Boston University
Department of Electrical and Computer Engineering

ENG EC 414 (Ishwar) Introduction to Machine Learning

HW 5

Issued: Fri 18 Mar 2022 **Due:** 10:55pm Boston time Fri 25 Mar 2022 in Gradescope + Blackboard.
Required reading: Slides on PCA + your notes from lectures & Discussion 7.

Important: Before you proceed, please read the documents pertaining to *Homework formatting and submission guidelines* and the *HW-grading policies* in the Homeworks section of Blackboard, especially guidelines for submitting reports in Gradescope. and code in Blackboard. In particular, for computer assignments *you are prohibited from using any online code or built-in MATLAB functions except as indicated in the problem or skeleton code (when provided).*
In order to receive full credit, all work should be supported by a concise explanation that is clear, relevant, specific, logical, and correct. In particular, for each part, you must clearly outline the key steps and provide proper justification for your calculations.

Note: Problem difficulty = number of coffee cups ☕

Problem 5.1 [35pts] *(PCA Analytical Exercises)*

(a) [10pts] Consider the following data points:

$$\mathbf{x}_1 = \begin{bmatrix}1\\3\end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix}1\\7\end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix}0\\5\end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix}2\\5\end{bmatrix},$$

Hand-compute the empirical covariance matrix of the above four data points. Then hand-compute the eigenvector of the covariance matrix corresponding to its largest eigenvalue (this is the first principal direction) and the corresponding principal components of the four data points along this direction.

(b) [25pts] ☕☕ Let $\mathbf{r} = (-1,\ 2,\ -1,\ 0,\ 2)^\top$ and $\mathbf{s} = (1,\ -1,\ -1,\ 1,\ 1)^\top$. Consider the following unlabeled dataset consisting of 8 feature vectors in 5 dimensional Euclidean space: $\mathbf{x}_1 = 3\mathbf{r}$, $\mathbf{x}_2 = \mathbf{r}$, $\mathbf{x}_3 = -\mathbf{r}$, $\mathbf{x}_4 = -3\mathbf{r}$, $\mathbf{x}_5 = 3\mathbf{s}$, $\mathbf{x}_6 = 2\mathbf{s}$, $\mathbf{x}_7 = \mathbf{s}$, $\mathbf{x}_8 = -6\mathbf{s}$.

(i) Without using computing devices, compute (with suitable explanation) all the eigenvalues of the empirical *covariance* matrix of the dataset.

(ii) Without using computing devices, compute (with suitable explanation) the top two principal directions.

Now suppose that the first four feature vectors have label +1 and the remaining have label −1.

(iii) If we decide the class based on whether the *first* principal component is above or below a threshold, what is the highest CCR we can get for this dataset and for what threshold?

(iv) If we decide the class based on whether the *second* principal component is above or below a threshold, what is the highest CCR we can get for this dataset and for what threshold?

Hint: If $A = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$ and all the $\mathbf{u}_i$'s are unit-norm vectors in $\mathbb{R}^d$ which are pairwise orthogonal, then they are the orthonormal eigenvectors of A with λ_i's as the corresponding eigenvalues.

Problem 5.2 [60pts] *(PCA of Images)* In this problem we will perform Principal Component Analysis (PCA) of two different image datasets: the AT&T Face Dataset and the MNIST Hand-Written Digits Dataset.

Dataset Information:

(1) **AT&T Face Dataset:** This dataset contains 400 images (10 images of 40 diffferent people) of size 112×92 (10304 pixels) in the Portable Grayscale Map (PGM) format. We have provided you this dataset as `att-database-of-faces.zip` and a helper function `load_faces.m` to load the dataset into your MATLAB workspace as a matrix of "vectorized" images of dimensions 400×10304. Unzip the dataset and place the resulting folder named `att-database-of-faces` in the same directory as your solution file to use `load_faces.m`.

