missing?

⬆ 3 Upvotes 💬 Hide 1 Reply

Tom Mosher Mentor · 3 years ago · Edited

Theta is size (n x 1). So if you perform theta * theta', you get (n x 1) times (1 x n),which gives you a (n x n) square matrix. But the answer you want is a scalar (size 1x1). So try swapping the order of the operands.

(1 x n) times (n x 1) will give the same result as sum(theta.^2).
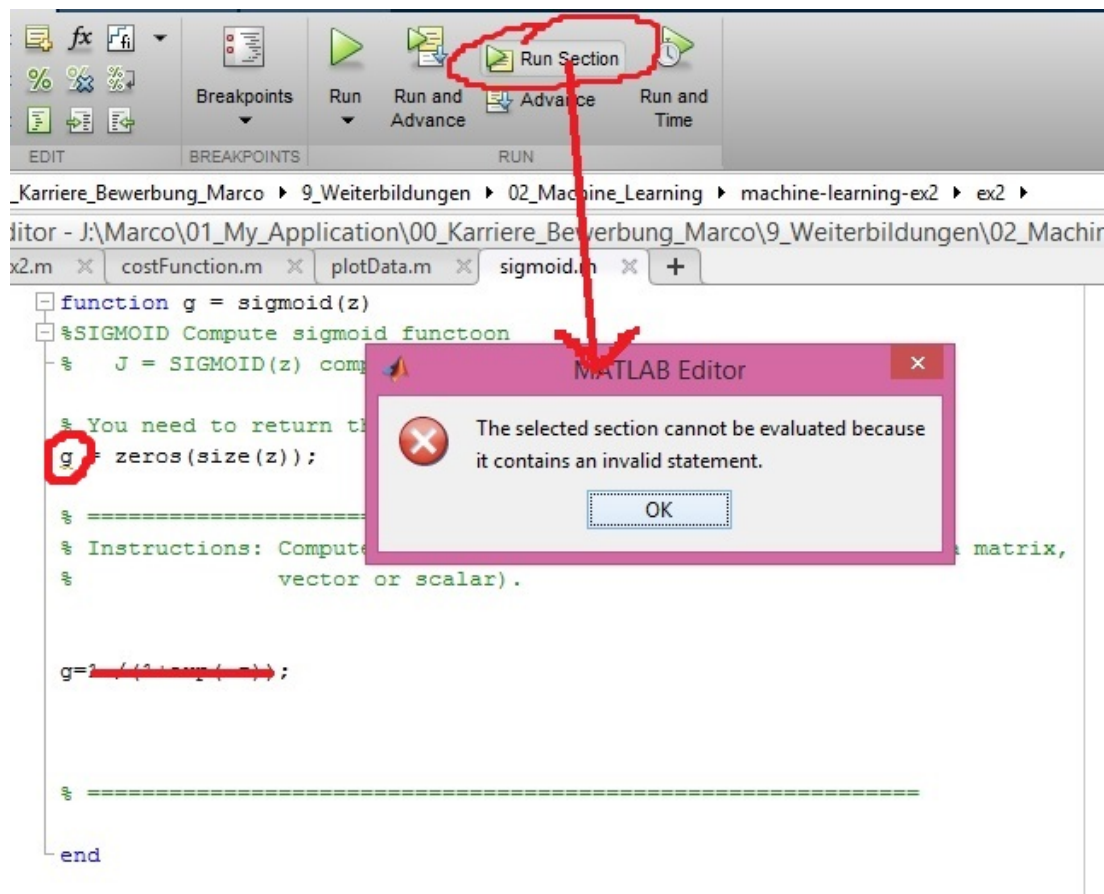
⬆ 7 Upvotes

Marco Moldenhauer · 3 years ago

Hello Tom, I already past the lecture ("compute sigmoid function"). So when I compute this in the comand window: g=sigmoid(0)=0.5 / this is ok!

But I dont understand when I press run section. There is always a error message and "g" is red underlined. Why?

⬆ 0 Upvotes 💬 Hide 2 Replies

Tom Mosher Mentor · 3 years ago

Don't use the run button on a function that requires data parameters. Unless you provide the parameter 'z', it will be undefined when the function runs, and give you that error.

⇧ 0 Upvotes

Tom Mosher Mentor · 3 years ago

I never use the run button at all for this work. I edit the scripts, then go to the command window and run one of the exercise scripts (ex2, ex2_multi, etc). Or I type in the whole function name including some data parameters.

⇧ 1 Upvote

Marco Moldenhauer · 3 years ago

Hallo Tom, I don't understand the last term in the cost function. The sum of theta??? Why we need that. I could find it in the lecture videos :(

⇧ 0 Upvotes   💬 Hide 1 Reply

Tom Mosher Mentor · 3 years ago

That is the regularization term.

The course materials are slightly out of order. The regularization method is part of the four videos that are just after the programming assignment.

