

Statistical Machine Learning

Assignment: 2

Max marks: 100

Instructions:

1. You are free to use math libraries like Numpy, Pandas; and use Matplotlib, Seaborn library for plotting.
 2. Add all the analysis related to the question in the written format, anything not in the report will not be marked.
 3. Implement code that is Modular in nature and generalized to be executed for any input.
 4. You can't use the inbuilt libraries to implement any algorithm except for calculation purposes like variance, Eigen decomposition, SVD etc.
 5. Code should be submitted in Python/Matlab file format only(.py/.m)
-

DIMENSIONALITY REDUCTION

1. Use the MNIST digit recognition dataset having 10 classes for the purpose of this assignment.

Dataset ([MNIST](#))

Perform the following tasks:

- a. Compute the global mean and covariance of the data.
- b. Implement PCA and FDA from scratch.
- c. Visualize data using a scatter plot after applying PCA & FDA. (You can transform the data into 2 dimensions and then plot it.)
- d. Implement the LDA discriminant function from scratch.
- e. Apply PCA with 95% eigen energy on MNIST and then LDA for classification and report the accuracy on test data.
- f. Visualize and analyze the eigenvectors obtained using PCA (only for eigenvectors obtained in part(e). I.e., Display eigenvectors by converting them into image form).
- g. Perform step(e) with different eigen energy mentioned below and show the comparisons and analysis on accuracy.

- 70% eigen energy
 - 90% eigen energy
 - 99% eigen energy
- h. Apply FDA on MNIST and then LDA for classification and report the accuracy on test data.
 - i. Perform PCA then FDA. Classify the transformed datasets using LDA. Analyze the results on Accuracy.

NOISE REDUCTION

2. Perform the following steps on the given dataset:
 - a. Add Gaussian noise to the dataset. (**NOTE:** You can take mean=0 and variance can be varied upon your choice such that the noise reduction can be seen clearly from the image.)
 - b. Perform PCA on the noisy dataset for Noise Reduction.
 - c. Visualize the dataset before & after noise reduction. (Report the images as shown below. Linear PCA in the below image refers to normal PCA only.).
 - d. Report the number of components for which PCA works the best in Noise Reduction.

Original data



Data corrupted with Gaussian noise



Result after linear PCA