(2) **MNIST Hand-Written Digits Dataset:** We provide you with `mnist256.mat` which is a data structure that contains the `mnist` dataset of 7291 gray scale hand-written images of size 16×16 (256 pixels). Use Matlab's `load` command to load `mnist256.mat` and extract all images into a 7291×257 matrix named `mnist` in your workspace. The first column of `mnist` will contain the image labels (the numeral represented by the image) and columns 2 through 257 will contain the vectorized pixel values of the image.

Tasks: Use the AT&T Face Dataset for parts (a) through (c) and the MNIST dataset for part (d).

(a) [5pts] *(Data Visualization)* Use the provided `load_faces.m` function and load the face-dataset into your workspace. Compute the mean face image, that is the mean of all the 400 face image vectors. Use `subplot` and `imshow(uint8(reshape(img_vector, img_size)))` to create a figure containing two images: (1) image #120 in the dataset, and (2) the mean face of the dataset. Does the mean face resemble a human face or is it a smooth shapeless "blob"?

completed in part A b

(b) [20pts] *(Eigenvalues of Covariance Matrix)* Subtract the vectorized mean face image $\hat{\boldsymbol{\mu}}_x$ from all the vectorized face images and compute the empirical covariance matrix $\hat{\mathrm{S}}_x$ of the mean-centered vectorized face images. Perform an Eigenvalue Decomposition of this covariance matrix using the inbuilt MATLAB fucntion `eig`. Arrange the eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots$ in the order of non-increasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$, into an orthonormal matrix $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots]$. Let Λ be the corresponding diagonal matrix of eigenvalues. wasn't in order confusing...

(1) [5pts] Report the values of the first five eigenvalues. (2) [5pts] Plot λ_k as a function of k, for $k = 1, 2, \ldots, 450$. Comment on the observed trends in your report. Explain why $\lambda_k = 0$ for all $k > 400$. (3) [5pts] Compute and plot (as a function of k) the values of the so-called "fraction of variance explained" by the top k principal components:

$$\rho_k := \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}.$$

Round the values of ρ_k to 2 decimal places. Comment on the observed trends in your report. (4) [5pts] Find and report the smallest values of k for which $\rho_k \geq 0.51, 0.75, 0.90, 0.95,$ and 0.99.

Notes: Use subplots to show plots of λ_k and ρ_k in the same figure. The reshaped principal directions of a human face dataset are often referred to as eigenfaces.

(c) [9pts] ☕ *(Image Approximation through Dimensionality Reduction)* The first eigenface corresponds to the largest eigenvalue and represents a direction that encompasses the largest variance in the training data, the second eigenface corresponds to the second largest eigenvalue and represents a direction which is orthogonal to the first eigenface and encompasses the second largest variance in the training data, and so on. A linear combination of eigenfaces along with a mean face, can be used to reconstruct a given face. Specifically, let $\mathbf{x}$ denote the vectorized representation of an image. An approximation to $\mathbf{x}$ based on the top k principal components is given by

$$\hat{\mathbf{x}} = \hat{\boldsymbol{\mu}}_x + \sum_{i=1}^{k} \hat{y}_{i,PCA} \cdot \mathbf{u}_i, \quad \text{where} \quad \hat{y}_{i,PCA} = \mathbf{u}_i^\top (\mathbf{x} - \hat{\boldsymbol{\mu}}_x)$$

[1pt per image] Generate a single figure (using subplots) showing the following: the mean face, the k principal component approximation face for image #43 in the face dataset for $k = 1$, $k = 2$, all the k values that you found in part (b) sub-part (4), and then show the original image in the same figure too. Add a descriptive title to each image in the figure and a title to the overall figure (see skeleton code for reference).

(d) [26pts] ☕☕ *(Connecting Principal Components with Image Features)* For this problem, load the MNIST data set. Filter the dataset using the label column to extract all images of the numeral '3' into one matrix (*Hint*: use conditional/logical based indexing into MATLAB matrices). We will use this filtered dataset of images of numeral '3' for this problem.

(1) [5pts] Compute the mean image of the numeral '3' and use subplot and `imshow(reshape(img_vector, img_size))` to report image #120 in the filtered dataset and the mean image of the filtered dataset in a single figure.

