APM120, quiz #2 solutions
Applied Linear Algebra and Big Data
Last updated: March 28, 2017

your name: ______________________________

Read these instructions carefully: Please solve all problems, **deriving, calculating and showing explicitly all stages of your solution and explaining each step of each question**. The number of points for each question is noted below, the total number of points is 115 and the final score is min(100, your points). Use the space below and attach additional pages if needed. Use a non-programmable calculator to convert your answers to decimal number format, carrying out calculations to four significant digits. Whenever relevant, use the numerical checks provided to verify your solution.

Start time: 7:00, end time: 9:00.

Good luck!

1. (35 pts) Consider $A\mathbf{x} = \mathbf{b}$ for A and $\mathbf{b}$ given below.

   (a) Explain what the pseudo inverse $A^\dagger$ of a matrix A is, and write and explain the expression for it. Why is it called a pseudo inverse? Why is it not a proper inverse of the matrix?

   (b) Do you expect to find a unique solution that satisfies these equations? Explain. If not, what would you need to assume to find one? Write down an expression for the *general* solution for $\mathbf{x}$ in such a case based on the additional assumption you made.

   (c) Solve for $\mathbf{x}$ given the specific $A, \mathbf{b}$ given below, using the approach you outlined in the previous section.

```
A=[ 1   2  3
    2  -2  1 ];
b=[ 1
    2 ];
```

Hint: The U, and Σ of SVD of A are given by,

```
[U,S,V]=svd(A);
U=[ -0.982  -0.189
    -0.189   0.982 ];
S=[ 3.767  0      0
    0      2.968  0 ];
```

Solution:

(a) The pseudo inverse $A^\dagger = V\Sigma^\dagger U^T$ where $\Sigma^\dagger$ has the dimensions of the transposed of Σ, with values on the diagonal only, equal to the inverse singular values. It is called a pseudo inverse because for an underdetermined problem, $AA^\dagger = I$. Because $A^\dagger A \neq I$, this is not a proper inverse.

(b) The problem is under-determined, so one expects an infinite number of solutions. To find a unique solution, we require the solution to have the smallest norm possible. The solution satisfying these requirements is $\mathbf{x} = A^\dagger \mathbf{b}$.

(c) Given the U, and Σ, we calculate the first two V vectors using $A = U\Sigma V^T$, so $A^T = V\Sigma^T U^T$, so $A^T U = V\Sigma^T$, or $A\mathbf{u}_i = \sigma_i \mathbf{v}_i$. The third one does not matter because in the expression for the pseudo inverse it is multiplied by a zero singular value, so set it to zero,

```
V(:,1)=A'*U(:,1)/S(1,1);
V(:,2)=A'*U(:,2)/S(2,2);
V(:,3)=[0 0 0]';
my_fprintf_array(V);
V=[ -0.361046  0.598035  0
    -0.42091  -0.789199  0
    -0.832154  0.139714  0 ];
Sdagger=[ 1/S(1,1)  0        0
             0      1/S(2,2)  0 ]';
Adagger=V*Sdagger*U';
x=Adagger*b;
r=A*x-b;
my_fprintf_array(Sdagger); my_fprintf_array(Adagger); my_fprintf_array(x);
Sdagger=[ 0.265  0
          0  0.337
          0  0 ];
Adagger=[ 0.056  0.216
          0.16  -0.24
          0.208  0.0880 ];
x=[ 0.488
    -0.320
    0.384 ];
```

and given the accuracy, the residual is zero so we found a solution.

2. (35 pts) Consider the following two data sets representing the size of two mountain glaciers in Switzerland (X) and of two glaciers in the Canadian Rockies (Y) during $N = 5$ years (the time average has been subtracted off, and you do not need to normalize by the standard deviation in this case). Analyze the relation between the two using **multivariate PCA** (not MCA) according to the following guidelines.

(a) Write down a single data matrix that combines both data sets as $F = \begin{pmatrix} X \\ Y \end{pmatrix}$, or, in Matlab's terms, `F=[X;Y]`. No need to remove mean or normalize by std. The covariance matrix of F is given below, how is it calculated from F? Interpret each of its elements.

(b) Given the SVD of the combined data matrix given below, interpret the connection between the two data sets.

(c) Write down a formula for the total variance of X, what does it represent? Write down a formula for the total covariance of X and Y, what does it represent? Can the total covariance vanish while the total variance of X or Y is not zero? Can the total variance of X or Y vanish while the total covariance is not zero?

