Assignment 4: CS 663

Due: 26th October before 11:55 pm

Remember the honor code while submitting this (and every other) assignment. You may discuss broad ideas with other students or ask me for any difficulties, but the code you implement and the answers you write must be your own. We will adopt a zero-tolerance policy against any violation.

Submission instructions: Follow the instructions for the submission format and the naming convention of your files from the submission guidelines file in the homework folder. However, please do *not* submit the face image databases in your zip file that you will upload on moodle. Please see assignment4_SVD_FaceRecognition.rar. Upload the file on moodle *before* 11:55 pm on 26th October. We will not penalize submission of the files till 10 am on 27th October. **No late assignments will be accepted after this time.** Please preserve a copy of all your work until the end of the semester.

1. The aim of this exercise is to help you understand the mathematics behind PCA more deeply. Do as directed: [5+5+5+5=20 points]

   (a) Prove that the covariance matrix in PCA is symmetric and positive semi-definite.

   (b) Prove that the eigenvectors of a symmetric matrix are orthonormal.

   (c) Consider a dataset of some N vectors in d dimensions given by $\{\boldsymbol{x_i}\}_{i=1}^N$ with mean vector $\bar{\boldsymbol{x}}$. Note that each $\boldsymbol{x_i} \in \mathbb{R}^d$ and also $\bar{\boldsymbol{x}} \in \mathbb{R}^d$. Suppose that only k eigenvalues of the corresponding covariance matrix are large and the remaining are very small in value. Let $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}$ be an approximation to $\boldsymbol{x_i}$ of the form $\tilde{\boldsymbol{x}}_{\boldsymbol{i}} = \bar{\boldsymbol{x}} + \sum_{l=1}^k \boldsymbol{V_l}\alpha_{il}$ where $\boldsymbol{V_l}$ stands for the lth eigenvector (it has d elements) and α_{il} (it is a scalar) stands for the lth eigencoefficient of $\boldsymbol{x_i}$. Argue why the error $\frac{1}{N}\sum_{i=1}^N \|\tilde{\boldsymbol{x}}_{\boldsymbol{i}} - \boldsymbol{x_i}\|_2^2$ will be small. What will be the value of this error in terms of the eigenvalues of the covariance matrix?

   (d) Consider two uncorrelated zero-mean random variables (X_1, X_2). Let X_1 belong to a Gaussian distribution with variance 100 and X_2 belong to a Gaussian distribution with variance 1. What are the principal components of (X_1, X_2)? If the variance of X_1 and X_2 were equal, what are the principal components?

   Solution:

   (a) A matrix $\boldsymbol{C}$ is positive semi-definite if $\boldsymbol{w}^t\boldsymbol{C}\boldsymbol{w} \geq 0$ for any vector $\boldsymbol{w}$. Since $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^T/(N-1)$ (see slides for the notation regarding N and $\boldsymbol{X}$) which yields $\boldsymbol{w}^t\boldsymbol{C}\boldsymbol{w} = \boldsymbol{w}^t\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{w}/(N-1) = \|\boldsymbol{X}^T\boldsymbol{w}\|^2/(N-1)$. This is a squared vector magnitude and it must be non-negative, which prove that C is positive semi-definite. The symmetry of $\boldsymbol{C}$ follows from the fact that $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^T/(N-1)$. **Marking scheme:** 1 point for symmetry, 1 point for definition of positive semi-definite and 2 points for the proof. An alternative proof will involve proving that eigenvalues are non-negative. For this, let $\boldsymbol{v}$ be an eigenvector. Clearly $\boldsymbol{v}^t\boldsymbol{C}\boldsymbol{v}$ must be non-negative by the definition of $\boldsymbol{C}$. But $\boldsymbol{C}\boldsymbol{v} = \lambda\boldsymbol{v}$, due to which $\boldsymbol{v}^t\boldsymbol{C}\boldsymbol{v} = \lambda$ as $\boldsymbol{v}$ is a unit vector. As $\boldsymbol{v}^t\boldsymbol{C}\boldsymbol{v}$ must be non-negative, then λ must be non-negative.