Percentile Values: the P-th percentile ($0 < P \leq 100$) of a list of ordered values (sorted from least to greatest) is the smallest value in the list such that no more than P percent of the data is strictly less than the value and at least P percent of the data is less than or equal to that value.

Compute the first two principal components for all images in the filtered dataset. Use the provided `percentile_values(v, percentiles)` function to compute the percentile values for the first principal components of all images and the second principal components of all images for the following percentiles: 5th, 25th, 50th, 75th, and 95th percentile.

Helpful functions for parts (2) and (3): `hold on, grid on, xticks, yticks, subplot`.

(2) [15pts] Generate a single figure containing the following plots. A scatter plot of the first two principal components of all images, i.e., image j will appear as a point with coordinates $(\hat{y}_{1j}, \hat{y}_{2j})$ in the scatter plot, where $\hat{y}_{1j}, \hat{y}_{2j}$ are, respectively, the first and second principal components of image j. Overlay a rectangular grid where the vertical grid lines occur at the percentile values of the first principal component and the horizontal grid lines occur at the percentile values of the second principal component. We will call this grid, the Percentiles Grid (PG). Find the intersection points of the grid lines in PG (*Hint:* Intersction points in PG are easily obtained via the Cartesian product of the percentile values of each principal component). Find the images whose first two principal components $(\hat{y}_{1j}, \hat{y}_{2j})$ are closest (in terms of Euclidean distance) to these intersection points and indicate their locations in the scatter plot using "filled" red dots (see documentation of `scatter`).

(3) [6pts] Display, in a single figure, the closest images from the filtered data that correspond to the red dots and discuss what image property (if any) of the numeral '3' does each principal component capture? (Be wary of the indices in this task).

Code-submission via Blackboard: Create one dot-m file. Name it as follows: `<yourBUemailID>_hw5_2.m` for Problem 5.2. All local functions and scripts pertaining to one problem should appear within the single dot-m file for that problem. When run, your scripts should be able to display in the command window whatever you are asked to compute and report in each part. One skeleton code is provided for your reference: `skeleton_hw5_2.m`. Reach-out to the TA via Piazza and office hours for questions related to coding. When submitting code, please include a ‘Readme’ text file in your source code directory describing approximate running times of different parts, any additional comments (as needed) explaining how to use your code, and any dependencies between different parts. Please do not include into the directory, any data files that are already provided. Write your code under the assumption that all data files are in the same directory as your source code.

Problem 5.1 [35pts] *(PCA Analytical Exercises)*

(a) [10pts] Consider the following data points:

$$\hat{S}_x \quad \mathbf{x}_1 = \begin{bmatrix}1\\3\end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix}1\\7\end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix}0\\5\end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix}2\\5\end{bmatrix},$$

Hand-compute the empirical covariance matrix of the above four data points. Then hand-compute the eigenvector of the covariance matrix corresponding to its largest eigenvalue (this is the first principal direction) and the corresponding principal components of the four data points along this direction.

(b) [25pts] Let $\mathbf{r} = (-1, 2, -1, 0, 2)^\top$ and $\mathbf{s} = (1, -1, -1, 1, 1)^\top$. Consider the following unlabeled dataset consisting of 8 feature vectors in 5 dimensional Euclidean space: $\mathbf{x}_1 = 3\mathbf{r}$, $\mathbf{x}_2 = \mathbf{r}$, $\mathbf{x}_3 = -\mathbf{r}$, $\mathbf{x}_4 = -3\mathbf{r}$, $\mathbf{x}_5 = 3\mathbf{s}$, $\mathbf{x}_6 = 2\mathbf{s}$, $\mathbf{x}_7 = \mathbf{s}$, $\mathbf{x}_8 = -6\mathbf{s}$.

(i) Without using computing devices, compute (with suitable explanation) all the eigenvalues of the empirical *covariance* matrix of the dataset.

(ii) Without using computing devices, compute (with suitable explanation) the top two principal directions.

