Digital Image Processing HW4

Parth Pujari and Aayushi Barve

1 Question 1

For the **ORL** images, the following is the reconstruction rate (accuracy) versus the number of principal components. All the results below are the same when we use `eig`, `svd` on the L matrix or `eig` on the L matrix. (All three implementations are in the PCA function with options 0, 1 and 2 for the three respectively).

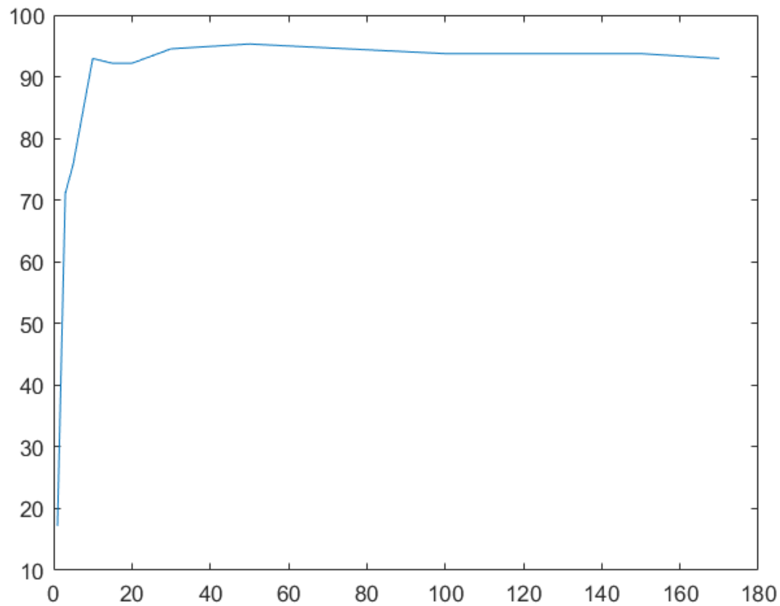

Figure 1: Accuracy vs k

The following are the top 25 eigen faces. Note that their spread is **normalized**. That is, their maximum intensity is 255 and the minimum is 0,

Figure 2: Eigenfaces

The following is am image from the ORL dataset.

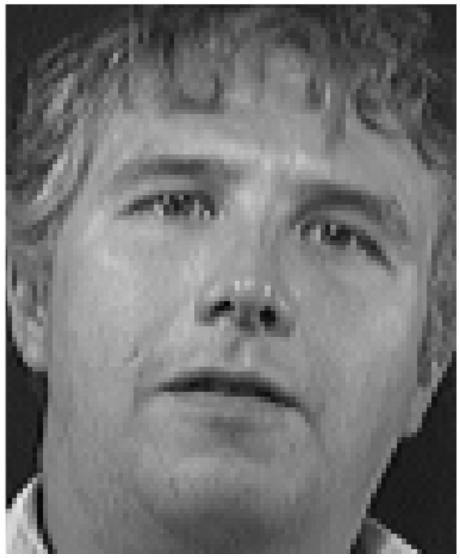

Figure 3: Face

The following are its reconstructions for values of k given in the question.

Figure 4: Reconstructions

For the Yale dataset, the accuracy versus k while considering the first 3 largest eigenvalues and (for k greater than 3) without the 3 largest eigenvalues are plotted below.

