Question 3: Assignment 4: CS 663, Fall 2024

Amitesh Shekhar
IIT Bombay
22b0014@iitb.ac.in

Anupam Rawat
IIT Bombay
22b3982@iitb.ac.in

Toshan Achintya Golla
IIT Bombay
22b2234@iitb.ac.in

October 22, 2024

1. The aim of this exercise is to help you understand the mathematics of SVD more deeply. Do as directed: [30 points – see split-up below]

   (a) Argue that the non-zero singular values of a matrix $\boldsymbol{A}$ are the square-roots of the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$ or $\boldsymbol{A}^T\boldsymbol{A}$. (Make arguments for both) [3 points]

   (b) Show that the squared Frobenius norm of a matrix is equal to the sum of squares of its singular values. [3 points]

   (c) A students tries to obtain the SVD of a $m \times n$ matrix $\boldsymbol{A}$ using eigendecomposition. For this, the student computes $\boldsymbol{A}^T\boldsymbol{A}$ and assigns the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ (computed using the `eig` routine in MATLAB) to be the matrix $\boldsymbol{V}$ consisting of the right singular vectors of $\boldsymbol{A}$. Then the student also computes $\boldsymbol{A}\boldsymbol{A}^T$ and assigns the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ (computed using the `eig` routine in MATLAB) to be the matrix $\boldsymbol{U}$ consisting of the left singular vectors of $\boldsymbol{A}$. Finally, the student assigns the non-negative square-roots of the eigenvalues (computed using the `eig` routine in MATLAB) of either $\boldsymbol{A}^T\boldsymbol{A}$ or $\boldsymbol{A}\boldsymbol{A}^T$ to be the diagonal matrix $\boldsymbol{S}$ consisting of the singular values of $\boldsymbol{A}$. He/she tries to check his/her code and is surprised to find that $\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ is not equal to $\boldsymbol{A}$. Why could this be happening? What processing (s)he do to the computed <u>eigenvectors</u> of $\boldsymbol{A}^T\boldsymbol{A}$ and/or $\boldsymbol{A}\boldsymbol{A}^T$ in order rectify this error? (Note: please try this on your own in MATLAB.) [8 points]

   (d) Consider a matrix $\boldsymbol{A}$ of size $m \times n, m \leq n$. Define $\boldsymbol{P} = \boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{Q} = \boldsymbol{A}\boldsymbol{A}^T$. (Note: all matrices, vectors and scalars involved in this question are real-valued). [4+4+4+4=16 points]

      i. Prove that for any vector $\boldsymbol{y}$ with appropriate number of elements, we have $\boldsymbol{y}^t\boldsymbol{P}\boldsymbol{y} \geq 0$. Similarly show that $\boldsymbol{z}^t\boldsymbol{Q}\boldsymbol{z} \geq 0$ for a vector $\boldsymbol{z}$ with appropriate number of elements. Why are the eigenvalues of $\boldsymbol{P}$ and $\boldsymbol{Q}$ non-negative?

      ii. If $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{P}$ with eigenvalue λ, show that $\boldsymbol{A}\boldsymbol{u}$ is an eigenvector of $\boldsymbol{Q}$ with eigenvalue λ. If $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{Q}$ with eigenvalue μ, show that $\boldsymbol{A}^T\boldsymbol{v}$ is an eigenvector of $\boldsymbol{P}$ with eigenvalue μ. What will be the number of elements in $\boldsymbol{u}$ and $\boldsymbol{v}$?

      iii. If $\boldsymbol{v}_i$ is an eigenvector of $\boldsymbol{Q}$ and we define $\boldsymbol{u}_i \triangleq \dfrac{\boldsymbol{A}^T\boldsymbol{v}_i}{\|\boldsymbol{A}^T\boldsymbol{v}_i\|_2}$. Then prove that there will exist some real, non-negative γ_i such that $\boldsymbol{A}\boldsymbol{u}_i = \gamma_i\boldsymbol{v}_i$.

      iv. It can be shown that $\boldsymbol{u}_i^T\boldsymbol{u}_j = 0$ for $i \neq j$ and likewise $\boldsymbol{v}_i^T\boldsymbol{v}_j = 0$ for $i \neq j$ for correspondingly distinct eigenvalues. (You did this in HW4 where you showed that the eigenvectors of symmetric matrices are orthonormal.) Now, define $\boldsymbol{U} = [\boldsymbol{v}_1|\boldsymbol{v}_2|\boldsymbol{v}_3|...|\boldsymbol{v}_m]$ and $\boldsymbol{V} = [\boldsymbol{u}_1|\boldsymbol{u}_2|\boldsymbol{u}_3|...|\boldsymbol{u}_m]$. Now show that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{V}^T$ where $\boldsymbol{\Gamma}$ is a diagonal matrix containing the non-negative values $\gamma_1, \gamma_2, ..., \gamma_m$. With this, you have just established the existence of the singular value decomposition of any matrix $\boldsymbol{A}$.