Now suppose that the first four feature vectors have label +1 and the remaining have label −1.

(iii) If we decide the class based on whether the *first* principal component is above or below a threshold, what is the highest CCR we can get for this dataset and for what threshold?

(iv) If we decide the class based on whether the *second* principal component is above or below a threshold, what is the highest CCR we can get for this dataset and for what threshold?

Hint: If $A = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$ and all the $\mathbf{u}_i$'s are unit-norm vectors in $\mathbb{R}^d$ which are pairwise orthogonal, then they are the orthonormal eigenvectors of A with λ_i's as the corresponding eigenvalues.

a) $x = \begin{vmatrix} 1 & 1 & 0 & 2 \\ 3 & 7 & 5 & 5 \end{vmatrix}$

$\hat{S}_x \equiv$ covariance matrix

$\hat{\mu}_x = \frac{1}{n} \sum_{j:y_j=1}^{n} x_j$

$\hat{\mu}_x = \frac{1}{4} \begin{vmatrix} 4 \\ 20 \end{vmatrix}$

$\hat{\mu}_x = \begin{vmatrix} 1 \\ 5 \end{vmatrix}$

$\bar{X} - \hat{\mu}_x = \begin{vmatrix} 0 & 0 & -1 & 1 \\ -2 & 2 & 0 & 0 \end{vmatrix}$
$\bar{x} =$

$\hat{S} = \frac{1}{n} \sum^{n} \bar{x}_j \cdot \bar{x}_j^T$

$\hat{S}_x = \underset{2\times4}{\begin{vmatrix} 0 & 0 & -1 & 1 \\ -2 & 2 & 0 & 0 \end{vmatrix}} \underset{4\times2}{\begin{vmatrix} 0 & -2 \\ 0 & 2 \\ -1 & 0 \\ 1 & 0 \end{vmatrix}} \left(\frac{1}{4}\right)$

$\hat{S}_x = \frac{1}{4} \begin{vmatrix} 2 & 0 \\ 0 & 8 \end{vmatrix}$

$$\boxed{\hat{S}_x = \begin{vmatrix} 1/2 & 0 \\ 0 & 2 \end{vmatrix}}$$

Remember $|A - \lambda \cdot I| = 0$

$\begin{vmatrix} 1/2 - \lambda & 0 \\ 0 & 2 - \lambda \end{vmatrix} = (1/2 - \lambda)(2 - \lambda) = 0$

$2(1/2) - 1/2\lambda - 2\lambda + \lambda^2 = 0$

$\lambda^2 - 5/2\lambda + 1 = 0$

$\lambda^2 - 9/2\lambda + 2 = 0$

$\lambda_1 = 2$, $\lambda_2 = 1/2$

largest eigenvalue

$|A - 2| = 0$

$\begin{vmatrix} 1/2 - 2 & 0 \\ 0 & 2 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = 0$

$\begin{vmatrix} -3/2 & 0 \\ 0 & 2 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = 0$

$e_2/2 = 1$

$e_1 \cdot -\frac{2}{3} = 1$

$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = 0$

$v_1 = \begin{vmatrix} 1 \\ 0 \end{vmatrix}$ ← already normalized

$|A - 1/2| = 0$

$\begin{vmatrix} 1/2 - 1/2 & 0 \\ 0 & 2 - 1/2 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = 0$

$\begin{vmatrix} 0 & 0 \\ 0 & 3/2 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = 0$ $e_3 \cdot 3/2$

$\begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = 0$

$v_2 = \begin{vmatrix} 0 \\ 1 \end{vmatrix}$ ← already normalized (Didn't need to find)

$\hat{S}_x = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} 1/2 & 0 \\ 0 & 2 \end{vmatrix} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$

$\boxed{\begin{vmatrix} 1/2 & 0 \\ 0 & 4 \end{vmatrix}}$

Note if a matrix is diagonal, do not need to find eigenvectors. It's the identity.