⇧ 0 Upvotes

Marco Moldenhauer · 3 years ago · Edited

Hello Tom why the hypothesis function is a vector?

I gues ist a scalar? because
Theta.'*x=Theta(0)*x1^(1)+Theta(1)*x2^(1)+Theta(2)*x2^(1)+Theta(4)*x4^(1) ....
Theta(n+1)*x(n+1)^(1)

⇧ 0 Upvotes   💬 Hide 2 Replies

Tom Mosher Mentor · 3 years ago · Edited

This tutorial teaches the vectorized method.

theta' * x is a scalar. Note that lower-case x is only one training example.

X * theta is a vector. X is the whole matrix of training examples.

⇧ 2 Upvotes

⌄

Marco Moldenhauer · 3 years ago

Thank you Tom

⇧ 0 Upvotes

⌄

Deleted Account · 3 years ago

Deleted Account I got stuck for a while using a double equals sign to set theta(1) equal to zero. A single equals seems to work. Does anyone know when you might use a double equals in Octave?

⇧ 0 Upvotes   💬 Hide 1 Reply

⌄

Tom Mosher   Mentor · 3 years ago

The '==' operator is the logical comparison.

The '=' operator is the assignment operator.

⇧ 2 Upvotes

⌄

Richard Guo · 3 years ago

I am able to calculate the cost function, but how to calculate grad? Is grad the new theta after X y?

⇧ 0 Upvotes   💬 Hide 6 Replies

⌄

Tom Mosher   Mentor · 3 years ago · Edited

See the formula on the top of Page 5 of ex2.pdf. grad is the partial derivative of Cost.

⇧ 0 Upvotes

⌄

Richard Guo · 3 years ago · Edited

Thanks for the quick reply

so its 1/m sum(J) * x (x in here is a vector or a matrix?) or just 1 /m* J * x

is its dimension n x 1(base on X's feature? and same length as theta?)

⇧ 0 Upvotes

Tom Mosher Mentor · 3 years ago

There is a separate tutorial for the gradient.

The dimensions of grad must be the same as theta.

⇧ 1 Upvote

Marilia Nunes Freire Ribeiro · 3 years ago

hello Tom, my sigmoid function passed on the test, then I tried to follow your tips but I made something wrong and couldn't find the mistake in my code yet, could you help me?

First I assumed h as equal to theta' times X;

after this, I calculated the two parts, the red and the blue, where the red is the minus y' times log(h); the blue is the (one minus y) times the log(one minus h);

then, I put the J equal to 1/m times(.*) (the red minus the blue);

Assuming theta(1)=0;

then grad equal to theta' times theta;

after scale equal to the grad times lambda/(2*m);

finally, the cosFunction equal to J+grad;

I assumed in my sigmoid function as g equal to 1/(1 plus e^(-z))) ;

and passed on the test. what could be wrong? Thank you since now.

⇧ 0 Upvotes

Suresh Kandulapati · 3 years ago

Dear Tom,

Thank you so much for your clarifications for doing the assignments.

Please tell me what value should be assigned to "lambda" in this example.

Tom Mosher Mentor · 3 years ago

In costFunctionReg(), the lambda value is passed to the function as a parameter. The script that is calling the function provides the value. That is "ex2_reg.m" in this case.

1 Upvote

Marcio Ribeiro · 3 years ago

Hi,

I have done the sigmoid function as mentioned, and it worked, but I didn't understood. Isn't the sigmoid(z) originated from g(theta' x) ? Why we change the order and don't transpose theta ?

Thanks!

1 Upvote   Hide 3 Replies

Tom Mosher Mentor · 3 years ago

theta and X are not used inside your sigmoid function. The script that calls your sigmoid function handles theta and X.

1 Upvote

Marcio Ribeiro · 3 years ago

I think I got it. The (z) is just the product between X and theta, and without doing like explained above, the matrix dimensions wouldn't fit for multiplication.

Thanks for your help!

0 Upvotes

Tom Mosher Mentor · 3 years ago

Nice work!

0 Upvotes

Dan Jeffrey · 3 years ago · Edited

I am stuck on the cost function. I suspect my limited experience with matrices is at the root of my problem. Why is initial_theta a 3 x 1 matrix and not 1 x 3? I thought it was supposed to represent the coefficients of the hypothesis function. If so why are they arranged in columns?

I believe I want to create a matrix of hypotheses of size 100 x 1 to use in the cost and gradient calculations. I do not see how to do that with the hypothesis function that was given in the exercise instructions: g(transpose(theta) * X) cannot work with matrices of the dimensions we are using. initial_theta is 3 x 1. Transpose(initial_theta) is 1 x 3. Neither of these can be multiplied by X (100 x 3).

Can you see where I might be misunderstanding the matrices?

