BNC: Homework 3

Zayid Oyelami

July 25, 2019

1 Problem 1: PCA vs LDA

1.1 Background

Principal Component Analysis and Linear Discriminant Analysis are both techniques useful for linear feature extraction. PCA relies on a signal representation criterion whereby the objective is to be able to represent samples in a lower dimension. In contrast, LDA relies on signal classification criterion whereby the objective is to increase the discrimination between classes.

1.2 Method

1.2.1 PCA

PCA was done by subtracting the mean of the data set. Then, finding the co-variance matrix of the mean-subtracted dat set to further compute the eigenvectors of that best explain the variance in the data to best reconstruct the data set.

1.2.2 LDA

LDA was done by finding the mean and the co-variance of each class in the data set. Then, the scatter with-in class for each class and the scatter between the classes were calculated. Finally, the Fisher criterion was calculated with the scatter between classes and then within class.

1.3 Results

1.3.1 Data Set I: Linearly Separable - PCA

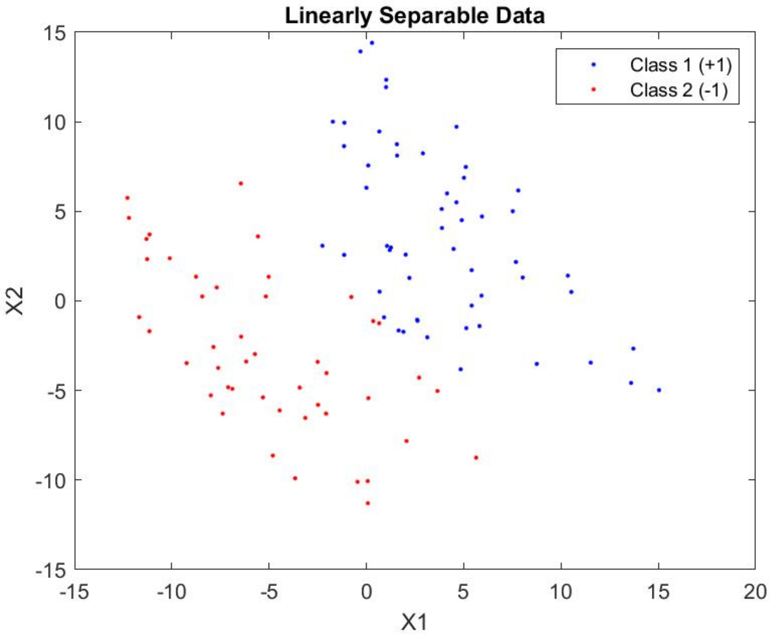

Figure 1: Linearly separable data.

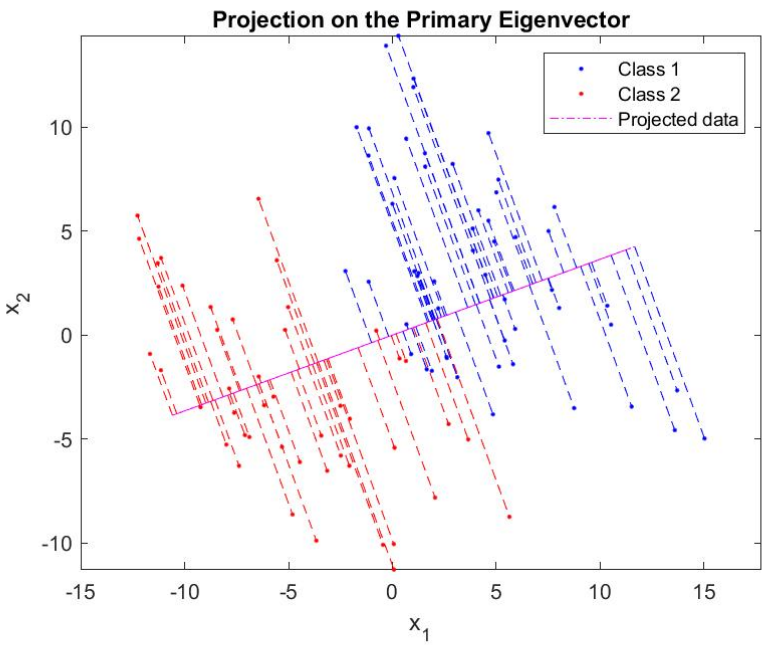

Figure 2: The projection of primary principle component on the separable data set.