$W_{PCA} = V = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$

Encode feature vector

$\langle (x_1 - \hat{\mu}_x), \mu_1 \rangle$

$\bar{X} - \hat{\mu}_x = \begin{vmatrix} 0 & 0 & -1 & 1 \\ -2 & 2 & 0 & 0 \end{vmatrix}$

$\bar{X} =$

x_1: $\langle \begin{pmatrix} 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle = 0$

$\underbrace{0 \begin{pmatrix} 1 \\ 0 \end{pmatrix}}_{\mu_1} + \underbrace{\begin{pmatrix} 1 \\ 5 \end{pmatrix}}_{\hat{\mu}_x} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$

x_2: $\langle \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle = 0$

$0 \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$

x_3: $\langle \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle = -2$

$-2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$

x_4: $\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle = 1$

$1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$

Principle Components:

$\boxed{\begin{vmatrix} 1 & 1 & -1 & 2 \\ 5 & 5 & 5 & 5 \end{vmatrix}}$

b) $X = \begin{vmatrix} 3r & r & -r & -3r & 3s & 2s & s & -6s \end{vmatrix}$

$3r + r - r - 3r$ $\quad 3s + 2s + s - 6s$

$0 \qquad 0$

Note $\hat{S}_x \equiv$ covariance matrix

$$\hat{\mu}_x = \frac{1}{n} \sum_{j : y_j = 1}^{n} x_j$$

$\hat{\mu}_x = \frac{1}{8} |0| = \tilde{0}$

$\bar{X} = X - \hat{\mu}_x$

$\bar{X} = X$

$$\hat{S} = \frac{1}{n} \sum^{n} \bar{x}_j \cdot x_j^T$$

$\hat{S}_x = \frac{1}{8} \underbrace{\begin{vmatrix} 3r & r & -r & -3r & 3s & 2s & s & -6s \end{vmatrix}}_{1\times 8} \underbrace{\begin{vmatrix} 3r \\ r \\ -r \\ -3r \\ 3s \\ 2s \\ s \\ -6s \end{vmatrix}}_{8\times 1}$

$\hat{S}_x = \frac{1}{8} \begin{vmatrix} 9rr' + rr' + rr' + 9rr' + 9ss' + 4ss' + ss' + 36ss' \end{vmatrix}$

$r \cdot s = \begin{vmatrix} -1 & 2 & -1 & 0 & 2 \end{vmatrix} \begin{vmatrix} 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{vmatrix}$

$r \cdot s = -1 + -2 + 1 + 2 = 0 \leftarrow$ Note r & s are orthogonal

$r \cdot r' = \underbrace{\begin{vmatrix} -1 & 2 & -1 & 0 & 2 \end{vmatrix}}_{1\times 5} \underbrace{\begin{vmatrix} -1 \\ 2 \\ -1 \\ 0 \\ 2 \end{vmatrix}}_{5\times 1}$

$r \cdot r' = 1 + 4 + 1 + 0 + 4$

$= 10$

$s \cdot s' = \underbrace{\begin{vmatrix} 1 & -1 & -1 & 1 & 1 \end{vmatrix}}_{1\times 5} \underbrace{\begin{vmatrix} 1 \\ -1 \\ -1 \\ 1 \\ 1 \end{vmatrix}}_{5\times 1}$

$s \cdot s' = \begin{vmatrix} 1+1+1+1+1 \end{vmatrix}$

$s \cdot s' = 5$

$$\hat{S}_x = \frac{1}{8}\left| 9rr' + rr' + rr' + 9rr' + 9ss' + 4ss' + ss' + 36ss' \right|$$

$$\hat{S}_x = \frac{1}{8}\left| 20rr' + 50ss' \right|$$

$$\hat{S}_x = \frac{20}{8}rr' + \frac{50}{8}ss'$$

Find μ_s and μ_r (eigenvectors)