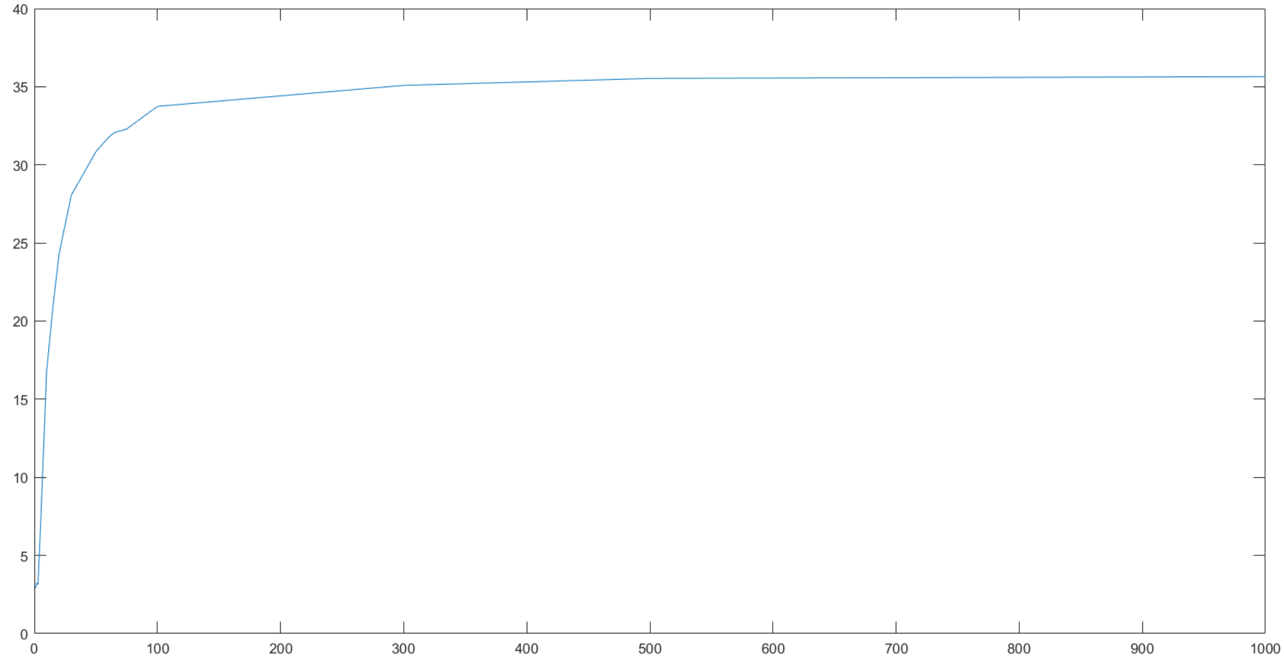

Figure 5: With top 3

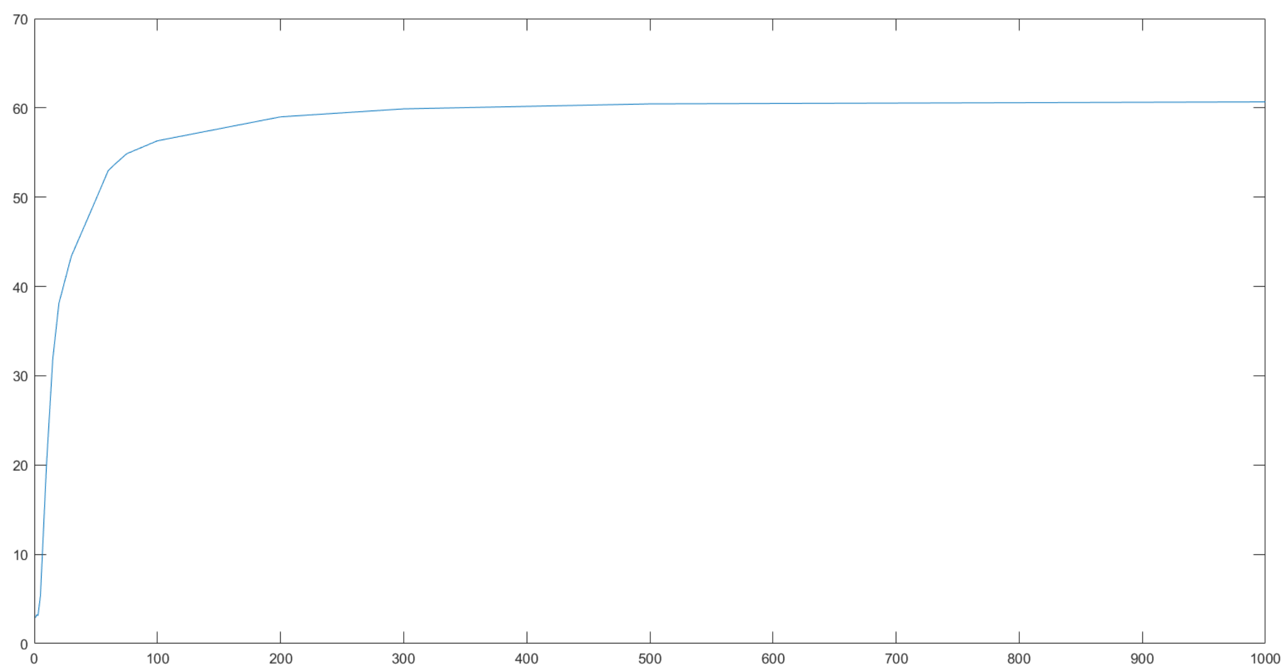

Figure 6: Without top 3

2 Question 2

Images that do not belong to the dataset need not have their image vector aligned along the principal components. What is certain is that images belonging to the dataset have small coefficients along eigenvalues corresponding to small eigenvalues. So we come up with the following ways to determine if an image belongs to a dataset:

1. Check if the eigen-coefficients corresponding to small eigenvalues are small. If not, the image is not a part of the dataset.
2. Check if the reconstruction error is large for all training images.
3. Check if the eigen-coefficients along the principal components are small. If so, the images are likely not from the dataset.

We use method 2 and 1. The false positive cases are 10-14 out of 80 and the false negatives are 28-33 out of 130. The implementation is in the same code.

3 Question 3

$\mathcal{X} = \{x_1, x_2, ...x_N\}_i \in \mathbb{R}^d, \bar{x} = (1/N)\sum_{x_i \in \mathcal{X}} x_i$. Consider the minimizing problem,

$$\min_e \sum_{i=1}^{N} ||x_i - \bar{x} - (e \cdot (x_i - \bar{x})e)||^2$$

as an equivalent to maxmizing:

$$e^T Ce$$

Where C is the covariance matrix of $\mathcal{X}$
We used Lagrangian multipliers to solve this inequality. I propose another method which also gives the direction f' perpendicular to e for which $f'Cf$ is maximized.