   (b) Let $\boldsymbol{v_1}, \boldsymbol{v_2}$ be the eigenvectors of symmetric matrix $\boldsymbol{C}$ corresponding to distinct eigenvalues λ_1, λ_2 respectively. Then $\boldsymbol{C}\boldsymbol{v_1} = \lambda_1\boldsymbol{v_1}$ and $\boldsymbol{C}\boldsymbol{v_2} = \lambda_2\boldsymbol{v_2}$. We have $\boldsymbol{v_2^t}\boldsymbol{C}\boldsymbol{v1} = \lambda_1\boldsymbol{v_2^t}\boldsymbol{v_1}$ and $\boldsymbol{v_1^t}\boldsymbol{C}\boldsymbol{v2} = \lambda_2\boldsymbol{v_1^t}\boldsymbol{v_2}$. Now $\boldsymbol{v_1^t}\boldsymbol{C}\boldsymbol{v2} = \sum_j v_{1j}\sum_i C_{ji}v_{2i} = \sum_{i,j} C_{ij}v_{1j}v_{2i} = \boldsymbol{v_2^t}\boldsymbol{C}\boldsymbol{v1}$ due to the symmetric nature of $\boldsymbol{C}$. Hence, we will have $\lambda_1\boldsymbol{v_2^t}\boldsymbol{v_1} = \lambda_2\boldsymbol{v_2^t}\boldsymbol{v_1}$. Since $\lambda_1 \neq \lambda_2$, we must have $\boldsymbol{v_2^t}\boldsymbol{v_1} = 0$ which means $\boldsymbol{v_1}, \boldsymbol{v_2}$ must be orthogonal to each other. As each eigenvector by convention has unit magnitude, the matrix of eigenvectors of $\boldsymbol{C}$ must be orthonormal. **Marking scheme:** Any reasonable argument should fetch 2.5 out of 5 even if the proof is not complete.

(c) We have $\boldsymbol{x_i} = \bar{\boldsymbol{x}} + \sum_{l=1}^{d} \boldsymbol{V_l}\alpha_{il} = \bar{\boldsymbol{x}} + \boldsymbol{V}\boldsymbol{\alpha_i}$. Likewise $\tilde{\boldsymbol{x}}_i = \bar{\boldsymbol{x}} + \sum_{l=1}^{k} \boldsymbol{V_l}\alpha_{il} = \bar{\boldsymbol{x}} + \boldsymbol{V}\tilde{\boldsymbol{\alpha}}_i$ where $\tilde{\alpha}_{il} = \alpha_{il}$ for $1 \leq l \leq k$ and $\tilde{\alpha}_{il} = 0$ for $k+1 \leq l \leq d$. Then $\frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{x_i} - \tilde{\boldsymbol{x}}_i\|^2 = \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{\alpha_i} - \tilde{\boldsymbol{\alpha}}_i\|^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{l=k+1}^{d} \alpha_{il}^2 \approx \sum_{l=k+1}^{d} E[\alpha_l^2] = \sum_{l=k+1}^{d} \lambda_l$ because eigenvalues of a covariance matrix equal the variance of the eigencoefficients and the eigencoefficients have 0 mean. The RHS is small, because all eigenvalues $\lambda_{k+1}, ..., \lambda_d$ are small. **Marking scheme:** Any reasonable argument should fetch 2.5 out of 5 even if the proof is not complete.

(d) The spread is much greater for X_1, hence the eigenvector corresponding to the larger eigenvalue is very close to $(1;0)$ and the other eigenvector is very close to $(0;1)$ (both are column vectors with 2 elements). The corresponding eigenvalues are 100 and 1 respectively.