$$\|r\| = \sqrt{(-1)^2 + (2)^2 + (-1)^2 + (2)^2}$$
$$= \sqrt{10}$$

$$\|s\| = \sqrt{5}$$

$$\mu_s = \frac{s}{\sqrt{5}}, \quad \mu_r = \frac{r}{\sqrt{10}}$$

Note $A = \sum_{i=1}^{d} \lambda_i \, \mu_i \mu_i^T$

$$\hat{S}_x = \lambda_r \left(\frac{r}{\sqrt{10}}\right)\left(\frac{r'}{\sqrt{10}}\right) + \lambda_s \left(\frac{s}{\sqrt{5}}\right)\left(\frac{s'}{\sqrt{5}}\right)$$

$$\hat{S}_x = \frac{\lambda_r rr'}{10} + \frac{\lambda_s ss'}{5}$$

$$\frac{20}{8}rr' + \frac{50}{8}ss' = \frac{\lambda_r rr'}{10} + \frac{\lambda_s ss'}{5}$$ ← ratios for transposed should equal

$$\frac{20}{8} = \frac{\lambda_r}{10}$$

$$\lambda_r = \frac{200}{8}$$

$$\boxed{\lambda_r = 25}$$

$$\frac{50}{8} = \frac{\lambda_s}{5}$$

$$\boxed{\lambda_s = \frac{250}{8}}$$

ii) The eigenvectors of $\hat{S}_x$ are the principal components

$$\boxed{\begin{aligned} y_{1,pca} &= \begin{vmatrix} \frac{-1}{\sqrt{10}} \\ \frac{2}{\sqrt{10}} \\ \frac{-2}{\sqrt{10}} \\ 0 \\ \frac{2}{\sqrt{10}} \end{vmatrix} \\ y_{2,pca} &= \begin{vmatrix} \frac{1}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{vmatrix} \end{aligned}}$$

Problem 5.2 [60pts] *(PCA of Images)* In this problem we will perform Principal Component Analysis (PCA) of two different image datasets: the AT&T Face Dataset and the MNIST Hand-Written Digits Dataset.

(a) [5pts] *(Data Visualization)* Use the provided `load_faces.m` function and load the face-dataset into your workspace. Compute the mean face image, that is the mean of all the 400 face image vectors. Use `subplot` and `imshow(uint8(reshape(img_vector, img_size)))` to create a figure containing two images: (1) image #120 in the dataset, and (2) the mean face of the dataset. Does the mean face resemble a human face or is it a smooth shapeless "blob"?

The mean face resembles a blob.

(b) [20pts] *(Eigenvalues of Covariance Matrix)* Subtract the vectorized mean face image $\hat{\boldsymbol{\mu}}_x$ from all the vectorized face images and compute the empirical covariance matrix $\hat{\mathrm{S}}_x$ of the mean-centered vectorized face images. Perform an Eigenvalue Decomposition of this covariance matrix using the inbuilt MATLAB fucntion `eig`. Arrange the eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots$ in the order of non-increasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$, into an orthonormal matrix $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots]$. Let Λ be the corresponding diagonal matrix of eigenvalues.

(1) [5pts] Report the values of the first five eigenvalues. (2) [5pts] Plot λ_k as a function of k, for $k = 1, 2, \ldots, 450$. Comment on the observed trends in your report. Explain why $\lambda_k = 0$ for all $k > 400$. (3) [5pts] Compute and plot (as a function of k) the values of the so-called "fraction of variance explained" by the top k principal components:

$$\rho_k := \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}.$$

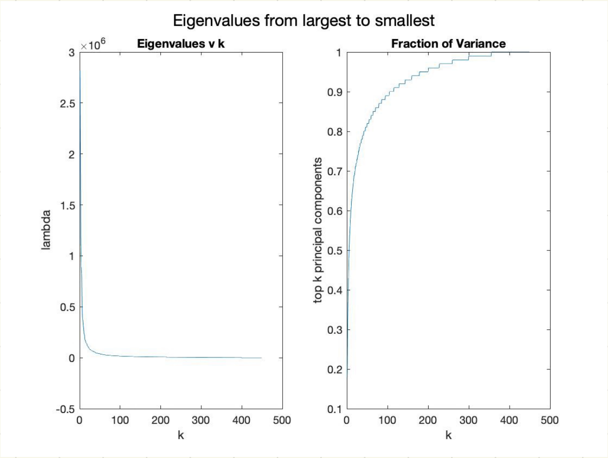
Round the values of ρ_k to 2 decimal places. Comment on the observed trends in your report. (4) [5pts] Find and report the smallest values of k for which $\rho_k \geq 0.51, 0.75, 0.90, 0.95,$ and 0.99.