Soln:

(a) We know that SVD of matrix $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{USV}^T$. $\mathbf{S}$ is a diagonal matrix that contains eigenvalues of $\mathbf{A}$. Similarly, the $\mathbf{A}^T$ is given by $\mathbf{A}^T = \mathbf{V}\ \mathbf{S}^T\ \mathbf{A}$ Hence, we have:

$$\mathbf{A} \cdot \mathbf{A}^T = \mathbf{USV}^T \cdot \mathbf{VS}^T\mathbf{U}^T$$

$$\mathbf{A}^T \cdot \mathbf{A} = \mathbf{VS}^T\mathbf{U}^T \cdot \mathbf{USV}^T$$

Since, $\mathbf{V}$ and $\mathbf{U}$ both are orthonormal and $\mathbf{S}$ is a diagonal matrix, $\mathbf{S}^T = \mathbf{S}$. Thus the above equations resolve to:

$$\mathbf{A} \cdot \mathbf{A}^T = \mathbf{US}^2\mathbf{U}^T$$

$$\mathbf{A}^T \cdot \mathbf{A} = \mathbf{VS}^2\mathbf{V}^T$$

And the $\mathbf{S}$ is the square root of the eigenvalues of $\mathbf{A} \cdot \mathbf{A}^T$ or $\mathbf{A}^T \cdot \mathbf{A}$, which proves the statement.

(b) We know that the Frobenius norm of a matrix $\mathbf{A}$ is given by:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} \tag{1}$$

The singular value decomposition of matrix $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{USV}^T$. The Frobenius norm of matrix $\mathbf{A}$ can be written as:

$$\|\mathbf{A}\|_F = \|\mathbf{USV}^T\|_F = \|\mathbf{S}\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2} \tag{2}$$

Thus, we have shown that the squared Frobenius norm of a matrix is equal to the sum of squares of its singular values.

(c) We must note that U is obtained from the eigenvectors of $\boldsymbol{AA}^T$ and V is obtained from the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$. If we independently use eig for obtaining U and V, then it can lead to inconsistencies in the signs, due to which A will not be equal to the matrix product $\boldsymbol{USV}^T$. This inconsistency occurs because of the fact that if a vector $\boldsymbol{k}$ is a eigenvector of matrix A with eigenvalue λ, then the vector $-\boldsymbol{k}$ is also a eigenvector of matrix A with same eigenvalue λ.
A simple way to solve this problem is to mark out those $\boldsymbol{u}$,$\boldsymbol{v}$ pairs which are having a sign inconsistency and then multiply either the $\boldsymbol{u}$ or $\boldsymbol{v}$ vector by -1 in all such cases.

(d)

$$P = A^TA = VS^2V^T$$

$$Q = AA^T = US^2U^T$$

i.

$$\mathbf{y}^T\mathbf{P}\mathbf{y} = \mathbf{y}^T(\mathbf{A}^T\mathbf{A})\mathbf{y} = (\mathbf{y}^T\mathbf{A}^T)(\mathbf{A}\mathbf{y}) = (\mathbf{A}\mathbf{y})^T(\mathbf{A}\mathbf{y}) = ||\mathbf{A}\mathbf{y}||^2 \geq 0$$

$$\mathbf{z}^T\mathbf{Q}\mathbf{z} = \mathbf{z}^T(\mathbf{A}\mathbf{A}^T)\mathbf{z} = (\mathbf{z}^T\mathbf{A})(\mathbf{A}^T\mathbf{z}) = (\mathbf{A}^T\mathbf{z})^T(\mathbf{A}^T\mathbf{z}) = ||\mathbf{A}^T\mathbf{z}||^2 \geq 0$$