Since C is real and symmetric, by the **spectral theorem** we have

$$C = UDU^T$$

where $U \in \mathbb{R}^{N \times N}$ is orthonormal (and hence full rank) and D is diagonal and its diagonal elements are eigen values of C (for the sake of simplicity assume we arrange eigen values in descending order).
Consider any vector v, $v^T Cv = v^T UDU^T v = w^T Dw$ for some $w \neq 0$ because U (and hence U^T) has 0 nullity. Further, any $w^T Dw = v^T UDU^T v$ because U (and hence U^T) spans $\mathbb{R}^N$. Thus e is a solution of

$$\max_v v^T Cv$$

if and only if it is a solution of

$$\max_w w^T Dw$$

Further, $||v|| = 1$ iff $||w|| = 1$ since $||w|| = ||U^T v|| = v^T UU^T v = v^T v = ||v|| = 1$.
We write $w = (w_1, w_2, ...w_N)^T$. Then $w^T Dw$ is:

$$w_1^2 e_1 + w_2^2 e_2 + ...w_N^2 e_N$$

where $e_1, e_2....e_N$ are the diagonal elements of D ($e_1 \geq e_2 \geq ... \geq e_N$) with $w_1^2 + w_2^2 + ...w_N^2 = 1$.
This is maximized when $w_1 = 1$ and $w_2, ...w_N = 0$.
Proof by induction on N. The case when $N = 1$ is trivially true since eigen values are positive. Assume true for $N = n$, i.e.,

$$\max w_1^2 e_1 + w_2^2 e_2 + ...w_n^2 e_n$$

where ($e_1 \geq e_2 \geq ... \geq e_n$) with $w_1^2 + w_2^2 + ...w_n^2 = 1$ is e_1, i.e., $w_1 = 1$ and $w_i = 0 \; \forall i > 0$. Then consider

