Programming Assignment: K-Means Clustering and PCA

Deadline   Pass this assignment by September 16, 11:59 PM PDT

Instructions

My submission

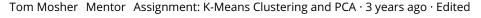**Discussions**

← Assignment: K-Means Clustering and PCA

Tutorials for ex7_pca

Tom Mosher   Mentor   Assignment: K-Means Clustering and PCA · 3 years ago · Edited

These are tutorials for all three functions in the ex7 PCA exercise. All of these functions have a vectorized implementation in one or two lines of code.

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

pca()

Compute the transpose of X times X, scale by 1/m, and use the svd() function to return the U, S, and V matrices.

X is size (m x n), so "X transpose X" and U are both size (n x n)

(note: the feature matrix X has already been normalized, see ex7_pca.m)

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

projectData()

Errata:

In projectData.m, make the following change in the Instructions section:

% projection_k = x' * U(:, **1:k**);

coursera

Return Z, the product of X and the first 'K' columns of U.

X is size (m x n), and the portion of U is (n x K). Z is size (m x K).

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

recoverData()

Return X_rec, the product of Z and the first 'K' columns of U.

Dimensional analysis:

- The original data set was size (m x n)

- Z is size (m x K), where 'K' is the number of features we retained.

- U is size (n x n), where 'n' is the number of features in the original set.

- So "U(:,1:K)" is size (n x K).

So to restore an approximation of the original data set using only K features, we multiply (m x K) * (K x n), giving a (m x n) result.

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

keywords: tutorial ex7_pca

⇧ 18 Upvotes   💬 Reply   Follow this discussion

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

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

Vivek Indrayan · 2 years ago   ⌄

I've finished all the assignments in Ex 7 and received grades for them all. But when I run ex7_pca.m and the script proceeds to the part for dimension reduction of the image (of faces), the script fails as follows:

Dimension reduction for face dataset.

The projected data Z has a size of: 5000 1

Program paused. Press enter to continue.

Visualizing the projected (reduced dimension) faces.

operator *: nonconformant arguments (op1 is 5000x1, op2 is 100x1024)

error: called from 'recoverData' in file /Users/vk2/Dropbox/Coursera/machine-learning-ex7/ex7/recoverData.m near line 24, column 7

error: called from:

error: /usr/local/octave/3.8.0/share/octave/3.8.0/m/miscellaneous/run.m at line 80, column 5

What could I be doing wrong?

⇧ 0 Upvotes   💬 Hide 1 Reply

Tom Mosher Mentor · 2 years ago · Edited

In recoverData() at that point in the ex7_pca.m script, here are the sizes of the variables:

X: 5000 x 1024

U: 1024 x 1024

K: 100 (scalar)

I'm not sure where you're getting a 5000 x 1 operand, but that's likely a problem.

Also, you should be using the columns of U, so when you reduce it to K features, its size is (1024 x 100), not (100 x 1024).

⇧ 0 Upvotes

Zach Brown · 2 years ago

I received credit from the submit script for everything except for projectData.m and recoverData.m. I've completed those two scripts. Now when I run ex7_pca I get the expected results for the first part, but I get the following error for the k-means part:

Index exceeds matrix dimensions.

Error in plotDataPoints (line 12)

scatter(X(:,1), X(:,2), 15, colors);

Error in ex7_pca (line 232)

plotDataPoints(Z(sel, :), idx(sel), K);

Is it possible that I messed something up in one of my k-means scripts even though they were all accepted by the grader? It doesn't seem like this error is coming from projectData or recoverData, but maybe it is?

⇧ 0 Upvotes   💬 Hide 1 Reply

Tom Mosher   Mentor   · 2 years ago

Passing the submit grader doesn't mean your code is perfect. It only means that is passes the test case that the grader uses.

It is common to find bugs in code that the submit grader has passed.

We have additional test cases to help with this. See this thread:

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

⇧ 0 Upvotes

Shivam Sharma · 2 years ago

Hello Tom,

Can't seem to figure out recover_data.m.... i multiplied Z with U' for every row of U' but can't seem to get the answer .... keep getting error as

Subscripted assignment dimension mismatch.

plz help

⇧ 0 Upvotes   💬 Hide 6 Replies

Tom Mosher   Mentor   · 2 years ago

Why the rows?

⇧ 0 Upvotes

Shivam Sharma · 2 years ago

the rows were earlier the coloumns right so that's why

⇧ 0 Upvotes

Tom Mosher Mentor · 2 years ago

I believe the tutorial says to use the columns.

⇧ 0 Upvotes

Shivam Sharma · 2 years ago

yeah so I tried that too but the answer is not coming... like look i figured Z is (m x K) and U is (n x K) so Z * U' for the first K rows should do the trick... but it isn't.... so where am I going wrong?

⇧ 0 Upvotes

Tom Mosher Mentor · 2 years ago

The derived features are in the columns of U - not the rows. Multiplying by the rows of U just gets you scrambled eggs.

⇧ 0 Upvotes

Shivam Sharma · 2 years ago

no no I was multiplying by row of U' which were columns of U so as to get my X_rec as size (m x n)....

but now I have overcome that got new idea and its working.... thanks anyway... but just can't get what is the error in my prior way to solve it

⇧ 0 Upvotes

Arturo Jesús Laflor Hernánez · 2 years ago

Hi Tom, I wrote the code into projectData.m. I did the computation of times X(mxn) and U(nxK) to obtain Z(mxK) in one line of code. I have the result:

Dimension reduction on example dataset.