⇧ 2 Upvotes   💬 Hide 12 Replies

> **Simon Middlemiss · 3 years ago**
>
> Did you solve this? I have come to the same conclusion. Theta as passed into 'costFunction' cannot be transposed and multiplied by X as definition of h(x) calls for.
>
> ⇧ 0 Upvotes

> **Dan Jeffrey · 3 years ago**
>
> Yes. I first figured it out "backwards" by studying the vectors in the programming assignment. Making the matrices fit by dimensions led me to the correct code. Then I studied the meaning of the correct code in terms of the content of those matrices.
>
> The biggest confusion for me came from the variable names, X and x. I was regarding lower case x in the same way as upper case X.
>
> ⇧ 1 Upvote

> **Tom Mosher   Mentor · 3 years ago**
> M
>
> @Simon:
>
> Prof Ng's definition of h(x) holds for x as a vector of one training example. (1 x n) times (n x 1) gives a scalar result.
>
> But if X is the whole matrix of training examples, size (m x n), then the hypothesis is a vector, formed from (m x n) times (n x 1) giving a (m x 1) result.

That is, "h = X * theta", if you are using the vectorized method.

As Dan mentions, dimensional analysis can be very helpful.

⇧ 1 Upvote

Richard Guo · 3 years ago

Hi Tom I am confused? such as the left part of the function we will have -y * log(sigmoid(X * theta)), is this a parallel multiplication(we will get m x 1 result)?

⇧ 0 Upvotes

Tom Mosher Mentor · 3 years ago

X * theta will give a (m x 1) result.

sigmoid() and log() are both element-wise functions, so the return will also be (m x 1).

To multiply by y, you can use element-wise multiplication with the log() result, and then use sum() to get the summation.

You can also use vector multiplication to compute the sum automatically, if you wish. Either method is good.

⇧ 2 Upvotes

Jacob Krajewski · 3 years ago

Tom, things like making vector math work are great, but is it just something we have the freedom to do at any time to just make the math work?

What I mean is that the formula doesn't call for -y' * log(sigmoid(X*h), yet for the sake of matlab, we need to do such operations... What I wonder is to what degree can we simply fuddle around until it works? and why does it work? And doesn't the formula reflect what we write in code? Sorry for the lengthy question, some things seem a bit "well how was I supposed to know to do that from the math, but for the vectors don't match up so I'm forced to?"

⇧ 1 Upvote

⌄

Tom Mosher Mentor · 3 years ago · Edited

The equations we're given are just the math. They aren't written with any specific implementation in mind. We have a very powerful tool that is good at matrix algebra (Octave and MATLAB). The skill of the computer (or machine learning) scientist lies in learning how to change theory into practice.

⇧ 2 Upvotes

Jacob Krajewski · 3 years ago

Good answer. So my skills are in fuddling till it works, right :P j/k

⇧ 0 Upvotes

Tom Mosher Mentor · 3 years ago

This gap in the instruction materials is the reason why the tutorials were written to emphasize the vectorized method.

⇧ 1 Upvote

Jacob Krajewski · 3 years ago

Thank you, the challenge is embraced. I think I was hurried because of an impending deadline!

⇧ 1 Upvote

Preetha Rajan Mentor · 3 years ago · Edited

Tom, thank you very very much, for the incredibly valuable information about the log() function and sigmoid() function, producing element-wise results! I was honestly, at my wits-end with this assignment, as I kept getting a matrix as the 'value' for the cost function J (upon running my codes on various test cases), rather than a scalar! My code works fine, now! :)

⇧ 0 Upvotes

Tom Mosher Mentor · 3 years ago

Nice work!

⇧ 0 Upvotes

I am still at the implement sigmoid function part

sir,

i have used element division,

broken up the code for g(z) in parts,

changed the name to sigmoid(z)

changed it to g as well

but all this time there were two errors:

undefined function or variable z

or

sigmoid needs another argument

i am at my wits end.any help would be greatly appreciated

⇧ 0 Upvotes   ▢ Hide 4 Replies

Tom Mosher   Mentor · 3 years ago

Have you been editing the sigmoid.m script template that came with the exercise scripts?

You did not need to change the name of the function or any of the variables.

⇧ 0 Upvotes

O   Oj · 3 years ago

i really dont think i changed anything in the already written part of the code, but for some unknown reason it is working...sooo hurraayyy!!

although i am a lil stuck on the next part. sir...

do we have to define g(z) again,or call the function g = sigmoid(z) from earlier task?

because when i just set the value of g with parameter of theta and X product,it shows g not rec says that costgrad needs more arguments(this keeps popping up in every ML assignment i do,is it a matlab software problem from my computers end?)

⇧ 0 Upvotes

O Oj · 3 years ago

thank u in advance for the help sir

⇧ 0 Upvotes

M

Tom Mosher Mentor · 3 years ago

It's not a problem in MATLAB. It's a problem in how you are writing the script files.

Once you have written the sigmoid() function and placed it in your working folder, you can call it from any other function. You do not need to write it again.

⇧ 0 Upvotes