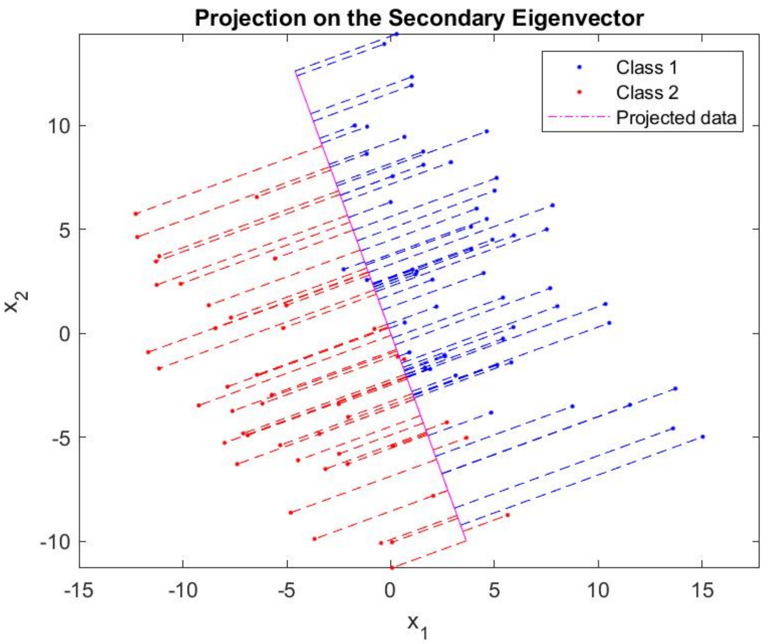

Figure 3: The projection of secondary principle component on the separable data set.

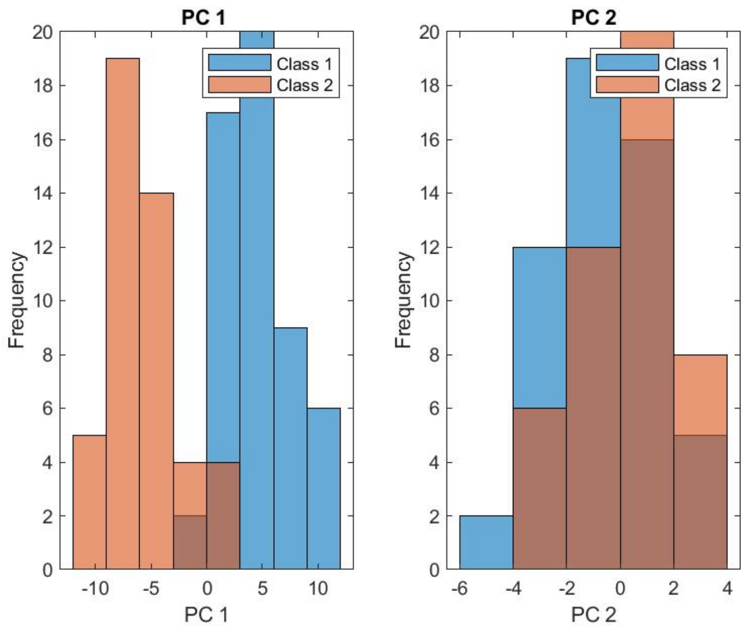

Figure 4: The histogram of both principle components' projected values on the separable data set.

