8850 – QUIZ #1: SOLUTIONS
Machine Learning

Instructor: Prof. Pavel Skums; TA: Kiril Kuzmin

09/14/2018

Name: ________________________________

Honor Code Statement: "I will not commit any act of academic dishonesty while completing this assignment. I am fully aware that any of my own personal actions while attempting this assignment that are interpreted as academic dishonesty, will be treated as such. I understand that if I am held accountable for an act of academic dishonesty that I will receive a grade of "0" (zero) for this assignment and the incident will be reported to the Dean of Students Office."

This quiz contains 11 pages (including this cover page) and 6 questions. Total of points is 100.
NB: For each problem you have to provide an explanation on how you have obtained the answer. Even the correct answer will receive a partial credit if there is no explanation or unsatisfactory explanation.

Good luck and productive work!

Distribution of grades

Question	Points	Score
1	10	
2	15	
3	20	
4	20	
5	25	
6	10	
Total:	100	

1. (10 points) Calculate a determinant of the matrix

$$A = \begin{pmatrix} 2 & 0 & 4 & 5 \\ 0 & 2 & 3 & 5 \\ 0 & 0 & -3 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Solution. Since A is an (upper) triangular matrix, its determinant is the product of the main diagonal entries:

$$\det A = 2 \cdot 2 \cdot (-3) \cdot 1 = -12.$$

Answer. -12.

2. (15 points) Find eigenvectors and eigenvalues of the matrix

$$B = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Solution. First of all we obtain the eigenvalues of the matrix B. To do this, we have to find the values of λ which satisfy the characteristic equation of the matrix, namely those values of λ for which

$$\det(B - \lambda I) = 0,$$

where I is an identity matrix. Thus, we have an equation

$$\begin{vmatrix} 6-\lambda & 0 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & -1-\lambda \end{vmatrix} = 0,$$

i.e.,

$$(6-\lambda)(2-\lambda)(-1-\lambda) = 0.$$

Therefore, the eigenvalues of B are $\lambda = 6$, $\lambda = 2$ and $\lambda = -1$.

Once the eigenvalues have been found, we can find the eigenvectors. For each eigenvalue λ, we have to solve an equation

$$(B - \lambda I)\mathbf{u} = \mathbf{0},$$

where $\mathbf{u}$ is the eigenvector associated with eigenvalue λ.

Case 1: $\lambda = 6$. We have an equation

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -7 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the system of linear equations:

$$\begin{cases} -4u_2 = 0 \\ -7u_3 = 0 \end{cases}$$

The solutions are $u_1 = \alpha \in \mathbf{R}$, $u_2 = u_3 = 0$. Let us choose $\alpha = 1$, then the eigenvectors of B associated with the eigenvalue $\lambda = 6$ is $(1, 0, 0)$.

Case 2: $\lambda = 2$. We have an equation

$$\begin{pmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the system of linear equations:

$$\begin{cases} 4u_1 = 0 \\ -3u_3 = 0 \end{cases}$$

The solutions are $u_1 = 0$, $u_2 = \alpha \in \mathbf{R}$, $u_3 = 0$. Let us choose $\alpha = 1$, then the eigenvectors of B associated with the eigenvalue $\lambda = 2$ is $(0, 1, 0)$.

Case 3: $\lambda = -1$. We have an equation

$$\begin{pmatrix} 7 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the system of linear equations:

$$\begin{cases} 7u_1 = 0 \\ 3u_2 = 0 \end{cases}$$

The solutions are $u_1 = u_2 = 0$, $u_3 = \alpha \in \mathbf{R}$. Let us choose $\alpha = 1$, then the eigenvectors of B associated with the eigenvalue $\lambda = -1$ is $(0, 0, 1)$.

Answer. The eigenvalues of B are $\lambda = 6$, $\lambda = 2$ and $\lambda = -1$. The eigenvectors of B associated with these eigenvalues are $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ correspondingly.

3. (20 points) Find eigenvalues of the matrix C and its *orthonormal* basis consisting of eigenvectors, where

$$C = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{pmatrix}.$$

Solution. First of all we obtain the eigenvalues of the matrix C. To do this, we have to find the values of λ which satisfy the characteristic equation

$$\det(C - \lambda I) = 0.$$

Thus, we have an equation

$$\begin{vmatrix} 5-\lambda & 0 & 0 \\ 0 & 2-\lambda & 2 \\ 0 & 2 & 2-\lambda \end{vmatrix} = 0,$$

i.e.,

$$(5-\lambda)((2-\lambda)^2 - 4) = 0.$$

Therefore, the eigenvalues of B are $\lambda = 5$, $\lambda = 4$ and $\lambda = 0$.