Notes: Use subplots to show plots of λ_k and ρ_k in the same figure. The reshaped principal directions of a human face dataset are often referred to as eigenfaces.

1)

```
The first five eigenvalues are: [-7.73e-08 4.21e+05 4.33e+05 4.78e+05 4.93e+05 ]
```

2&3)



Eigenvalues v k:

log graph. Values are at a peak @ 0 and then quickly decrease to zero and stay there. Since rank($\bar{X}X^T$) = rank($\bar{X}^T\bar{X}$) then there are only 400 eigenvalues according to the n dimension matrix

Fraction of Variance explained:

log graph also but instead starts at low point and then hits high and stays there. High $\approx$ 1.

4)

```
The smallest values of k for which pk >= 0.51, 0.75, 0.90, 0.95 and 0.99 are: [6 29 105 179 300 ]
```

(c) [9pts] ☕ *(Image Approximation through Dimensionality Reduction)* The first eigenface corresponds to the largest eigenvalue and represents a direction that encompasses the largest variance in the training data, the second eigenface corresponds to the second largest eigenvalue and represents a direction which is orthogonal to the first eigenface and encompasses the second largest variance in the training data, and so on. A linear combination of eigenfaces along with a mean face, can be used to reconstruct a given face. Specifically, let **x** denote the vectorized representation of an image. An approximation to **x** based on the top k principal components is given by

$$\hat{\mathbf{x}} = \hat{\boldsymbol{\mu}}_x + \sum_{i=1}^{k} \hat{y}_{i,PCA} \cdot \mathbf{u}_i, \quad \text{where} \quad \hat{y}_{i,PCA} = \mathbf{u}_i^\top (\mathbf{x} - \hat{\boldsymbol{\mu}}_x)$$

[1pt per image] Generate a single figure (using subplots) showing the following: the mean face, the k principal component approximation face for image #43 in the face dataset for $k = 1$, $k = 2$, all the k values that you found in part (b) sub-part (4), and then show the original image in the same figure too. Add a descriptive title to each image in the figure and a title to the overall figure (see skeleton code for reference).

Skipped.

(d) [26pts] *(Connecting Principal Components with Image Features)* For this problem, load the MNIST data set. Filter the dataset using the label column to extract all images of the numeral '3' into one matrix (*Hint*: use conditional/logical based indexing into MATLAB matrices). We will use this filtered dataset of images of numeral '3' for this problem.

(1) [5pts] Compute the mean image of the numeral '3' and use subplot and `imshow(reshape(img_vector, img_size))` to report image #120 in the filtered dataset and the mean image of the filtered dataset in a single figure.

Percentile Values: the P-th percentile ($0 < P \leq 100$) of a list of ordered values (sorted from least to greatest) is the smallest value in the list such that no more than P percent of the data is strictly less than the value and at least P percent of the data is less than or equal to that value.

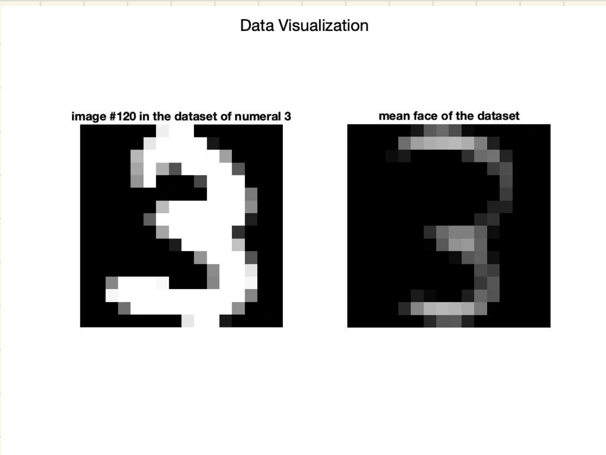
Compute the first two principal components for all images in the filtered dataset. Use the provided `percentile_values(v, percentiles)` function to compute the percentile values for the first principal components of all images and the second principal components of all images for the following percentiles: 5th, 25th, 50th, 75th, and 95th percentile.