1.3.2 Data Set I: Linearly Separable - LDA

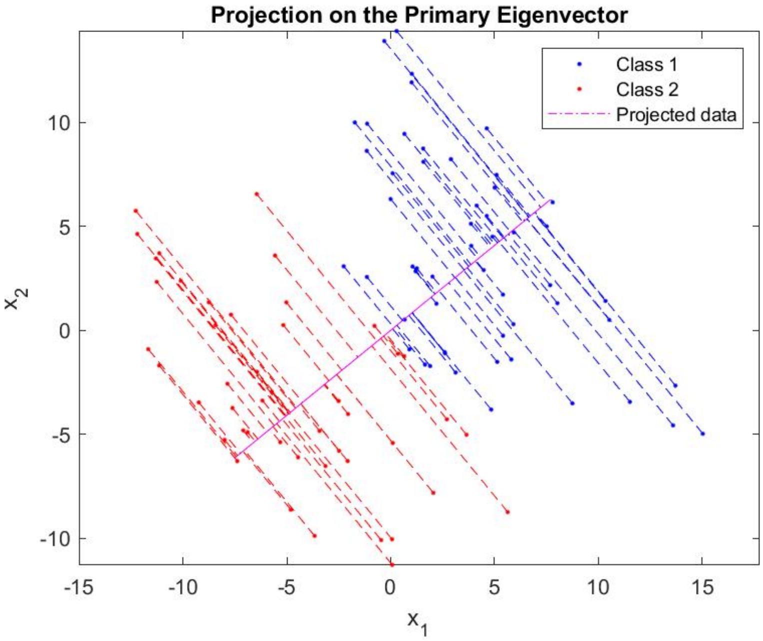

Figure 5: The projection of primary principle component on the separable data set.

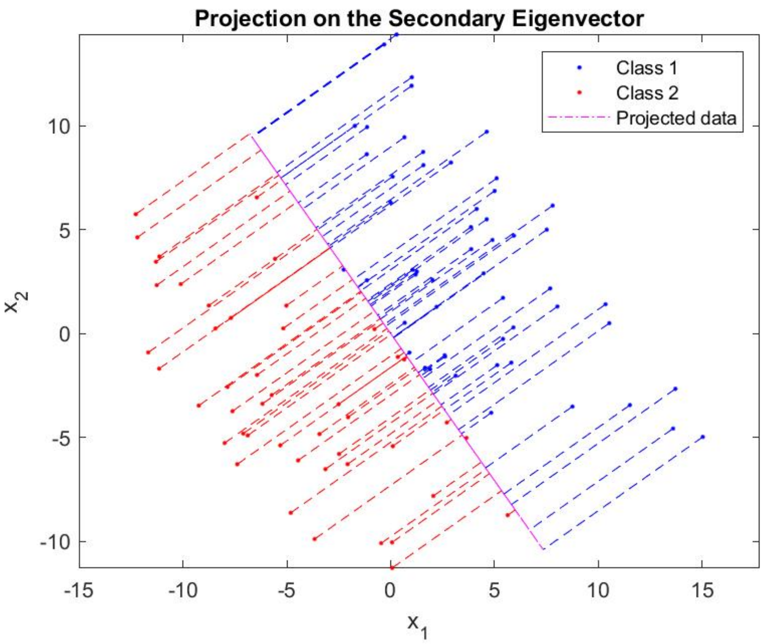

Figure 6: The projection of secondary principle component on the separable data set.

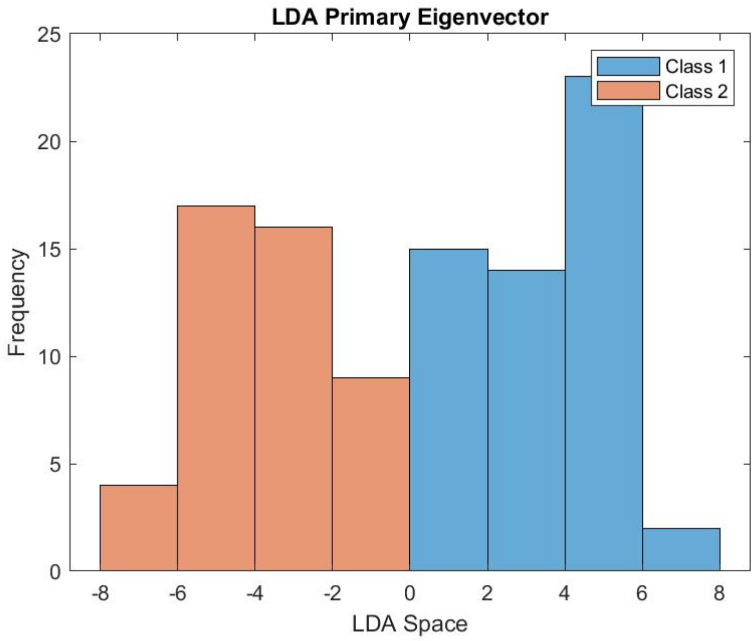

Figure 7: The histogram of the primary principle component's projected values on the separable data set.