$$\max w_1^2 e_1 + w_2^2 e_2 + ...w_{n+1}^2 e_{n+1}$$

Assume $w_1 \neq 1$ for optimality, then atleast one of $w_2, ...w_{n+1} \neq 0$.
$\max w_2^2 e_2 + ...w_{n+1}^2 e_{n+1}$ subject to $w_2^2 + ...w_{n+1}^2 = (1 - w_1^2)$ is $(1 - w_1^2)e_2$ by induction hypothesis (scaling all weights by $(1 - w_1^2)$). Thus, $\max w_1^2 e_1 + w_2^2 e_2 + ...w_{n+1}^2 e_{n+1} = w_1^2 e_1 + (1 - w_1^2)e_2, w_1^2 \leq 1$ is max when $w_1 = 1$, contradiction. Thus $w_1 = 1$.
Note that $w = (1, 0, ...0)^T$ gives $v = Uw$ which is the first column of U or the eigenvector corresponding to the largest eigenvalue.
Now, f orthonormal to e corresponds to w' orthonormal to $w = (1, 0, ...0)^T$. Thus w'_1 (first element of w') must be 0. This corresponds to

$$\max w_2^2 e_2 + ...w_N^2 e_N$$

${w'_2}^2 + {w'_3}^2 + ...{w'_N}^2 e_N 2 = 1$. As we've seen this is equal to e_2 with $w'_2 = 1$ and $w'_3, ...w'_N = 0$. Thus $w' = (0, 1, 0, ...1)^T$ and $f = Uw'$ is the second column of U or the eigenvector corresponding to the second largest eigen value. The proof for g orthonormal to e and f maximizing $g^T Cg$ is similar.

Since g is orthonormal to e and f, $w'' = U^T g$ has its first and second elements as 0 and the maximization is over $e_3, e_4, ...e_N$. This gives us e_3 as the maximum and g is the eigenvector corresponding to the third largest eigen value.

4 Question 4

a) The covariance matrix $C = XX^T$ where X is the mean subtracted image matrix. A symmetric matrix A is said to be positive semi-definite if $\forall v, v^T A v \geq \mathbf{0}$.
Then for any vector v we have $v^T C v = v^T X X^T v = (X^T v)^T (X^T v) = w^T w \geq \mathbf{0}$ (since $x^T x \geq \mathbf{0}\ \forall x$).

b) For any symmetric matrix A, if we have 2 distinct eigen vectors $\lambda_1 \neq \lambda_2$, then for some $v_1, v_2 \neq \mathbf{0}$

$$Av_1 = \lambda_1 v_1, \quad Av_2 = \lambda_2 v_2$$
$$v_2^T A v_1 = \lambda_1 v_2^T v_1$$

But,

$$\begin{aligned} v_2^T A v_1 &= v_2^T A^T v_1 \\ &= (v_1^T A v_2)^T \\ &= v_1^T A v_2 \end{aligned}$$

Since $v_1^T A v_2$ is a scalar. But this is equal to:

$$\begin{aligned} &v_1^T(\lambda_2 v_2) \\ &= \lambda_2 v_1^T v_2 \\ &= \lambda_2 v_2^T v_1 \end{aligned}$$

Thus, $\lambda_1 v_2^T v_1 = \lambda_2 v_2^T v_1$ and since $\lambda_1 \neq \lambda_2$, $v_2^T v_1 = 0$

