coursera

Discussion Forums

Get help and discuss course material with the community.

THIS WEEK'S FORUM

Week 3

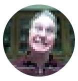
Discuss this week's modules: Logistic Regression & Regularization.

57 threads · Last post 42 minutes ago

Go to forum

Forums | **All Threads** Search

← All Threads



ex4 tutorial for nnCostFunction and backpropagation

Tom Mosher Mentor Week 5 · a year ago · Edited

Keywords: ex4 tutorial backpropagation nnCostFunction

(note: if you have a question about this tutorial, please start a new thread. This one is full and is closed to additional replies)

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

You can design your code for backpropagation based on analysis of the dimensions of all of the data objects. This tutorial uses the vectorized method, for easy comprehension and speed of execution.

Reference the four steps outlined on Page 9 of ex4.pdf.

Let:

Help Center

m = the number of training examples

n = the number of training features, including the initial bias unit.

h = the number of units in the hidden layer - NOT including the bias unit

r = the number of output classifications

1: Perform forward propagation, see the separate tutorial if necessary.

2: δ_3 or d3 is the difference between a3 and the y_matrix. The dimensions are the same as both, (m x r).

3: z2 came from the forward propagation process - it's the product of a1 and Theta1, prior to applying the sigmoid() function. Dimensions are (m x n) ⋅ (n x h) --> (m x h)

4: δ_2 or d2 is tricky. It uses the (:,2:end) columns of Theta2. d2 is the product of d3 and Theta2(no bias), then element-wise scaled by sigmoid gradient of z2. The size is (m x r) ⋅ (r x h) --> (m x h). The size is the same as z2, as must be.

5: Δ_1 or Delta1 is the product of d2 and a1. The size is (h x m) ⋅ (m x n) --> (h x n)

6: Δ_2 or Delta2 is the product of d3 and a2. The size is (r x m) ⋅ (m x [h+1]) --> (r x [h+1])

7: Theta1_grad and Theta2_grad are the same size as their respective Deltas, just scaled by 1/m.

Now you have the unregularized gradients. Check your results using ex4.m, and submit this portion to the grader.

===== Regularization of the gradient ===========

Since Theta1 and Theta2 are local copies, and we've already computed our hypothesis value during forward-propagation, we're free to modify them to make the gradient regularization easy to compute.

8: So, set the first column of Theta1 and Theta2 to all-zeros. Here's a method you can try in your workspace console:

```
Q = rand(3,4)       % create a test matrix
Q(:,1) = 0          % set the 1st column of all rows to 0
```

9: Scale each Theta matrix by λ/m. Use enough parenthesis so the operation is correct.

10: Add each of these modified-and-scaled Theta matrices to the un-regularized Theta gradients that you computed earlier.

You're done. Use the test case (linked below) to test your code, and the ex4 script, then run the submit script.

Here is a link to the test cases, so you can check your work:

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

The test cases for ex4 include the values of the internal variables discussed in the tutorial.

Appendix:

Here are the sizes for the Ex4 digit recognition example, using the method described in this tutorial.

NOTE: The submit grader, the gradient checking process, and the additional test case all use different sized data sets.

a1: 5000x401

z2: 5000x25

a2: 5000x26

a3: 5000x10

d3: 5000x10

d2: 5000x25

Theta1, Delta1 and Theta1_grad: 25x401

Theta2, Delta2 and Theta2_grad: 10x26

131 Upvote · Follow 89 · Reply to Tom Mosher

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

Earliest

Top

Most Recent

Thomas M. Snell · a year ago

Could you post delta1 and delta2 for this example? My J value checks, but my grad values are roughly two times too big.

1 Upvote · Hide 28 Replies

See earlier replies

Mark Flocco · a year ago

Thanks for posting these intermediate values for the test case; despite having my sigmoidGradient function pass the submission, I had a small error which was throwing everything off in the nnCostFunction. This saved me from pulling out the other half of my hair.

1 Upvote

Tom Mosher Mentor · a year ago

Can you be more specific about what error the submit script didn't catch?

0 Upvote

Mark Flocco · a year ago

Sorry, I thought I replied to this earlier. I was too quick to respond. My issue was that I'd taken a shortcut and applied the sigmoid function on my z2, then submitted that to the sigmoidGradient function which resulted in incorrect answers. The intermediate values were very helpful in isolating this issue regardless, and I thank you.

