REPORT - ASSIGNMENT 3
STATISTICAL MACHINE LEARNING

VAIBHAV VARSHNEY (MT17065)

Dataset information

- Here, **dataset 1** is Face dataset which consist of images of 11 types. There is total of 704 images. Each having dimensions of 192x168. While **dataset 2** is CIFAR-10 dataset which consist of 60000 images each having dimensions 32x32x3.
- In case of dataset 1, each image is resized into 50x50.
- PCA works on the principle of choosing the directions in which the variables have the most spread, not the dimensions that have the most relative distances between clustered subclasses.
- LDA tries to minimise the within class variance and maximise the between class variance.

Dataset 1:

(1) LDA : On performing LDA projection on the dataset, following metrics were obtained.

K-fold mean accuracy	0.776
K-fold standard deviation	0.030
Test set accuracy	0.727

(2) PCA : On performing PCA projection, a total of **121** projections were obtained. The following table shows the stats:

K-fold mean accuracy	0.958
K-fold standard deviation	0.015
Test set accuracy	0.967

(3) LDA after PCA : 0.886 120 eigenvectors

(4) PCA after LDA : 0.594 7eigenvectors

Observations and inferences

- In case of **PCA,** while projecting the dataset in that eigenvector subspace. Those eigenvectors are retained in the 99% eigenenergy (121) which best describes the features for dataset.
- In case of **LDA,** the dataset is projected in c-1 eigenvector subspace.
- Here in dataset 1, the **PCA is performing better than LDA** because in case of PCA, it is trying to find best distinction of the images. While **LDA** is trying to maximize the distinction between the distinct classes. In this case of dataset 1, PCA is able to find better features as compared to LDA.
- **In case of LDA after PCA,** pca performs feature extraction. In this particular case, initially applying PCA, may be extracting the best features. Then applying LDA is distincting and providing good performance as compared to PCA after LDA.
- In **PCA after LDA,** we are extracting distinct eigenvectors for the given classes. Further applying PCA is further refining the feature space reducing performance of the model.