c) Consider $\frac{1}{N}\sum_{i=1}^{N} ||x_i - \tilde{x_i}||$.
Let $x_i = \bar{x} + V\alpha_i$ and $\tilde{x_i} = \bar{x}V\tilde{\alpha_i}$ Where $\tilde{\alpha_i} = (\alpha_{i1}, \alpha_{i2}, ...\alpha_{ik}, 0, 0, \ldots)^T$ where α_{ij} is the jth element of α.
$\frac{1}{N}\sum_{i=1}^{N} ||x_i - \tilde{x_i}|| = \frac{1}{N}\sum_{i=1}^{N} ||V(\alpha_i - \tilde{\alpha_i})|| = \frac{1}{N}\sum_{i=1}^{N} ||\alpha_i - \tilde{\alpha_i}||$ (because V is orthonormal)
$= \frac{1}{N}\sum_{i=1}^{N} ||(0, 0, \ldots, \alpha_{ik+1}, \ldots \alpha_{id})|| = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=k+1}^{d} \alpha_{i\,j}^2 = \frac{1}{N}\sum_{j=k+1}^{d}\sum_{i=1}^{N} \alpha_{i\,j}^2$.
Now we show that this is small.
Denote the covariance matrix XX^T by C.
$XX^T = C = VDV^T$ where the elements in the principal diagonal of D are arranged in decreasing order. The coefficient matrix $J = V^T X \in \mathbb{R}^{d \times N}$ is such that the jth column of J contains the d coefficients of the jth image. Then, $JJ^T = (V^T X)(V^T X)^T = V^T (XX^T) V = D$.
But the $(JJ^T)_{p,p} = \sum_{i=1}^{N} \alpha_{i\,p}^2$, i.e., the sum over all images of the pth eigen coefficient squared. This is the pth diagonal element of D or the pth eigen value $_p$ and is small for $p > k$. Therefore the sum over eigen coefficients for $p > k$ is also small. Precisely,
$\frac{1}{N}\sum_{j=k+1}^{d}\sum_{i=1}^{N} \alpha_{i\,j}^2 = \frac{1}{N}\sum_{j=k+1}^{d} \lambda_j$

The final anwer is

$$\frac{1}{N}\sum_{i=1}^{N} ||x_i - \tilde{x_i}|| = \frac{1}{N}\sum_{j=k+1}^{d} \lambda_j$$

d) Consider N samples of $(X1, X2)$. Then the covariance matrix is $\begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}$ This has eigen values 100 and 1 with eigen vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ Since 100 >> 1, the principal component is along $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

For the second case, the variances are equal and both eigen values are the same. Hence the principal components are along $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

5 Question 5

a) $A = USV^T$ where S is $(m \times n, (m \leq n))$ diagonal (not necessarily square), U and V are orthonormal. Then consider AA^T.

$$AA^T = USS^TU^T \tag{1}$$

But SS^T is square and diagonal. $(SS^T)_{i,i} = (S_{i,i})^2$. But (4) is the eigen decomposition of AA^T. Hence the diagonal elements of S are the square roots of the eigen values of AA^T.
Similarly $A^TA = VS^TSV^T$. But $(S^TS)_{i,i} = (S_{i,i})^2$ for $i \leq m$ and a similar argument follows.

b) Frobenius norm of A is the sum of squares of all elements in A. This is equal to the trace of AA^T.

$$tr(AA^T) = tr(USS^TU^T) = tr(SS^TU^TU) = tr(SS^T)$$

using the fact that $tr(AB) = tr(BA)$ for all matrices A, B.
This is equal to the sum of squares of diagonal elements in S or the sum of squares of singular values of A.

c) Consider AA^T. Perform eig on it to get U. Consider each column of U. It corresponds to an eigenvector of AA^T. However, for each column u of U, $-u$ could also be an eigen vector of AA^T.
We note that changing a sign of a column of U gives U'. Then, U' is orthonormal since $u^Tv = 0 = -u^Tv$ if $u \neq v$ and $||u|| = ||-u|| = 1$. Further, $UDU^T = U'DU'^T$.

Therefore the sign of each column of U could be anything. Similarly for any column of V found from eig on A^TA. Then, $UD^{1/2}V^T$ is not necessarily A (where $D^{1/2}$ is the rectangular diagonal matrix formed from square roots of eigen values of AA^T or A^TA). A simple example is:

$$\begin{aligned}
A &= UD^{1/2}V^T, \\
AA^T &= (-U)D(-U)^T \\
A^TA &= VD'V^T \\
-A &= (-U)D^{1/2}V^T \neq A
\end{aligned}$$