1 Upvote

JF

Jordan Fleming · a year ago · Edited

These are the z2 values provided above are:

```
z2 =
   0.054017   0.166433
  -0.523820  -0.588183
   0.665184   0.889567
```

And these are the z2 values I get:

```
z2 =
   0.51350   0.54151
   0.37196   0.35705
   0.66042   0.70880

```

My results from submission seem to indicate that my forward prop is working so I'm struggling to understand how my z2 values can be so far out.

== Part Name | Score | Feedback

== --------- | ----- | --------

== Feedforward and Cost Function | **30 / 30** | Nice work!

== Regularized Cost Function | 15 / 15 | Nice work!

== Sigmoid Gradient | 5 / 5 | Nice work!

== Neural Network Gradient (Backpropagation) | 0 / 40 |

== Regularized Gradient | 0 / 10 |

== --------------------------------

== | 50 / 100 |

Basically I'm trying to figure out where I went wrong and z2 is, mathematically, the earliest I see my results deviate from from the provided test case.

3 Upvote

Tom Mosher Mentor · a year ago

Are you adding the bias units before you use the sigmoid function?

1 Upvote

Tom Mosher Mentor · a year ago

No, that's not it. Here's the issue:

z2 doesn't include the sigmoid function at all.

4 Upvote

JF

Jordan Fleming · a year ago · Edited

Are you really still awake? Or am I talking to a sophisticated ANN? :)

Yes I am adding the bias units. I'm using the for loop method of implementation so I am syntactically pre-pending a column [1 X(i,:)]. I'm trying to not be too revealing with code so please let me know if I should edit this.

And I know I have to disregard the bias units during backprop so I am using the syntax 2:end that you've mentioned elsewhere.

0 Upvote

JF

Jordan Fleming · a year ago

I'd like to say that I've been staring at it for too long but I just checked and I made the same mistake in my ex3. Thanks for pointing that out.

0 Upvote

Tom Mosher Mentor · a year ago

So are you OK now?

0 Upvote

JF

Jordan Fleming · a year ago

Hey, sorry for the belated response. You were correct about my use of the sigmoid function and that change gave me the correct implementation.

1 Upvote

Tom Mosher Mentor · a year ago

Cool.

0 Upvote

CS

Clara Giner Sanfrancisco · 10 months ago

Dear Tom, my **nnCostFunction** is calculating everything right with the exception of d2, and because of this error Delta1 and Theta1_grad. Instead of using a for loop, I've used matrix products. I am a bit lost with d2. Here you can see my results for the test case with regularization:

d3 =

0.8887 0.9074 0.9233 -0.0634

0.8382 -0.1397 0.8798 0.8969

0.9234 0.9386 -0.0491 0.9609

d2 =

0.7444 0.9860

```
0.7618 1.0036

0.7689 1.0020

a2 =

1.0000 0.5135 0.5415

1.0000 0.3720 0.3571

1.0000 0.6604 0.7088

a3 =

0.8887 0.9074 0.9233 0.9366

0.8382 0.8603 0.8798 0.8969

0.9234 0.9386 0.9509 0.9609

Delta1 =

2.2751 -0.1339 -0.0694

2.9917 -0.1766 -0.1043

Delta2 =

2.6503 1.3779 1.4350

1.7063 1.0339 1.1068

1.7540 0.7689 0.7793

1.7944 0.9357 0.9670

J =

19.4736

grad =

0.7584

0.9972

0.3554

0.4745
```

```
0.6435

0.7652

0.8834

0.5688

0.5847

0.5981

1.9260

1.9446

1.9896

2.1786

2.4783

2.5023

2.5264

2.7223
```

0 Upvote

CS

Clara Giner Sanfrancisco · 10 months ago

Ok, I've solved it, in the d2 equation I was using g'(a2(:,2:end)) instead of g'(z2). Thank you very much for all you help, and for providing all these helpful test cases.

0 Upvote

Tom Mosher Mentor · 10 months ago

Good catch!

0 Upvote

Min Yang · 9 months ago

Hi,

I used this test case, what might be wrong?

```
il = 2;              % input layer
hl = 2;              % hidden layer
nl = 4;              % number of labels
nn = [ 1:18 ] / 10;  % nn_params
X = cos([1 2 ; 3 4 ; 5 6]);
y = [4; 2; 3];
lambda = 4;
[J grad] = nnCostFunction(nn, il, hl, nl, X, y, lambda)

J =

   19.4736


grad =

    0.7661
    0.9799
    0.3725
    0.4975
    0.6417
    0.7461
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0
         0

```

0 Upvote

Simon Crase · a year ago

Roughly? Or exactly? Remember what differentiation does to the 0.5 in the cost.