Projection of the first example: 1.481274

(this value should be about 1.481274)

but... when I do the submission, the response is the same:

Project Data (PCA) | 0 / 10 |

any suggestions on this?

Thank you.

⇧ 0 Upvotes   💬 Hide 9 Replies

Tom Mosher   Mentor   · 2 years ago

The submit grader uses a test set with many features. Verify that your projectData() function works with any size of feature set.

⇧ 1 Upvote

Tom Mosher   Mentor   · 2 years ago

Try the additional test case for this function. Use this thread to find it:

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

⇧ 1 Upvote

Arturo Jesús Laflor Hernánez · 2 years ago

Thank you Tom, I'll verify my code.

⇧ 0 Upvotes

Arturo Jesús Laflor Hernánez · 2 years ago

Thank you a lot Tom. I could see my mistake by using the test cases. My error was that instead of a range for the second parameter of matriz U, I only set a number K.

I could not see my novice octave mistake :-).

Thank's.

David · 2 years ago

Instructions should say projection_k = x' * U(:, **1:k**); rather than projection_k = x' * U(:, **k**);. Thanks for your comments. It helped me realised I made the same mistake.

⇧ 0 Upvotes

Tom Mosher Mentor · 2 years ago

Which instructions are incorrect? I'll fix it if I can, or add it to the Course Wiki.

⇧ 0 Upvotes

David · 2 years ago

In projectData.m

% ===================== YOUR CODE HERE =====================

% Instructions: Compute the projection of the data using only the top K

% eigenvectors in U (first K columns).

% For the i-th example X(i,:), the projection on to the k-th

% eigenvector is given as follows:

% x = X(i, :)';

% projection_k = x' * U(:, k);

%

Thanks

⇧ 1 Upvote

Tom Mosher Mentor · 2 years ago

Thank you.

⇧ 0 Upvotes

Zhuochuan Huang · 2 years ago

The original instruction looks correct to me as projection_k refers to a projection on k and Z is a vector of K of such projections: projection_1, ... , projection_K.

⇧ 0 Upvotes

PR **Paul rogers** · 2 years ago

trouble running ex7_pca.m

getting following error:

octave:29> ex7_pca.m

error: invalid call to script /Users/paulrogers/Desktop/MeFein/Machine Learning/ex7/ex7_pca.m

any suggestions ? Note ex7.m runs fine.

⇧ 0 Upvotes   ▢ Reply

DI **daniele iovinelli** · 3 years ago

Hi! My exercise works fine in project_data and recover_data, but when I make the submission it says :

!! Submission failed: unexpected error: operator *: nonconformant arguments (op1 is 15x5, op2 is 1x11)

!! Please try again later.

⇧ 0 Upvotes   ▢ Hide 7 Replies

Tom Mosher Mentor · 3 years ago · Edited

The submit grader uses a test case with 11 features. Your code needs to work with any number of features.

⇧ 0 Upvotes

DI **daniele iovinelli** · 3 years ago

so I have to change K??

Tom Mosher   Mentor   · 3 years ago

Your code needs to automatically work with any size of data set.

⇧ 0 Upvotes

DI   daniele iovinelli · 3 years ago

OK! I'll try to improve it! Thanks

⇧ 0 Upvotes

DI   daniele iovinelli · 3 years ago

The algorithm works fine in test exercise, but there are small differences in the numbers of recover data function..and the same happens with the all_theta exercise..the matrices are well in dimensions ,but there are small differences in number results...

⇧ 0 Upvotes

DI   daniele iovinelli · 3 years ago

ok I have result the problem...was only a trasposition problem!

⇧ 0 Upvotes

Tom Mosher   Mentor   · 3 years ago

Thanks for your report.

⇧ 0 Upvotes

K   ku21fan · 3 years ago

feeling weird.. about the size of U = (n x k)

What I learned about SVD is

if X is size (m x n) , then U (m x k) S (k x k) V' (k x n)

so X = U * S * V'

From this, I thought U should be (m x k) instead of (n x k) .

whatever... I did ex7_pca by using the fact ' U (n x k) '

Just wondering .. !

⇧ 1 Upvote   💬 Reply

Tom Anderson · 3 years ago · Edited

In the lecture, it is given to "always perform mean normalization" with the meaning "Replace each $x_j^{(i)}$ with $x_j^{(i)} - \mu_j$"

At time 0:23 in Week 8's video on Pcincipal Component Analysis Algorithm
https://www.coursera.org/learn/machine-learning/lecture/ZYIPa/principal-component-analysis-algorithm

The MATLAB cov function does this demeaning, as well as all the other tutorials I have found. Took me a while to figure out why my answer was off. Why don't we do this preprocessing step?

In other words, why aren't your instructions, instead of "Compute the transpose of X times X", something like "Compute newX = X - avg(X) and only then compute the transpose of newX times newX..."

```
1  |
```

⇧ 0 Upvotes   💬 Hide 5 Replies

Ezra Darshan · 3 years ago

The script ex7_pca.m already does this normalization - see line 49

⇧ 3 Upvotes

Tom Mosher Mentor · 3 years ago · Edited

Thanks Ezra.

I will add a note that clarifies this point.

⇧ 1 Upvote

Ian Ferreira · 3 years ago

What does this mean

warning: your version of GraphicsMagick limits images to 8 bits per pixel

How to address?

⇧ 0 Upvotes

Tom Mosher   Mentor   · 3 years ago

It's just a warning. You don't have to do anything.

⇧ 0 Upvotes

Ocxs · 2 years ago

Thanks!

⇧ 0 Upvotes

‹ | 1 | ›