For each eigenvalue λ, we solve an equation

$$(C - \lambda I)\mathbf{u} = \mathbf{0}.$$

Case 1: $\lambda = 5$. We have an equation

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -3 & 2 \\ 0 & 2 & -3 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the system of linear equations:

$$\begin{cases} -3u_2 + 2u_3 = 0 \\ 2u_2 - 3u_3 = 0 \end{cases}$$

The solutions are $u_1 = \alpha \in \mathbf{R}$, $u_2 = u_3 = 0$. We must choose $\alpha = 1$ due to the fact that all eigenvectors have to be orthonormal which means, in particular, that their length is 1. Thus, the eigenvector of C associated with the eigenvalue $\lambda = 5$ is $(1, 0, 0)$.

Case 2: $\lambda = 4$. We have an equation

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & 2 & -2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the system of linear equations:

$$\begin{cases} u_1 = 0 \\ -2u_2 + 2u_3 = 0 \\ 2u_2 - 2u_3 = 0 \end{cases}$$

In its turn the last system is the same as

$$\begin{cases} u_1 = 0 \\ 2u_2 - 2u_3 = 0 \end{cases}$$

The solutions are $u_1 = 0$, $u_2 = u_3$. Let u_3 be $\alpha \in \mathbf{R}$, Then the eigenvector of C associated with the eigenvalue $\lambda = 4$ is $(0, \alpha, \alpha)$. We choose $\alpha = \dfrac{1}{\sqrt{2}}$ to make sure that the length of the vector is 1. Therefore, the eigenvector in this case is $\left(0, \dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}\right)$.

Case 3: $\lambda = 0$. We have an equation

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the system of linear equations:

$$\begin{cases} 5u_1 = 0 \\ 2u_2 + 2u_3 = 0 \\ 2u_2 + 2u_3 = 0 \end{cases}$$

In its turn the last system is the same as

$$\begin{cases} 5u_1 = 0 \\ 2u_2 + 2u_3 = 0 \end{cases}$$

The solutions are $u_1 = 0$, $u_2 = -u_3$. Let u_3 be $\alpha \in \mathbf{R}$, Then the eigenvector of C associated with the eigenvalue $\lambda = 0$ is $(0, -\alpha, \alpha)$. We choose $\alpha = \dfrac{1}{\sqrt{2}}$ to make sure that the length of the vector is 1. Therefore, the eigenvector in this case is $\left(0, -\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}\right)$.

It is easy to check that a dot product of each pair of obtained eigenvector is equal to zero which means that they are orthogonal. In addition each of them is of length one. Thereby, the orthonormal basis has been found.

Answer. The eigenvalues of C are $\lambda = 5$, $\lambda = 4$ and $\lambda = 0$. The orthonormal basis consists of the following eigenvectors: $(1, 0, 0)$, $\left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $\left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

4. (20 points) Find singular value decomposition of the data represented by the matrix
$$D = \begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 3 \end{pmatrix}.$$

Solution. The singular value decomposition method is based on a theorem from linear algebra which says that a (rectangular) matrix $D = [d_{ij}] \in \mathbf{R}^{n\times m}$ can be broken down into the product of three matrices – an orthogonal matrix $U = [u_{ij}] \in \mathbf{R}^{n\times n}$, a diagonal matrix $\Omega = [\omega_{ij}] \in \mathbf{R}^{n\times m}$, and the transpose of an orthogonal matrix $V[v_{ij}] \in \mathbf{R}^{m\times m}$:
$$D = U\Omega V^T.$$

Here, the matrix Ω is a diagonal matrix containing the square roots of eigenvalues from U or V in descending order. The columns of U are orthonormal (or at least orthogonal) eigenvectors of the first covariance matrix DD^T ordered by the size of the corresponding eigenvalue, i.e., the eigenvector u_1 that corresponds to the highest eigenvalue is the first column of U; the eigenvector u_2 that corresponds to the second highest eigenvalue is the second column of U etc. The rows of V^T are orthonormal (or at least orthogonal) eigenvectors of the second covariance matrix D^TD, and they are also have to be written accordingly to the location of eigenvalues in Ω.

In our case, we have
$$DD^T = \begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 3 \end{pmatrix}\begin{pmatrix} 0 & 2 \\ 1 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix},$$
$$D^TD = \begin{pmatrix} 0 & 2 \\ 1 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 2 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 6 \\ 0 & 1 & 0 \\ 6 & 0 & 9 \end{pmatrix}.$$
Matrix DD^T has two eigenvalues $\lambda = 13$ and $\lambda = 1$ that correspond to the eigenvectors $(0, 1)$ and $(1, 0)$ respectively.