2. Consider a set of N vectors $\mathcal{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ each in $\mathbb{R}^d$, with average vector $\bar{\boldsymbol{x}}$. We have seen in class that the direction $\boldsymbol{e}$ such that $\sum_{i=1}^{N} \|\boldsymbol{x_i} - \bar{\boldsymbol{x}} - (\boldsymbol{e} \cdot (\boldsymbol{x_i} - \bar{\boldsymbol{x}}))\boldsymbol{e}\|^2$ is minimized, is obtained by maximizing $\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{e}$, where $\boldsymbol{C}$ is the covariance matrix of the vectors in $\mathcal{X}$. This vector $\boldsymbol{e}$ is the eigenvector of matrix $\boldsymbol{C}$ with the highest eigenvalue. Prove that the direction $\boldsymbol{f}$ perpendicular to $\boldsymbol{e}$ for which $\boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f}$ is maximized, is the eigenvector of $\boldsymbol{C}$ with the second highest eigenvalue. For simplicity, assume that all non-zero eigenvalues of $\boldsymbol{C}$ are distinct and that $\text{rank}(\boldsymbol{C}) > 2$. Extend the derivation to handle the case of a unit vector $\boldsymbol{g}$ which is perpendicular to both $\boldsymbol{e}$ and $\boldsymbol{f}$ which maximizes $\boldsymbol{g}^t\boldsymbol{C}\boldsymbol{g}$. [15 points]
Solution: Refer to lecture notes on Lagrange multipliers. We have $J_1(\boldsymbol{f}) = \boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f} - \lambda_1(\boldsymbol{f}^t\boldsymbol{f} - 1) - \lambda_2(\boldsymbol{f}^t\boldsymbol{e})$. Taking the derivatives w.r.t. $\boldsymbol{f}$ and setting to 0, we have $2\boldsymbol{C}\boldsymbol{f} - 2\lambda_1\boldsymbol{f} - \lambda_2\boldsymbol{e} = \boldsymbol{0}$ (note: this is a zero vector on the RHS). Taking dot product on both sides by $\boldsymbol{e}$, we get $\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{f} - \lambda_1\boldsymbol{e}^t\boldsymbol{f} - \lambda_2 = 0$. The first term is equal to $\lambda_1\boldsymbol{e}^t\boldsymbol{f}$ (why?) which is 0, giving $\lambda_2 = 0$. This yields $\boldsymbol{C}\boldsymbol{f} = \lambda_1\boldsymbol{f}$ proving that $\boldsymbol{f}$ is an eigenvector of $\boldsymbol{C}$, and it must correspond to the second highest eigenvalue since we assumed distinct eigenvalues, and because $\boldsymbol{f}^t\boldsymbol{C}\boldsymbol{f} = \lambda_1$.