(d) Reconstruct the data using the first PC only. Calculate the total variance of the original data and of the reconstructed data directly from the corresponding data matrices. Use these total variances to calculate the fraction explained by the first mode, and compare to that predicted by the singular values.

```
X=[ -1  2  0  -2  1
    3  4  0  -4  -3 ];
Y=[ 6  8  0  -8  -6
    -2  4  0  -4  2 ];
```

Hint: The SVD of F is given by

```
F=[X;Y];
%% covariance matrix of F:
C=[ 2  2    4  4
    2  10  20  4
    4  20  40  8
    4   4   8  8 ];
[U,S,V]=svd(F);
U=[ -0.103 -0.435  -0.874   0.189
    -0.435  0.103  -0.189  -0.874
    -0.871  0.205   0.095   0.437
    -0.205 -0.871   0.437  -0.095 ];
S=[ 16.18  0      0  0  0
    0      6.18  0  0  0
    0      0      0  0  0
    0      0      0  0  0 ];
V=[ -0.372   0.602  0.539  -0.457  0
    -0.602  -0.372  0.457   0.539  0
    0        0      0       0     -1
    0.602    0.372  0.457   0.539  0
    0.372   -0.602  0.539  -0.457  0 ];
```

Solution:

(a) The data matrix and covariance matrix:

```
F=[X;Y]; C=F*F'/5; my_fprintf_array(F); my_fprintf_array(C);
F=[-1  2  0  -2  1
    3  4  0  -4 -3
    6  8  0  -8 -6
   -2  4  0  -4  2 ];
C=[ 2  2  4  4
    2 10 20  4
    4 20 40  8
    4  4  8  8 ];
```

the covariance matrix indicates that all variables are positively correlated. The maximum variance is of variable 3, the smallest of variable 1.

(b) The columns of U are the PCs, only two of them matter because only two singular values are nonzero. The first PC has a uniform sign to all elements and therefore corresponds to a co-variability of all of the variables, this is the dominant variability, explaining why the covariance matrix is positive throughout. The second PC corresponds to variability mode where the 1st and 4th variables are anti correlated with the 2nd and 3rd. It is less dominant as seen by the corresponding singular value.

(c) The total variance of X is given by the sum of the diagonal elements of $C_X = XX^T/N$. It represent a measure of the amplitude of the variability of both variables in X. The total covariance of X and Y: let $C = X * Y^T/N$ and the total covariance is $\sum_{i,j} C_{ij}^2$. This represents a measure of to what degree the two data sets vary together. It is possible for the total covariance to vanish while the total variance is not zero if the two data sets vary independently of each other. The opposite situation is not possible, because if the variance vanishes there is also no variability and therefore no correlation.

(d) Reconstruct the data using the first PC only,

```
F1=U(:,1)*S(1,1)*V(:,1)'; my_fprintf_array(F1)
F1=[ 0.618034  1        -0  -1        -0.618034
     2.61803   4.23607  -0  -4.23607  -2.61803
     5.23607   8.47214  -0  -8.47214  -5.23607
     1.23607   2        -0  -2        -1.23607 ];
```

Calculate the total variance of the reconstructed data

```
sum(diag(F1*F1'/5))
   52.3607
sum(diag(F*F'/5))
    60
% alternatively and equivalently, no need to calculate the
% full covariance matrix, one can instead,
sum(sum(F1.^2,2))/5
=  52.3607
```

Calculate the fraction explained by the first mode,

```
52.3607/60= 0.8727
```

and compare to that predicted by the singular values,

```
S(1,1)^2/sum(diag(S).^2)
 = 0.8727
```

3. (35 pts) Consider the following sets of items bought by $N = 6$ different customers on eBay.

   (a) Find the frequent items, pairs of items and triplets of items, and their support, based on a threshold support of $s \geq 0.5$.

   (b) Explain what are association rules, how confidence and interest of such rules are calculated and what they represent. Give an example in which a rule with negative interest could be useful.

   (c) Find the association rules of the form $\{i, j\} \rightarrow k$ that can be deduced from frequent triplets and calculate their confidence and interest.

