

Seoul National University

M1522.001400 Introduction to Data Mining

Spring 2016, Kang

Homework 10: Dimensionality Reduction (Chapter 11)

Due: June 15, 09:30 AM

## Reminders

- The points of this homework add up to 100.
- Like all homeworks, this has to be done individually.
- Lead T.A.: Minsoo Jung ([qtyp456987@gmail.com](mailto:qtyp456987@gmail.com))
- Please type your answers *in English*. Illegible handwriting may get no points, at the discretion of the graders.
- If you have a question about assignments, please upload your question in eTL.
- If you want to use slipdays or consider late submission with penalties, please note that you are allowed one week to submit your assignment after the due date.
- You can submit the homework by hand on the final exam date, June 13, or via e-mail until the due date, June 15.

Remember that:

- Whenever you are making an assumption, please state it clearly.

**Question 1**

Find the eigenvalues and its corresponding eigenvectors for the following matrix  
[20 points]:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 5 \end{pmatrix}$$

## Question 2

In **Figure 1** is a matrix  $M$ . It has rank 2, as you can see by observing that the first column plus the third column minus twice the second column equals 0.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix}$$

**Figure 1:** Matrix  $M$  for **Question 2**.

- (a) Compute the matrices  $M^T M$  and  $MM^T$ . [5 points]
- (b) Find the eigenvalues for the matrices of **Question 2** - (a). [5 points]
- (c) Find the eigenvectors for the matrices of **Question 2** – (a). [5 points]

- (d) Find the SVD for the original matrix  $M$  from **Question 2** - (b) and (c). Note that there are only two nonzero eigenvalues, so your matrix  $\Sigma$  should have only two singular values, while  $U$  and  $V$  have only two columns. [5 points]
- (e) Set the smaller singular value to 0 and compute the one-dimensional approximation to the matrix  $M$  from **Figure 1**. [5 points]
- (f) How much of the energy of the original singular values is retained by the one-dimensional approximation? [5 points]

### Question 3

Use the ratings of movies from **Figure 2** and the SVD from **Figure 3**. Suppose Leslie assigns rating 3 to Alien and rating 4 to Titanic, giving us a representation of Leslie in “movie space” of  $[0, 3, 0, 0, 4]$ . Find the representation of Leslie in concept space. [30 points]

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

**Figure 2:** Ratings of movies by users for **Question 3**.

$$\begin{array}{c}
 \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} .14 & 0 \\ .42 & 0 \\ .56 & 0 \\ .70 & 0 \\ 0 & .60 \\ 0 & .75 \\ 0 & .30 \end{bmatrix} \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix} \begin{bmatrix} .58 & .58 & .58 & 0 & 0 \\ 0 & 0 & 0 & .71 & .71 \end{bmatrix} \\
 M \qquad \qquad \qquad U \qquad \qquad \qquad \Sigma \qquad \qquad \qquad V^T
 \end{array}$$

**Figure 3:** SVD for the matrix  $M$  of **Figure 2** for **Question 3**.

#### Question 4

Demonstrate that the rank of the matrix in **Figure 4** is 3. [20 points]

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	2	0	4	4
Jenny	0	0	0	5	5
Jane	0	1	0	2	2

**Figure 4:** The matrix with ratings of movies for **Question 4**.