Thus,
$$U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Matrix D^TD has three eigenvalue $\lambda = 13$, $\lambda = 1$ and $\lambda = 0$. Pay your attention that having calculated eigenvalues for DD^T, you can at once name eigenvalues for $D^T D$. This trick is based on the following fact from linear algebra. Matrices DD^T and $D^T D$ have the same non-zero eigenvalues, and if one has more eigenvalues than the other, then these are all equal to zero. In our case, we have identical eigenvalues $\lambda = 13$ and $\lambda = 1$ and an additional eigenvalue that is equal to zero.

Going back to the point, we calculate the eigenvectors for $D^T D$. The vector $(2, 0, 3)$ corresponds to $\lambda = 13$, $(0, 1, 0)$ – to $\lambda = 1$ and $(3, 0, -2)$ – to $\lambda = 0$. If we normalize them, we have the following set of vectors:

$$\left(\frac{2}{\sqrt{13}}, 0, \frac{3}{\sqrt{13}}\right), (0, 1, 0), \left(\frac{3}{\sqrt{13}}, 0, -\frac{2}{\sqrt{13}}\right).$$

Hence,

$$V^T = \begin{pmatrix} \frac{2}{\sqrt{13}} & 0 & \frac{3}{\sqrt{13}} \\ 0 & 1 & 0 \\ \frac{3}{\sqrt{13}} & 0 & -\frac{2}{\sqrt{13}} \end{pmatrix}.$$

Finally, we form the matrix Ω. For this purpose, we take the square roots of the non-zero eigenvalues and populate the diagonal with them, putting the largest in ω_{11} and the next largest in ω_{22}. So we have

$$\Omega = \begin{pmatrix} \sqrt{13} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

The diagonal entries in S are the singular values of A, the columns in U are called left singular vectors, and the columns in V are called right singular vectors.

Answer.

$$U\Omega V^T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{13} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{13}} & 0 & \frac{3}{\sqrt{13}} \\ 0 & 1 & 0 \\ \frac{3}{\sqrt{13}} & 0 & -\frac{2}{\sqrt{13}} \end{pmatrix}.$$

5. (25 points) Assume you have conducted an experiment 6 times and each i-th time you observed two characteristics x_i and y_i; see the table below:

i	x_i	y_i
1	-1	1
2	1	2
3	-2	3
4	2	2
5	-3	-2
6	3	6

Perform a principal components analysis on the set of data. Choose the most important principal component and find a reduced data set.

Solution. First of all we put the data into a matrix:

$$A = \begin{pmatrix} -1 & 1 & -2 & 2 & -3 & 3 \\ 1 & 2 & 3 & 2 & -2 & 6 \end{pmatrix}.$$

Then subtract the mean. The mean of x-s is $\bar{x} = 0$, the mean of y-s is $\bar{y} = 2$. The matrix with adjasted data look like

$$\bar{A} = \begin{pmatrix} -1 & 1 & -2 & 2 & -3 & 3 \\ -1 & 0 & 1 & 0 & -4 & 4 \end{pmatrix}.$$

Now we calculate the covariance matrix $\bar{A}\bar{A}^T$:

$$\bar{A}\bar{A}^T = \begin{pmatrix} 28 & 23 \\ 23 & 34 \end{pmatrix}.$$

After we calculate the eigenvalues of the covariance matrix $\bar{A}\bar{A}^T$:

$$\lambda \approx 54, \quad \lambda \approx 8.$$

We take the biggest eigenvalue and find an eigenvector corresponding to it:

$$v_1 = (23, 26).$$

Finally, we deriving the reduced data set R:

$$R = v_1\bar{A} = (23, 26)\begin{pmatrix} -1 & 1 & -2 & 2 & -3 & 3 \\ -1 & 0 & 1 & 0 & -4 & 4 \end{pmatrix} = (-49, 23, 20, 46, -173, 173).$$

Answer. $(-49, 23, 20, 46, -173, 173)$.

6. (10 points) Construct an example (in $\mathbf{R}^2$) on 8 points where the optimal regression line is $y = -x$, even though *none of the input points* lie directly on this line.

Solution. One of possible solutions may be the following. We can take any two lines that are symmetrical to each other with $y = -x$ as a axis of their symmetry. For instance, let us take lines $y = -x+1$ and $y = -x-1$. Now, we may choose any four points from each line. Let choose $(0, 1)$, $(1, 0)$, $(2, -1)$, $(3, -2)$ from $y = -x+1$; and $(0, -1)$, $(1, -2)$, $(2, -3)$, $(3, -4)$ from $y = -x-1$. These points are the solution we wanted to find.

Answer. $(0, 1)$, $(1, 0)$, $(2, -1)$, $(3, -2)$, $(0, -1)$, $(1, -2)$, $(2, -3)$, $(3, -4)$.