0 Upvote · Reply

Thomas M. Snell · a year ago

Good thought, but no, mine are only roughly two times the grad values Tom Mosher shows in his Test Case 1.

0 Upvote · Reply

Daniel Mulally · a year ago

I seem to be getting the cost function but my d2 isn't coming out right and my Delta matrices aren't right. d3 is OK. When I submit my results I get credit for the first 3 items. Can you please send me Z2 and the sigmoid gradient of Z2 for your example?

== Part Name | Score | Feedback

```
== --------- | ----- | --------

== Feedforward and Cost Function | 30 / 30 | Nice work!

== Regularized Cost Function | 15 / 15 | Nice work!

== Sigmoid Gradient | 5 / 5 | Nice work!

== Neural Network Gradient (Backpropagation) | 0 / 40 |

== Regularized Gradient | 0 / 10 |
```

Results of your example:

```
J =

19.473636522732420

grad =

0.768228339819711

0.983978840331851

0.366105956425469

0.483456740910655

0.681001650553878

0.831520601089803

0.883417207679397

0.568762344914512

0.584667662135129

0.598139236978449

1.924278118545129

1.943092629583105

2.037269588504872

2.128392202547797

2.476398591555435
```

2.500516724398037

2.580661406705335

2.665218559809105

0 Upvote · Reply

Tom Mosher Mentor · a year ago

z2: (truncated to three decimal places)

```
 0.054  0.166
-0.523 -0.588
 0.665  0.889
```

sigmoidGradient(z2):

```
0.249  0.248
0.233  0.229
0.224  0.206
```

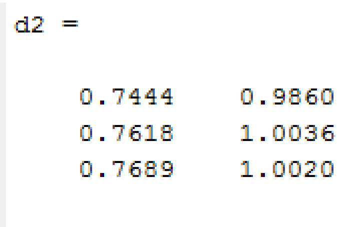
3 Upvote · Hide 18 Replies

See earlier replies

BR

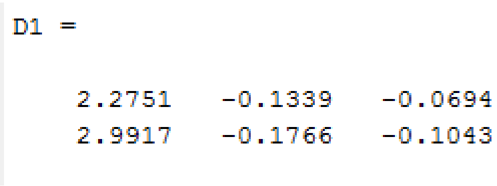
Babalola Rotimi · a year ago

Hi Tom, when I use the test case the values I get for d3 and D2 are correct but d2 and D1 are wrong. What could be causing this problem?



```
d2 =

    0.7444    0.9860
    0.7618    1.0036
    0.7689    1.0020
```



```
D1 =

    2.2751   -0.1339   -0.0694
    2.9917   -0.1766   -0.1043
```

0 Upvote

Tom Mosher Mentor · a year ago

This is most often a problem with how you are handling the bias column of Theta2.

1 Upvote

BR

Babalola Rotimi · a year ago

Yes I have solved the problem. I was using sigmoid(Z2) instead of sigmoidGradient(Z2) to compute d2. Thanks for your continued assistance

2 Upvote

NK

NISHA KHULBE · a year ago

Tom, I am getting every value as per your test cases for deltas also but still when I submit my code I am not getting grades for Neural Network backpropagation and regularization. Please help. z2 / delta1/2 are coming correct

1 Upvote

NK

NISHA KHULBE · a year ago

Got the problem, there is a slight change in d2.. Thanks for the

0 Upvote

Drozhnikov Alexander · 6 months ago

I compute d3*Theta2*g'(a2), then cut first column from result. Where ia a mistake? Help me please. I lost 2 hours, but without result

0 Upvote

Tom Mosher Mentor · 6 months ago · Edited

Use only the "(:,2:end)" columns of Theta2, and then element-wise multiplication by the sigmoid gradient.

And you should use the sigmoid gradient of z2, not a2. z2 doesn't include the hidden layer bias units, which is good because we don't want to backpropagate those since they do not connect to the input layer.

0 Upvote

Drozhnikov Alexander · 6 months ago

Thanks for the help!

0 Upvote

Drozhnikov Alexander · 6 months ago

I complied the cod and try to test. My results are re arrange

```
J = 19.474

grad =



0.76614

0.37246

0.64174

0.97990

0.49749

0.74614

0.88342

1.92598

2.47834

0.56876

1.94462

2.50225

0.58467

1.98965

2.52644

0.59814

2.17855

2.72233
```

Whats the problem?