   (d) Write the data matrix for the N sets given below. Find the Jaccard similarity of baskets 1 and 2.

```
1  {cashews,orange,avocado,tomato}
2  {orange,tomato,chocolate,pepper,cashews}
3  {cashews,orange,avocado,lemon}
4  {apple,tomato}
5  {cashews,avocado}
6  {orange,tomato,cashews,chocolate,pepper}
```

Solution:

   (a) Start with finding the support (percent occurrence) of individual items, (multiply each value by $100/N$ to get percentage)

```
tomato: 4
pepper: 2
orange: 4
lemon: 2
cashews: 5
apple: 1
avocado: 3
```

   Frequent items are:

```
tomato
orange
cashews
avocado
```

Possible frequent pairs and their counts are, therefore,

```
cashews,avocado: 3
cashews,orange: 3
cashews,tomato: 3
avocado,orange: 2
avocado,tomato: 1
orange,tomato: 3
```

The items that appear in the frequent sets are therefore,

```
cashews,avocado,orange,tomato
```

there are several possible triplets that can be formed out of these four items. However, we are only interested in those triplets that satisfy the condition that all possible sub-pairs are frequent. This leads to only one triplet and its frequency is calculated as,

```
cashews,orange,tomato: 3
```

and the one frequent triplet is therefore

```
cashews,orange,tomato: 3
```

(b) Association rules attempt to predict the presence of one or more items j in a basket based on the presence of other items I there. The confidence is given by the ratio of number of baskets in which both I and j are present to those in which items I are present. The interest is the confidence minus the fraction of baskets in which j appears. An example of a useful negative interest: if people buying diapers buy beers significantly less than other customers, the rule diapers $->$ beer has negative interest. It is useful because we know not to place beer near diapers on the shelves.

(c) The possible association rules and their support (fraction of occurrence), note that `{I,J}` support means the fraction of sets that include both `I` and `j`; the confidence (fraction of baskets with I that also contain j); and interest (confidence in $I \to j$ minus fraction of baskets that contain j),

I -> j	: I support,	{I,J} support,	j fraction,	confidence,	interes
cashews,orange -> tomato:	4	3	2/3	3/4	1/
orange,tomato -> cashews:	3	3	5/6	3/3	1/
cashews,tomato -> orange:	3	3	2/3	3/3	1/

The third rule is the most interesting one…

(d) the data matrix for the given sets:

```
items={cashews,orange,avocado,apple,tomato,chocolate,pepper,lemon}
1  {cashews,orange,avocado,tomato}
2  {orange,tomato,chocolate,pepper,cashews}
3  {cashews,orange,avocado,lemon}
4  {apple,tomato}
5  {cashews,avocado}
```

```
6  {orange,tomato,cashews,chocolate,pepper}
data matrix:
                 1       2       3       4       5       6
----------------------------------------------------------
cashews          1       1       1       0       1       1
orange           1       1       1       0       0       1
avocado          1       0       1       0       1       0
apple            0       0       0       1       0       0
tomato           1       1       0       1       0       1
chocolate        0       1       0       0       0       1
pepper           0       1       0       0       0       1
lemon            0       0       1       0       0       0
```

The Jaccard similarity between baskets 1 and 2: 3/6=0.5

4. (10 pts) Find the smallest norm solution to $A\mathbf{x} = \mathbf{b}$, that also satisfies $|\mathbf{x}|^2 \geq 4.8$.

```
A=[1 2];
b=[2];
```

Hints: (1) Using tools learned in class, find an appropriate solution $\mathbf{x}_1$ to $A\mathbf{x} = \mathbf{b}$, and then add to it a vector $\mathbf{y}$ such that $\mathbf{x} = \mathbf{x}_1 + \mathbf{y}$ is still a solution, and that the inequality is satisfied. (2) This question is not asking you to repeat a calculation from HW or class, but to think independently based on material covered in class.

Solution:

First, find the smallest norm solution satisfying the equation,

```
 [U,S,V]=svd(A); my_fprintf_array(U); my_fprintf_array(S); my_fprintf_array(V);
U=[ -1 ];
S=[ 2.23607  0 ];
V=[ -0.447214  -0.894427
    -0.894427  0.447214 ];
Sdagger=[ 1/2.23607 ; 0 ];
pseudo_inverse=V*Sdagger*U';
x1=pseudo_inverse*b;
x1 =[0.4000;
     0.8000];
```

Now need to add a vector from the null space of A. This null space is the second V vector, so we need to add a multiple of that with some unknown multiplying factor a, $\mathbf{x} = \mathbf{x}_1 + a\mathbf{v}_2$. Because the $\mathbf{x}_1$ and the null space are orthogonal, we have $|\mathbf{x}|^2 = |\mathbf{x}_1|^2 + a^2|\mathbf{v}_2|$. To calculate this, first,

```
normX_squared=norm(x1)^2
 =0.8000
```

so, $|\mathbf{x}|^2 = 0.8 + a^2$. To satisfy the inequality, we need to choose $a = 2$, and the requested solution is,

```
x=x1+2*V(:,2)
 =[ -1.3889;
    1.6944];
```