Helpful functions for parts (2) and (3): `hold on, grid on, xticks, yticks, subplot`.

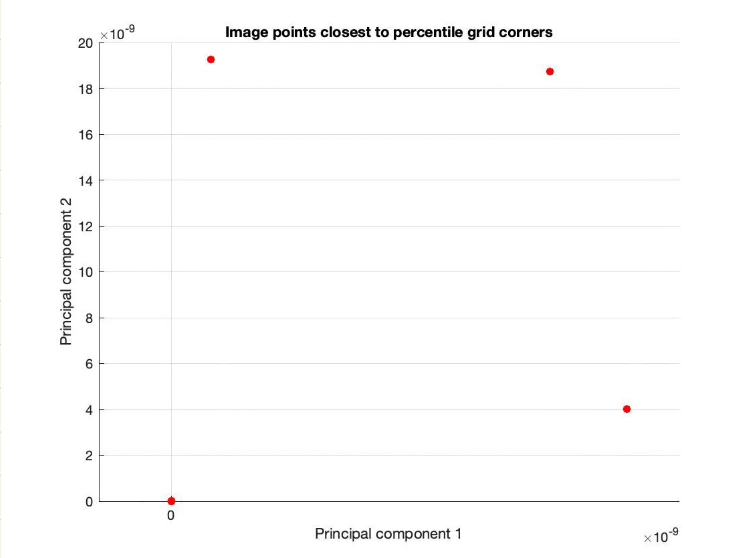
(2) [15pts] Generate a single figure containing the following plots. A scatter plot of the first two principal components of all images, i.e., image j will appear as a point with coordinates $(\hat{y}_{1j}, \hat{y}_{2j})$ in the scatter plot, where $\hat{y}_{1j}, \hat{y}_{2j}$ are, respectively, the first and second principal components of image j. Overlay a rectangular grid where the vertical grid lines occur at the percentile values of the first principal component and the horizontal grid lines occur at the percentile values of the second principal component. We will call this grid, the Percentiles Grid (PG). Find the intersection points of the grid lines in PG (*Hint:* Intersction points in PG are easily obtained via the Cartesian product of the percentile values of each principal component). Find the images whose first two principal components $(\hat{y}_{1j}, \hat{y}_{2j})$ are closest (in terms of Euclidean distance) to these intersection points and indicate their locations in the scatter plot using "filled" red dots (see documentation of `scatter`). } On plot

(3) [6pts] Display, in a single figure, the closest images from the filtered data that correspond to the red dots and discuss what image property (if any) of the numeral '3' does each principal component capture? (Be wary of the indices in this task).

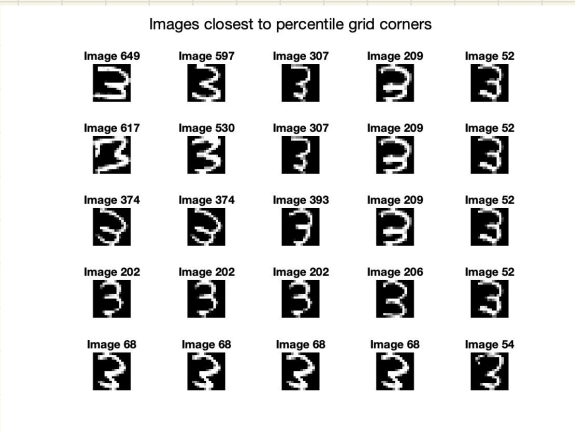
1)



2)



3)



These images all have a long "hood" for the number 3

"hood"