Graphical Observations

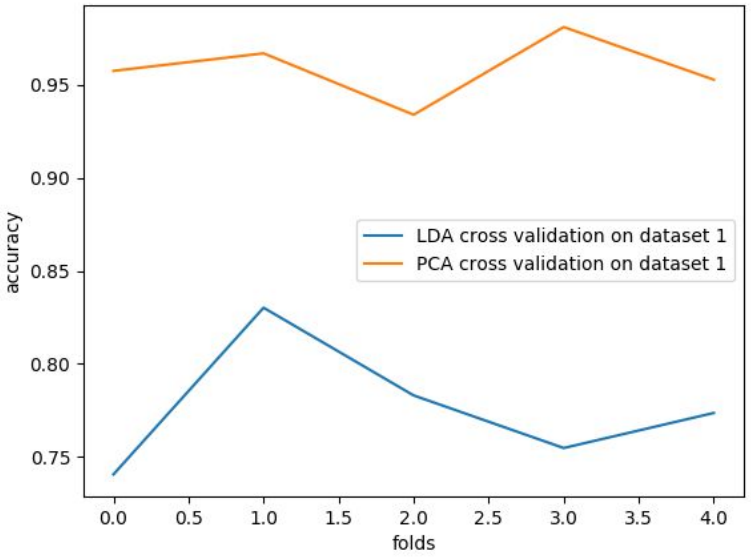


Fig : cross-validation of LDA and PCA

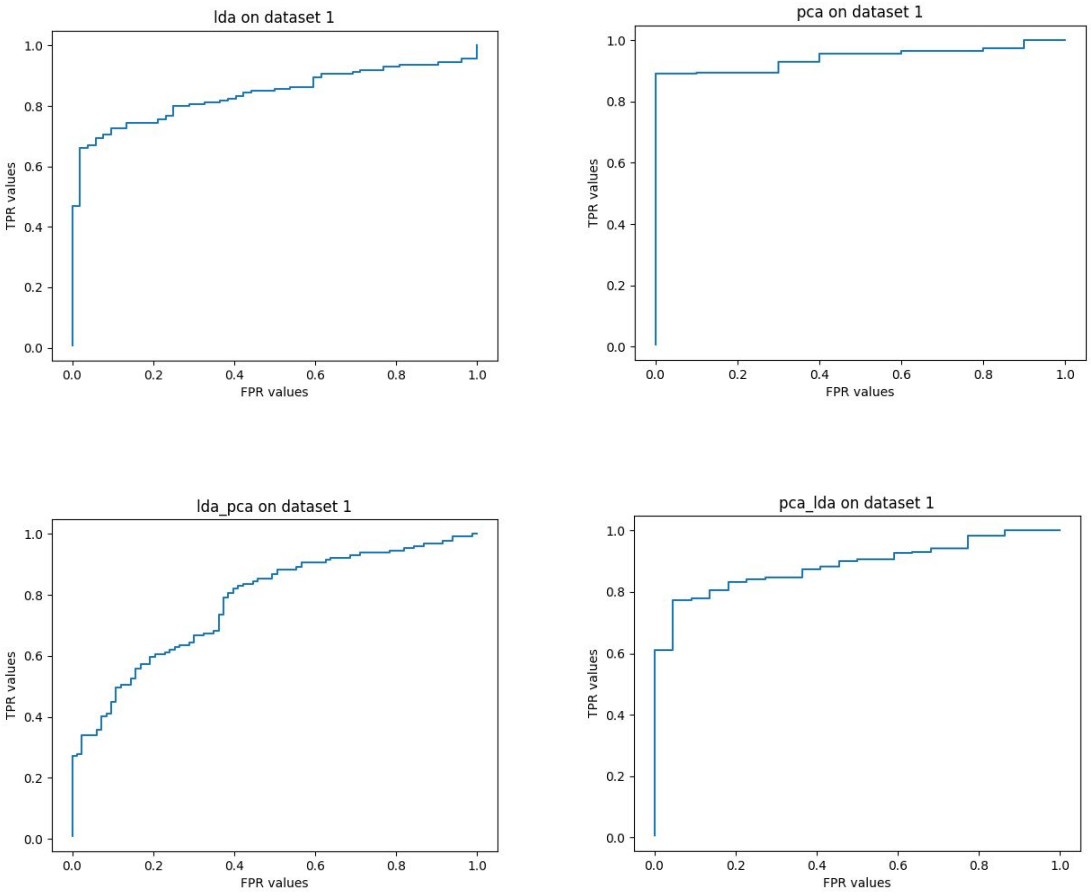


Fig : ROC curves

Dataset 2:

This is CIFAR-10 dataset. It consists of 10 classes. Each class has 6000 images. All the images after reading, have been dumped in files - 'data.pkl' and 'label.pkl'.

(1) LDA : On performing LDA projection on the dataset, following metrics were obtained.

K-fold mean accuracy	0.327
K-fold standard deviation	0.002
Test set accuracy	0.332

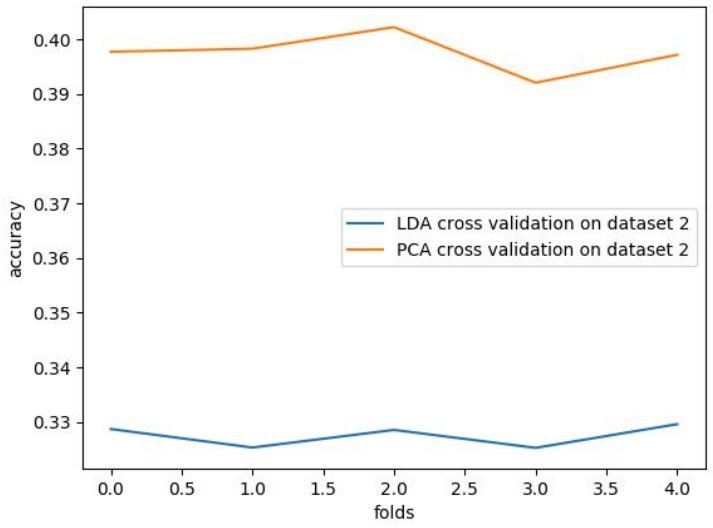
(2) PCA : On performing PCA projection, a total of **657** projections were obtained. The following table shows the stats:

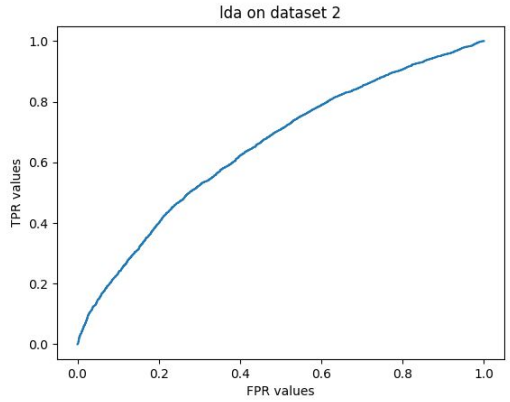
K-fold mean accuracy	0.397
K-fold standard deviation	0.003
Test set accuracy	0.406

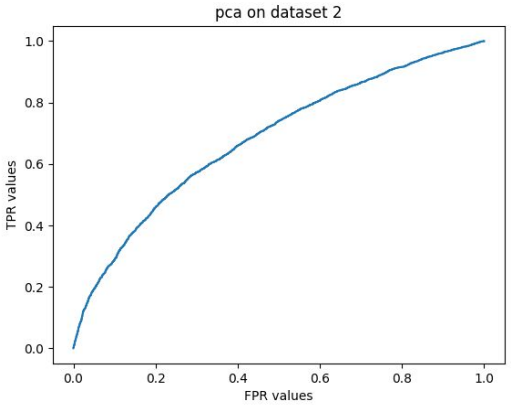
(3) LDA after PCA : accuracy = 0.358

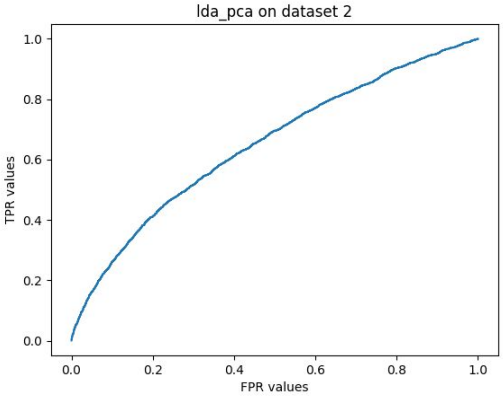
(4) PCA after LDA : accuracy = 0.299

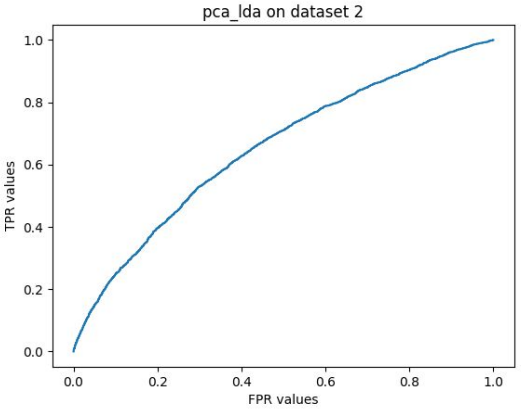
Graphical Observations











Observation and Inferences

- In case of dataset 2, images are of low quality. The background images in this dataset of each image are contributing enough noisy data. Hence, PCA may not be able to find the best describing projection space providing better feature extraction.
- In case of LDA, it is assumed that normal distribution is followed. But seeing the results, we can state that that dataset may not be following this assumption. Hence giving low accuracies and results for LDA.
- For the case of PCA after LDA and LDA after PCA, nothing can be stated (also considering the observation results of cross-validation and test accuracies).