Likewise, we have $J_2(\boldsymbol{g}) = \boldsymbol{g}^t\boldsymbol{C}\boldsymbol{g} - \tilde{\lambda}_1(\boldsymbol{g}^t\boldsymbol{g} - 1) - \lambda_2(\boldsymbol{f}^t\boldsymbol{g}) - \lambda_3\boldsymbol{e}^t\boldsymbol{g}$. Taking the derivative w.r.t. $\boldsymbol{g}$ and setting it to zero, we get $2\boldsymbol{C}\boldsymbol{g} - 2\tilde{\lambda}_1\boldsymbol{g} - \lambda_2\boldsymbol{f} - \lambda_3\boldsymbol{e} = \boldsymbol{0}$. Taking dot product on both sides by $\boldsymbol{e}$, we obtain $2\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{g} - 2\tilde{\lambda}_1\boldsymbol{e}^t\boldsymbol{g} - \lambda_2\boldsymbol{e}^t\boldsymbol{f} - \lambda_3\boldsymbol{e}^t\boldsymbol{e} = 0$ which gives us $\lambda_3 = 0$ since $\boldsymbol{e}^t\boldsymbol{C}\boldsymbol{g} = \tilde{\lambda}_1\boldsymbol{e}^t\boldsymbol{g} = 0$. Likewise, taking a dot product on both sides with $\boldsymbol{f}$, we obtain $\lambda_2 = 0$. This finally yields $\boldsymbol{C}\boldsymbol{g} = \tilde{\lambda}_1\boldsymbol{g}$ which shows that $\boldsymbol{g}$ is an eigenvector of $\boldsymbol{C}$ which now must correspond to the third largest eigenvalue, since $\boldsymbol{g}^t\boldsymbol{C}\boldsymbol{g}$ is to be maximized.
Marking scheme: 9 points for $\boldsymbol{f}$ and 6 points for $\boldsymbol{g}$. Out of the 9 points for $\boldsymbol{f}$, there are 3 points for $J_1(.)$, 3 points for the derivative, 1.5 points for arguing that $\lambda_2 = 0$ and 1.5 points for the final conclusion that $\boldsymbol{f}$ is an eigenvector with the second highest eigenvalue. Out of the 6 points for $\boldsymbol{g}$, there are 1.5 points for $J_2(.)$, 1.5 points for the derivative of $J_2(.)$ w.r.t. $\boldsymbol{g}$, 1.5 points for arguing that $\lambda_2 = \lambda_3 = 0$ and 1.5 points for the final conclusion that $\boldsymbol{g}$ is an eigenvector.

3. 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 *eigenvectors* 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 $\boldsymbol{A}$.

Solution:

(a) Let the SVD of $\boldsymbol{A}$ give us $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$. We have $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{S}^T\boldsymbol{U}^T = \boldsymbol{U}\tilde{\boldsymbol{S}}\boldsymbol{U}^T$ where $\tilde{\boldsymbol{S}}$ is defined in the manner described in the subquestion below. Clearly the diagonal elements of $\tilde{S}$ contain the squares of the diagonal elements of $\boldsymbol{S}$. The matrix $\tilde{\boldsymbol{S}}$ contains the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$. Hence the singular values of $\boldsymbol{S}$ are the positive square-roots of the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$. Likewise, we have $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{V} = \boldsymbol{V}^T\tilde{\boldsymbol{S}}\boldsymbol{V}$. Clearly the diagonal elements of $\tilde{S}$ contain the squares of the diagonal elements of $\boldsymbol{S}$. The matrix $\tilde{\boldsymbol{S}}$ contains the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$.

(b) Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$. Then $\|\boldsymbol{A}\|_F^2 = \text{trace}(\boldsymbol{A}^T\boldsymbol{A}) = \text{trace}(\boldsymbol{V}\boldsymbol{S}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T) = \text{trace}(\boldsymbol{V}\tilde{\boldsymbol{S}}\boldsymbol{V}^T) = \text{trace}(\tilde{\boldsymbol{S}}\boldsymbol{V}\boldsymbol{V}^T) = \text{trace}(\tilde{\boldsymbol{S}})$ which is the sum of the squares of the singular values of $\boldsymbol{A}$. The derivation uses the fact that $\boldsymbol{S}^T\boldsymbol{S} = \tilde{\boldsymbol{S}}$. Here $\tilde{\boldsymbol{S}} = \boldsymbol{S}$ is $\boldsymbol{S}$ is a square matrix. If $\boldsymbol{S}$ has size $m \times n$ where $m < n$, we see that $\tilde{\boldsymbol{S}}$ is a diagonal matrix of size $n \times n$ whose first m entries contains the square of the diagonal elements of $\boldsymbol{S}$ and the rest are zeros. If $m > n$, then $\tilde{\boldsymbol{S}}$ is a diagonal matrix of size $n \times n$ whose diagonal entries contain the square of the diagonal elements of $\boldsymbol{S}$.