0 Upvote

Tom Mosher Mentor · 6 months ago

Which test case are you using?

0 Upvote

Drozhnikov Alexander · 6 months ago

https://www.coursera.org/learn/machine-learning/module/Aah2H/discussions/uPd5FJqnEeWWpRIGHRsuuw

0 Upvote

Drozhnikov Alexander · 6 months ago

I done it. There was a mistake in Delta1&2

0 Upvote

Tom Mosher Mentor · 6 months ago

I presume you're using the regularized test case.

Do you get the correct gradients for the un-regularized case? It's in the same post.

0 Upvote

Tom Mosher Mentor · 6 months ago

That test case also includes data for all of the variables inside the function. You can set a breakpoint in your code and inspect the values.

0 Upvote

Tom Mosher Mentor · 6 months ago

Good news!

0 Upvote

L

lanlu · a year ago

Hi,

I have

delta_2 (d2)=

0.7939 1.0528

0.7367 0.9513

0.7677 0.9356

delta_3 (d3)=

0.8887 0.9074 0.9233 -0.0634

0.8382 -0.1397 0.8798 0.8969

0.9234 0.9386 -0.0491 0.9609

right. But,

D2 = delta_3' * a2

D2 =

1.9375 1.3779 1.4350

1.2474 1.0339 1.1068

1.2823 0.7689 0.7793

1.3118 0.9357 0.9670

is not right, while

D1 = delta_2' * a1

D1 =

2.2984 -0.0826 -0.0748

2.9397 -0.1075 -0.1616

is right. I am confused now. Hope you could help me with this.

👍 0 Upvote · Hide 2 Replies

L

lanlu · a year ago

I've found the problem, thx.

0 Upvote

CAMARA Mamoudou · a year ago

Thank you

0 Upvote

Angelina Yang · a year ago

hi Tom, just want to thank you. This post is very helpful. :-)

1 Upvote · Hide 2 Replies

Tom Mosher Mentor · a year ago

I'm glad it was useful.

0 Upvote

Angelina Yang · a year ago

hi Tom,

Btw, I am confused about how the error term is calculated of the final layer. In the final layer, the calculated probabilities are deducting the y vector consisted of zeros and one. Even for a single logistic regression, we don't compute prediction errors this way. So this deduction doesn't really make sense to me intuitively.

Just wondering if you have any insights on this?

Thank you very much!

Yang

0 Upvote

BD

Brad Deutsch · a year ago

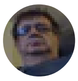
This was extremely helpful. The assignment strongly advocates doing the back-propagation using a for loop, which ended up seriously confusing be because the matrix sizes didn't make intuitive sense. This cleared everything up. Thanks!

4 Upvote · Hide 3 Replies

Tom Mosher Mentor · a year ago

Thanks. Using for-loops makes sense for Prof Ng's target audience (rudimentary programmers who don't know matrix algebra). For everyone else, for-loop are confusing and complicated.

1 Upvote



Harry Lewin · a year ago

I am extremely grateful for this tutorial. You provided an elegant and intuitive solution and the test cases needed to debug it. (Free from loopy confusion!) Many thanks.

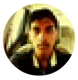
0 Upvote

BQ

Brian Quinif · 6 months ago

Another +1 for the vectorized approach over the loop approach. I'd add that the loop itself is not so confusing to me but rather what the underlying elements are. I guess I spent enough time in Econometrics classes a long time ago that beta_hat = (X'X)^(-1)*X'y is forever ingrained in my memory.

0 Upvote



Mohammed Ashmil · a year ago

Using the above test case

i get the following result

J =

19.4736

grad =

0.7661

0.9799

0.3725

0.4975

0.6417

0.7461

0.8834

0.5688

0.5847

0.5981

1.5411

1.8488

1.5855

1.9220

2.1588

2.4894

2.1308

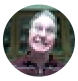
2.5055

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

the first few of gradients are right but others are not!

can someone please figure it out.

0 Upvote · Reply



Tom Mosher Mentor · a year ago

No, solving the problem is your job as a student.

The first few values in grad are from Theta1.

The rest of the values are from Theta2.

0 Upvote · Reply

KK

Karen Krohne · a year ago · Edited by moderator

Hi Tom,

I've been stuck with the backpropagation for quite a while, i think my implantation of the error function J works fine, since its giving me the correct results above,

When checking my gradient the relative difference is still too big, around 0.018071

My code is;

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

.... Could you maybe help me, giving a clue what I am doing wrong?

