STATISTICAL PATTERN RECOGNITION (FALL 2021)

HOMEWORK#4

Non-Parametric Density Estimation ,PCA and Fisher LDA

Due date: 27th December 2021

In order to do this homework, you have to go through density estimations, principal component analysis, and Fisher Linear Discriminant Analysis theories and concepts.

Part A: Non-Parametric Density Estimation

For this part of homework, you have to generate a dataset for three classes, each with 500 samples from three Gaussian distributions described below:

$$\text{Class1:} \quad \mu = \begin{pmatrix} 2 \\ 5 \end{pmatrix} \quad \sum = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\text{Class2:} \quad \mu = \begin{pmatrix} 8 \\ 1 \end{pmatrix} \quad \sum = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$$

$$\text{Class3:} \quad \mu = \begin{pmatrix} 5 \\ 3 \end{pmatrix} \quad \sum = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Use your generated data and estimate the density without pre-assuming a model for the distribution, which is done by a non-parametric estimation.

- Implement the below PDF estimation methods using h=0.09,0.3,0.6 and answer the following questions for all of them:

1. Histogram
2. Parzen Window
3. Gaussian kernel (Standard Deviations of 0.2,0.6,0.9)
4. KNN Estimator (Fork=1,9,99)

- Estimate P(X) and Plot the true and estimated PDF, and compare them for each model.

- Find the best value for h in the Gaussian kernel model with the standard deviation of 0.6 using 5-Fold cross-validation.

Part B: PCA

Dataset: Jaffe

In this part, you will compute a PCA from a set of images of faces. The database provides facial images of different people. For simplicity, convert the images into gray scale and resize them into 64*64. Attention that eigenfaces are sets of eigenvectors that can be used to work with face recognition applications. Each eigenface, as we will see in a bit, appears as an array vector of pixel intensities. We can use PCA to determine which eigenfaces contribute the largest variance in our data and eliminate those that do not contribute much. This process lets us determine how many dimensions are necessary to recognize a face as 'familiar.'

- Visualize the dataset.
- Preprocess and normalize the dataset.
- Implement the PCA function, then apply it to the dataset.
- Visualize the reduced dataset using 2D and 3D plots.
- Reconstruct the original data using K principle components (Show reconstructed images of each individual for K=1,40,120).
- Plot the MSE between the original and reconstructed images in terms of the number of eigenvectors.
- Visualize some of the first principal components.
- How many principal components are enough so that you have acceptable reconstruction? How do you select them?

Part C: Fisher LDA

Dataset: Jaffe

In this part, you will apply a Fisher LDA from a set of images of faces.

- Implement and apply the Fisher LDA function for a multi-class problem.
- What is the problem of applying Fisher LDA to the dataset?
- Reconstruct the original data by using K basis vectors obtained from LDA. (Show reconstructed images of one person for k=1, 6, 29).
- What would happen if we had a large number of outliers in the dataset?
- Plot the MSE between the original and reconstructed images in terms of the number of eigenvectors.

Notes:

- ✓ Pay extra attention to the due date. It will not extend.
- ✓ Be advised that submissions after the deadline would not grade.
- ✓ Prepare your full report in PDF format and include the figures and results.
- ✓ Do not use sklearn or any similar library and write your own code.
- ✓ Submit your assignment using a zipped file with the name of "StdNum_FirstName_LastName.zip"
- ✓ Feel free to use your preferred programming languages.
- ✓ Using other students' codes or the codes available on the internet will lead to zero