(c) The trick here is to observe that $\boldsymbol{U}$ is obtained from the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{V}$ is obtained from the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$. A naive implementation with independently using eig for obtaining $\boldsymbol{U}$ and $\boldsymbol{V}$ will lead to sign inconsistencies leading to erroneous results (i.e. $\boldsymbol{A}$ will not be equal to the computed $\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$. Why does a sign inconsistency arise? It arises because if if $\boldsymbol{u_x}$ is an eigenvector of matrix $\boldsymbol{X}$ with eigenvalue λ, then $-\boldsymbol{u_x}$ is also an eigenvector of $\boldsymbol{X}$ with the *same* eigenvalue λ. However, you can get around the sign inconsistencies by observing that if $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}\boldsymbol{A}^T$ with eigenvalue λ (thus a left singular vector in $\boldsymbol{U}$), then $\boldsymbol{A}^T\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}^T\boldsymbol{A}$ with the same eigenvalue (thus the corresponding right singular vector in $\boldsymbol{V}$). This effectively takes care of all sign ambiguities. See the code MySVD.m for further details.
There is another solution to this: Check out a $\boldsymbol{u}, \boldsymbol{v}$ pair which have a sign inconsistency. Multiply either the $\boldsymbol{u}$ or $\boldsymbol{v}$ vector in such cases by -1. **Marking scheme:** For pointing out the issue of sign inconsistency, there are 4 points. There are 4 points for suggesting how to overcome the sign inconsistency problem.

(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?
ANSWER:
$\boldsymbol{y^t P y} = \boldsymbol{y^t A^T A y} = \boldsymbol{(Ay)^t(Ay)}$, which is the squared magnitude of the vector $\boldsymbol{Ay}$. Hence $\boldsymbol{y^t P y} \geq 0$. Similarly, $\boldsymbol{z^t Q z} \geq 0$. Now suppose $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{P}$ with eigenvalue λ, then $\boldsymbol{Pu} = \lambda\boldsymbol{u}$, and hence $\boldsymbol{u^t P u} = \lambda\boldsymbol{u^t u} = \lambda$. As $\boldsymbol{u^t P u} \geq 0$, we must have $\lambda \geq 0$. Likewise for $\boldsymbol{Q}$. Thus $\boldsymbol{P}$ and $\boldsymbol{Q}$ are positive semi-definite matrices.

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 $\boldsymbol{v}$?
ANSWER:

$$\begin{aligned}
\boldsymbol{Pu} &= \lambda \boldsymbol{u} \\
\therefore \boldsymbol{APu} &= \lambda \boldsymbol{Au} \\
\therefore \boldsymbol{A(A^TA)u} &= \lambda \boldsymbol{Au} \\
\therefore \boldsymbol{(AA^T)(Au)} &= \lambda(\boldsymbol{Au}) \\
\therefore \boldsymbol{Q(Au)} &= \lambda(\boldsymbol{Au})
\end{aligned} \tag{1}$$

$$\begin{aligned}
\boldsymbol{Qv} &= \mu \boldsymbol{v} \\
\therefore \boldsymbol{A^TQv} &= \mu \boldsymbol{A^Tv} \\
\therefore \boldsymbol{A^T(AA^T)v} &= \mu \boldsymbol{A^Tv} \\
\therefore \boldsymbol{(A^TA)(A^Tv)} &= \mu\boldsymbol{(A^Tv)} \\
\therefore \boldsymbol{P(A^Tv)} &= \mu\boldsymbol{(A^Tv)}
\end{aligned} \tag{2}$$

Note that $\boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m$.