Thank you so much for you help so far, you tutorials are a great help !

Best

Karen

0 Upvote · Hide 3 Replies

Tom Mosher Mentor · a year ago

@Karen,

Sorry, students aren't allowed to post their program code. That is a violation of the course Honor Code. I have edited your post.

These equations are a little tricky, be sure you're using enough sets of parenthesis to keep the order of execution correct.

0 Upvote

KK

Karen Krohne · a year ago

@Tom,

Hi tom, sorry, Ive been totally of the course for a while, but i am still stuck with the same problem. Ive been trying to check the the parenthesis, and I dont think thats the problem.

I've been trying to run your test case, but i get the following (regularized)

```
J =

21.4670

grad =

1.6886

2.3296

0.1206
```

```
0.2291

0.2417

0.3874

0.8834

0.5688

0.5847

0.5981

0.4593

0.3446

0.2563

0.3119

0.4783

0.3689

0.2598

0.3223
```

I really dont know whats wrong. The J functions were giving the right values in the previous parts of the siignment.

You write that for many its a problem with using theta 2 correctly. When i am calculating my d2 , i use Theta2(: , 2:end) - is that wrong?

I hope you can help me, and that i am not violating the Honor code this time .

Thank you so so much for taking you time, it is such a great help you are giving!(i just wish i was bette in finding my own mistake ;))

Best

Karen

0 Upvote

Tom Mosher Mentor · a year ago · Edited

On Page 2 or Page 3 of this thread, you can find a post that gives the values you should have for all of the variables inside your cost function for this test case. Set a breakpoint in your cost function, just before it returns, and compare your values vs the ones in that post.

0 Upvote

JS

James Singleton · a year ago

if J is being calculated correctly, then it must be due to calculation of the backpropogation gradients: have a look at your formulae, there may be a bug in there!

0 Upvote · Reply

JS

James Singleton · a year ago

@ Tom Mosher, I was mucking about with For loops until I found this thread, per the instructions in the pdf...its almost like on this exercise its been designed to force you to look to the discussion boards?!

0 Upvote · Hide 6 Replies

Tom Mosher Mentor · a year ago · Edited

I think the deal is that Prof Ng feels that for-loops are more intuitive for novice programmers, and those with a math background will know to implement vectorized methods (it is mentioned in a couple of the PDF files, and in one of the video lectures). There is a gap regarding experienced programmers who automatically think in terms of loops. The tutorials exist to plug the gap.

2 Upvote

AG

Anil Gupta · 10 months ago

Thanks Tom for your tutorials and test cases. After two days of tizzy head the stuff went through the grader, whew. After doing the matrix I would feel too lazy to go into looping loops. The matrix is better visualization of weights. The rows connect to the receiving layer and the columns connect to the sending layer; and you all one column for the additional bias node. Thanks.

0 Upvote

Tom Mosher Mentor · 10 months ago

Nice work.

0 Upvote

GK

Gautam Karmakar · 8 months ago

Hi Tom, in the tutorial it says delta is d2 multiplied by a1 transpose but I think that would be wrong as mxr and rxh resulting mxh as we needed. can you let me know where am I wrong here.

0 Upvote

Tom Mosher Mentor · 8 months ago

The tutorial is correct.

It doesn't tell you what transpositions to use, you have to figure those for yourself.

0 Upvote

XL

Xinghou Liu · 3 months ago

What I found is that: the for-loop is actually more complicated and difficult to implement, while using the vectorised approach is more clear and easy to follow.

0 Upvote

JF

Jonas Fleischer · a year ago

Dear Tom,

Im very very grateful for this tutorial.

I still had to struggle a bit because of a stupid typo and or brain dysfunction where I accidentally set the whole second column of Theta2 to 0.

It can get rather frustrating when all the test values you provide seem to fit (at least all before the regularization) but its still wrong. Anyways what this is all about.

Thank you Mr Mosher.

2 Upvote · Reply

Tom Mosher Mentor · a year ago · Edited

I have updated the tutorial to include a link to the test cases (since the Forum hides things when the replies to a post occupy more than one page).

0 Upvote · Reply

‹ 1 2 3 4 ... 8 ›

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

TM M YM M KK M VA M M MS M M GN M

MS M M MA M M MR M CU M XD M JN M FV M AN I DS M

JM M M M M

Learn more about becoming a Mentor

Forum guidelines ›