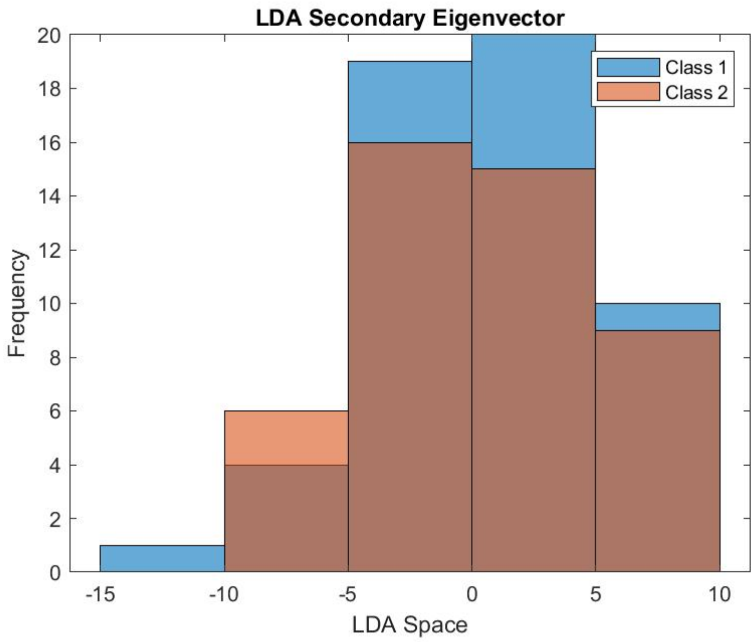

Figure 8: The histogram of the secondary principle component's projected values on the separable data set.

1.3.3 Data Set II: Non-linear & Overlapping - PCA

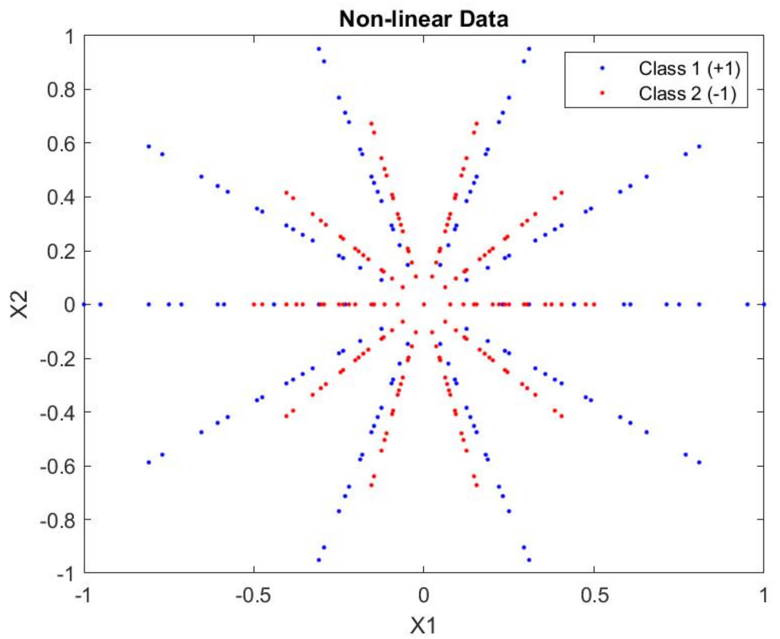

Figure 9: Non-linear data.