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 $\boldsymbol{Au}_i = \gamma_i \boldsymbol{v}_i$. **ANSWER:**

$$\begin{aligned}
\boldsymbol{Qv_i} &= \mu_i \boldsymbol{v_i} \\
\therefore \boldsymbol{AA^Tv_i} &= \mu_i \boldsymbol{v_i} \\
\therefore \boldsymbol{A(A^Tv_i)} &= \mu_i \boldsymbol{v_i} \\
\therefore \boldsymbol{A(A^Tv_i)/\|A^Tv_i\|_2} &= \mu_i \boldsymbol{v_i}/\|\boldsymbol{A^Tv_i}\|_2 \\
\therefore \boldsymbol{Au_i} &= (\mu_i/\|\boldsymbol{A^Tv_i}\|_2)\boldsymbol{v_i} \\
\therefore \boldsymbol{Au_i} &= \gamma_i \boldsymbol{v_i}
\end{aligned} \tag{3}$$

where we define the scalar $\gamma_i = \mu_i/\|\boldsymbol{A^Tv_i}\|_2$. Note that γ_i is non-negative as μ_i is non-negative (see part (a)) and $\|\boldsymbol{A^Tv_i}\|_2$ is non-negative since it is the magnitude of a vector.

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.[1]. 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 $\boldsymbol{A} = \boldsymbol{U\Gamma V}^T$ 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 $\boldsymbol{A}$. This is a key result in linear algebra and it is widely used in image processing, computer vision, computer graphics, statistics, machine learning, numerical analysis, natural language processing and data mining. [7.5 + 7.5 + 7.5 + 7.5 = 30 points]

ANSWER: We saw that $\boldsymbol{Au_i} = \gamma_i\boldsymbol{v_i}$. Assume $m \leq n$ without loss of generality. Then we can write $\forall i, 1 \leq i \leq m, \boldsymbol{Au_i} = \gamma_i \boldsymbol{v_i}$ and $\forall i, m+1 \leq i \leq n, \boldsymbol{Au_i} = \boldsymbol{0}$. So we can write $\boldsymbol{AV} = \boldsymbol{U\Gamma}$, where $\boldsymbol{V}$ is an orthonormal matrix of size $n \times n$, $\boldsymbol{\Gamma}$ is a diagonal matrix of size $m \times n$ containing at the most m non-zero values along its diagonal, and $\boldsymbol{U}$ is an orthonormal matrix of size $m \times m$. Hence we can say $\boldsymbol{A} = \boldsymbol{U\Gamma V^T}$. (Strictly speaking, we need to define a matrix $\tilde{\boldsymbol{V}} = (\boldsymbol{u_1}|\boldsymbol{u_2}|...|\boldsymbol{u_m}|\boldsymbol{u_{m+1}}|...|\boldsymbol{u_n})$ with the understanding that for i lying from $m+1$ to n, you will have to use the fact that $\boldsymbol{Au_i} = \boldsymbol{0}$. Likewise, we will define a $m \times n$ matrix $\tilde{\boldsymbol{\Gamma}}$ which is diagonal. This will yield $\boldsymbol{A\tilde{V}} = \boldsymbol{U\tilde{\Gamma}}$, and then $\boldsymbol{A} = \boldsymbol{U\tilde{\Gamma}\tilde{V}}^T$.) Note that the footnote in the question argues why $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal, and this is an important step. It is also important to realize that values in $\boldsymbol{\Gamma}$ are non-negative, and we have made that argument previously. **MARKING SCHEME:** 4 points for each part. In the last part, if the student has shown that $\boldsymbol{AV} = \boldsymbol{U\Gamma}$, that is sufficient. There are no points to be deducted if the student hasn't shown $\boldsymbol{A} = \boldsymbol{USV}^T$.

[1]This follows because $\boldsymbol{P}$ and $\boldsymbol{Q}$ are symmetric matrices. Consider $\boldsymbol{Pu}_1 = \lambda_1\boldsymbol{u}_1$ and $\boldsymbol{Pu}_2 = \lambda_2\boldsymbol{u}_2$. Then $\boldsymbol{u}_2^T\boldsymbol{Pu}_1 = \lambda_1\boldsymbol{u}_2^T\boldsymbol{u}_1$. But $\boldsymbol{u}_2^T\boldsymbol{Pu}_1$ also equal to $(\boldsymbol{P}^T\boldsymbol{u}_2)^T\boldsymbol{u}_1 = (\boldsymbol{Pu}_2)^T\boldsymbol{u}_1 = (\lambda_2\boldsymbol{u}_2)^T\boldsymbol{u}_1 = \lambda_2\boldsymbol{u}_2^T\boldsymbol{u}_1$. Hence $\lambda_2\boldsymbol{u}_2^T\boldsymbol{u}_1 = \lambda_1\boldsymbol{u}_2^T\boldsymbol{u}_1$. Since $\lambda_2 \neq \lambda_1$, we must have $\boldsymbol{u}_2^T\boldsymbol{u}_1 = 0$.