P and Q are symmetric matrices and also, as above, $\mathbf{y}^T\mathbf{P}\mathbf{y} \geq 0$ and $\mathbf{z}^T\mathbf{Q}\mathbf{z} \geq 0$ for any arbitrary vectors x and y. Thus, we can say that P and Q are positive semi-definite matrices. And since the eigenvalues of a positive semi-definite matrix are non-negative, we can say that eigenvalues of P and Q are non-negative.

ii.

$$\mathbf{Pu} = \mathbf{A}^T\mathbf{Au} = \lambda\mathbf{u}$$

$$\mathbf{Q} \cdot \mathbf{Au} = \mathbf{AA}^T \cdot \mathbf{Au} = \mathbf{A}(\mathbf{A}^T\mathbf{Au}) = \mathbf{A}(\lambda\mathbf{u}) = \lambda \cdot \mathbf{Au}$$

Thus, we can say that $\mathbf{Au}$ is an eigenvector of $\mathbf{Q}$, with corresponding eigenvalue as λ.

$$\boldsymbol{Qv} = \boldsymbol{AA}^T \cdot \boldsymbol{v} = \mu\boldsymbol{v}$$

$$\boldsymbol{P} \cdot \boldsymbol{A^Tv} = (\boldsymbol{A}^T\boldsymbol{A}) \cdot \boldsymbol{A^Tv} = \boldsymbol{A}^T \cdot (\boldsymbol{AA^Tv}) = \boldsymbol{A}^T \cdot \mu\boldsymbol{v} = \mu\boldsymbol{A^Tv}$$

Thus, we can say that $A^T v$ is an eigenvector of $\mathbf{P}$, with corresponding eigenvalue as μ.
We know that $\boldsymbol{A}$ is a m×n matrix, and $\boldsymbol{A}^T$ is a $n \times m$ matrix. Since, we are doing a product of $\boldsymbol{A}$ and $\boldsymbol{u}$, and $\boldsymbol{u}$ is a column vector, it must have **n** elements. Similarly, since we are doing a multiplication of $\boldsymbol{A^T}$ and $\boldsymbol{v}$, $\boldsymbol{v}$ must have **m** elements.

iii. Let λ_i be the eigenvalue of the eigenvector $\boldsymbol{v_i}$ of the matrix $\boldsymbol{Q}$.

$$\boldsymbol{Q v_i} = \boldsymbol{A A}^T \boldsymbol{v_i} = \lambda_i \boldsymbol{v_i}$$

$$\boldsymbol{A u_i} = \boldsymbol{A} \left(\frac{\boldsymbol{A}^T \boldsymbol{v_i}}{\|\boldsymbol{A}^T \boldsymbol{v}_i\|_2} \right) = \frac{\boldsymbol{A A}^T \boldsymbol{v_i}}{\|\boldsymbol{A}^T \boldsymbol{v}_i\|_2} = \frac{\lambda_i \boldsymbol{v}_i}{\|\boldsymbol{A}^T \boldsymbol{v}_i\|_2}$$

We define a scalar, $\gamma_i \triangleq \dfrac{\lambda_i}{\|\boldsymbol{A}^T \boldsymbol{v}_i\|_2}$.
Since $\boldsymbol{Q}$ is a positive semi-definite matrix, we can say for sure that λ_i is real and non-negative eigenvalue, and since $\|\boldsymbol{A}^T \boldsymbol{v}_i\|_2$ is a norm of a vector (which is also non-negative), it follows that λ_i is real, non-negative scalar.
Thus, we can say that $\boldsymbol{A u}_i = \lambda_i \boldsymbol{v}_i$.

iv. As proved above, we know that $\boldsymbol{A u}_i = \lambda_i \boldsymbol{v}_i$. Without loss of generality, taking m $\leq$ n, we can say that $\boldsymbol{A u}_i = \lambda_i \boldsymbol{v}_i$ for all $1 \leq i \leq m$, and $\boldsymbol{A u}_i = 0$ for all $m+1 \leq i \leq n$. This means that we can write the relation $\boldsymbol{AV = U\Gamma}$ such that $\boldsymbol{V}$ is an orthonormal matrix of order n×n, $\boldsymbol{\Gamma}$ is a diagonal matrix of size m×n having at most m non-zero values along the diagonal and U is an orthonormal matrix of size m×m. Now, if we multiply $\boldsymbol{V}^T$ on both sides, then since $\boldsymbol{V}$ is orthonormal, we can say that $\boldsymbol{V V}^T = I$. Thereby, we can conclude that $\boldsymbol{A} = \boldsymbol{U \Gamma V}^T$.