

CS246: Mining Massive Datasets

Homework 2

Answer to Question 1(a)

From the problem,

$$r'_i = \sum_j m_{ij} r_j, \quad (1)$$

and

$$w(\mathbf{r}) = \sum_i r_i. \quad (2)$$

Thus,

$$\begin{aligned} w(\mathbf{r}') &= \sum_i \sum_j m_{ij} r_j, \\ &= \sum_j \sum_i m_{ij} r_j, \\ &= \sum_j r_j \left(\sum_i m_{ij} \right), \\ &= \sum_j r_j, \\ &= w(\mathbf{r}). \end{aligned} \quad (3)$$

Answer to Question 1(b)

From the problem,

$$r'_i = \beta \sum_j m_{ij} r_j + (1 - \beta)/n. \quad (4)$$

As before,

$$\begin{aligned} w(\mathbf{r}') &= \sum_i r'_i = \sum_i (\beta \sum_j m_{ij} r_j + (1 - \beta)/n) \\ &= (1 - \beta) + \beta \sum_j r_j \sum_i m_{ij}, \\ &= (1 - \beta) + \beta \sum_j r_j. \end{aligned} \quad (5)$$

From the last equivalence, it follows that $w(\mathbf{r}') = w(\mathbf{r})$ if $\sum_i r_i = 1$.

Answer to Question 1(c)

Let \mathcal{L} be the set of live nodes, and \mathcal{D} the set of dead nodes. From the problem definition we can write our update rule as:

$$r'_i = \sum_j m_{ij} r_j \beta + T_j, \quad (6)$$

where,

$$T_j = \begin{cases} (1 - \beta)r_j/n & \text{if } j \in \mathcal{L}, \\ r_j/n & \text{if } j \in \mathcal{D}. \end{cases}$$

Note that for the case of dead nodes, we can separate the T_j component into two parts,

$$r_j/n = \beta r_j/n + (1 - \beta)r_j/n. \quad (7)$$

With this information, we can create a surrogate matrix \hat{M} such that

$$\hat{m}_{ij} = \begin{cases} m_{ij} & \text{if } j \in \mathcal{L}, \\ 1/n & \text{if } j \in \mathcal{D}. \end{cases} \quad (8)$$

Thus, we can write our update rule in terms of this surrogate matrix:

$$\begin{aligned} r'_i &= \sum_j \hat{m}_{ij} r_j \beta + (1 - \beta)r_j/n, \\ &= \beta \sum_j \hat{m}_{ij} r_j + (1 - \beta)/n. \end{aligned} \quad (9)$$

Which is identical in form to the previous question, from which it follows that $w(\mathbf{r}') = 1$.

Answer to Question 2(a)

Assuming M is real (an assumption validated by an instructor comment in Piazza), then both $M^T M$ and MM^T are real.

For a matrix A to be symmetric, $A^T = A$ must hold. Clearly, since $(MM^T)^T = (M^T)^T M = MM^T$ and $(M^T M)^T = M^T (M^T)^T$, then both are symmetric.

Finally, if M is a $p \times q$ matrix, then $M^T M$ is $q \times q$ and MM^T is $p \times p$, making them both square.

Answer to Question 2(b)

Parting from the MMDS book, section 11, let's use SVD to answer this question. Let $M = U\Sigma V$ be a singular value decomposition of M , where Σ is a diagonal matrix with the singular values in its diagonal. Then,

$$M^T M = V\Sigma U^T U \Sigma V^T = V\Sigma^2 V^T, \quad (10)$$

since U is an orthonormal matrix and $U^T U = I$. Rearranging, and using that fact that V is also orthonormal, we get that

$$M^T M V = V\Sigma^2 \quad (11)$$

which implies that V is the matrix of eigenvectors of $M^T M$ and Σ^2 a diagonal matrix with the eigenvalues in the diagonal. Similarly, we can arrive to $MM^T U = U\Sigma^2$. This shows that the eigenvalues of $M^T M$ and MM^T are the same, but their eigenvectors are different.

Answer to Question 2(c)

Since $M^T M$ is square, real and symmetric, we can write

$$M^T M = Q \Lambda Q^T \tag{12}$$

Answer to Question 2(d)

From question b), we see that Σ^2 contains the eigenvalues and V the eigenvectors, thus we can simply write

$$M^T M = V \Sigma^2 V^T \tag{13}$$

Answer to Question 2(e)

From the SVD of M results (code submitted to SNAP):

$$U = \begin{bmatrix} -0.27854301 & 0.5 & -0.75033067 & -0.33078343 \\ -0.27854301 & -0.5 & 0.12733222 & -0.81006191 \\ -0.64993368 & 0.5 & 0.57233111 & 0.00482762 \\ -0.64993368 & -0.5 & -0.30533177 & 0.4841061 \end{bmatrix} \quad (14)$$

$$\Sigma(\text{only non-zero elements}) = \begin{bmatrix} 7.61577311 & 1.41421356 \end{bmatrix} \quad (15)$$

$$V = \begin{bmatrix} -0.70710678 & -0.70710678 \\ -0.70710678 & 0.70710678 \end{bmatrix} \quad (16)$$

From the Eigendecomposition:

$$Evals = \begin{bmatrix} 58. & 2. \end{bmatrix} \quad (17)$$

$$Vecs = \begin{bmatrix} 0.70710678 & -0.70710678 \\ 0.70710678 & 0.70710678 \end{bmatrix} \quad (18)$$

V and $Vecs$ are the same in magnitude, but not in sign. Technically, one can arrive to a different SVD decomposition of M using Eigendecompositions of $M^T M$ and MM^T and viceversa.

The singular values of M are the square root of the eigenvalues of $M^T M$ and MM^T , that is $\sqrt{\Lambda_{ii}} = \Sigma_{ii}$

Answer to Question 3(a)

Answer to Question 3(b)

Answer to Question 4(a)

Answer to Question 4(b)

Cover Sheet

Assignment Submission Fill in and include this cover sheet with each of your assignments. Assignments are due at 11:59pm. All students (SCPD and non-SCPD) must submit their homeworks via GradeScope (<http://www.gradescope.com>). Students can typeset or scan their homeworks. Make sure that you answer each question on a separate page. Students also need to upload their code at <http://snap.stanford.edu/submit>. Put all the code for a single question into a single file and upload it. Please do not put any code in your GradeScope submissions.

Late Day Policy Each student will have a total of *two* free late periods. *One late period expires at the start of each class.* (Homeworks are usually due on Thursdays, which means the first late periods expires on the following Tuesday.) Once these late periods are exhausted, any assignments turned in late will be penalized 50% per late period. However, no assignment will be accepted more than *one* late period after its due date.

Honor Code We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down their solutions independently i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their discussion group. Using code or solutions obtained from the web (github/google/previous year solutions etc.) is considered an honor code violation. We check all the submissions for plagiarism. We take the honor code very seriously and expect students to do the same.

Your name: Ramon Iglesias
Email: rdit@stanford.edu **SUID:** rdit

Discussion Group (People with whom you discussed ideas used in your answers):

On-line or hardcopy documents used as part of your answers:

I acknowledge and accept the Honor Code.

(Signed) RDIT