4. In this part, you will implement a mini face recognition system. Download the ORL face database from the homework folder. It contains 40 sub-folders, one for each of the 40 subjects/persons. For each person, there are ten images in the appropriate folder named 1.pgm to 10.pgm. The images are of size 92 by 110 each. Each image is in the pgm format. You can view/read the images in this format, either through MATLAB or through image viewers like IrfanView on Windows, or xv/display/gimp on Unix. Though the face images are in different poses, expressions and facial accessories, they are all roughly aligned (the eyes are in roughly similar locations in all images). For the first part of the assignment, you will work with the images of the first 32 people (numbers from 1 to 32). For each person, you will include the first six images in the training set (that is the first 6 images that appear in a directory listing as produced by the `dir` function of MATLAB) and the remaining four images in the testing set. You should implement the recognition system by using the `eig` or `eigs` function of MATLAB on an appropriate data matrix. Record the recognition rate using squared difference between the eigencoefficients while testing on all the images in the test set, for $k \in \{1, 2, 3, 5, 10, 15, 20, 30, 50, 75, 100, 150, 170\}$. Plot the rates in your report in the form of a graph. Now modify the required few lines of the code but using the `svd` function of MATLAB (on the $\boldsymbol{L}$ matrix as defined in class) instead of `eig` or `eigs`.

   Repeat the same experiment (using just the `eig` or `eigs` routine) on the Yale Face database from the homework folder. This database contains about 64 images each of 38 individuals (*labeled from 1 to 39, with number 14 missing; some folders have slightly less than 64 images*). Each image is in pgm format and has size 192 by 168. The images are taken under different lighting conditions but in the same pose. Take the first 40 images of every person for training and test on the remaining 24 images (that is the first 40 images that appear in a directory listing as produced by the `dir` function of MATLAB). Plot in your report the recognition rates for $k \in \{1, 2, 3, 5, 10, 15, 20, 30, 50, 60, 65, 75, 100, 200, 300, 500, 1000\}$ based on (a) the squared difference between all the eigencoefficients and (b) the squared difference between all *except* the three eigencoefficients corresponding to the eigenvectors with the three largest eigenvalues. Display in your report the reconstruction of any one face image from the ORL database using $k \in \{2, 10, 20, 50, 75, 100, 125, 150, 175\}$ values. Plot the 25 eigenvectors (eigenfaces) corresponding to the 25 largest eigenvalues using the subplot or subimage commands in MATLAB. [30 points]
   Answer: See model code part1orl.m and part1yale.m. The typical recognition rates are also mentioned there. No penalty if the rates are within a range of $\pm 10\%$ of the ones mentioned there. See model code part2.m.

5. What will happen if you test your system on images of people which were not part of the training set? (i.e. the last 8 people from the ORL database). What mechanism will you use to report the fact that there is no matching identity? Work this out carefully and explain briefly in your report. Write code to test whatever you propose on all the 32 remaining images (i.e. 8 people times 4 images per person), as also the entire test set containing 6 images each of the first 32 people. How many false positives/negatives did you get? [10 points]
   Answer: See model code part3.m for a strategy that determines a threshold τ for the distance between the top k eigencoefficients of different images of the same person. An image of the same person producing a distance greater than τ w.r.t. all trainings images of the same person is a false negative. An image of a person producing a distance less than τ w.r.t. all training images of a different person yields a false positive.