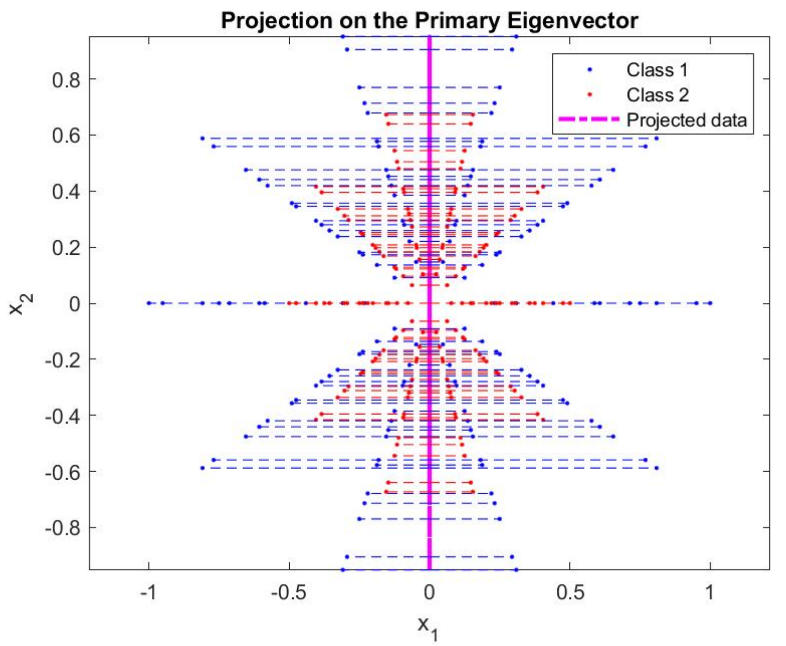

Figure 10: The projection of primary principle component on the non-linear data set.

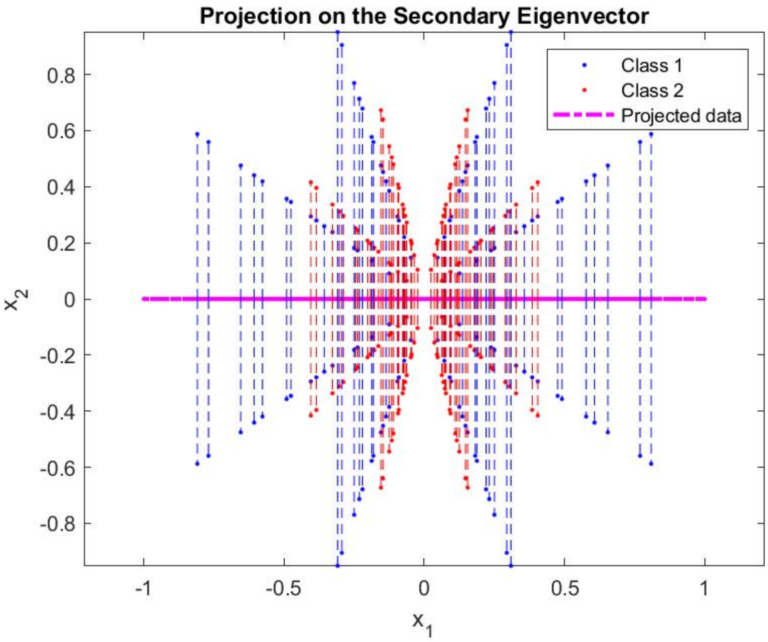

Figure 11: The projection of secondary principle component on the non-linear data set.

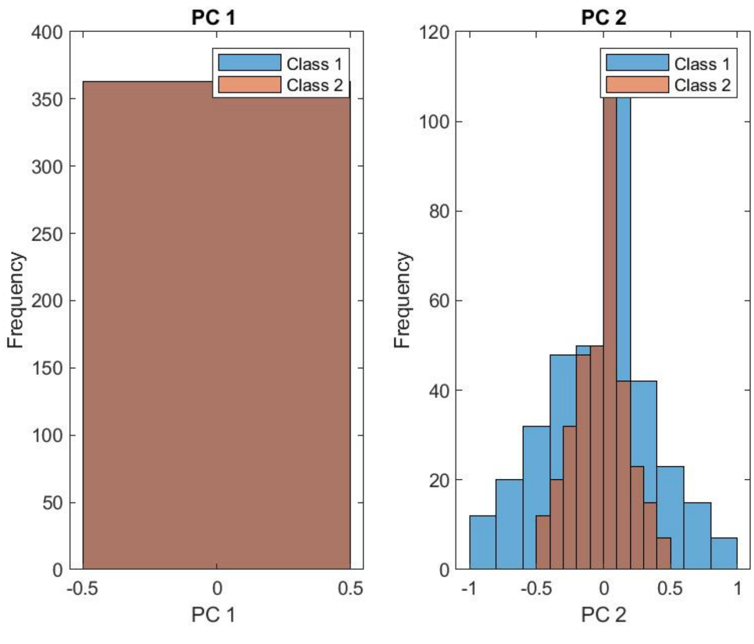

Figure 12: The histogram of both principle components' projected values on the non-linear data set.

1.3.4 Data Set II: Non-linear & Overlapping - LDA

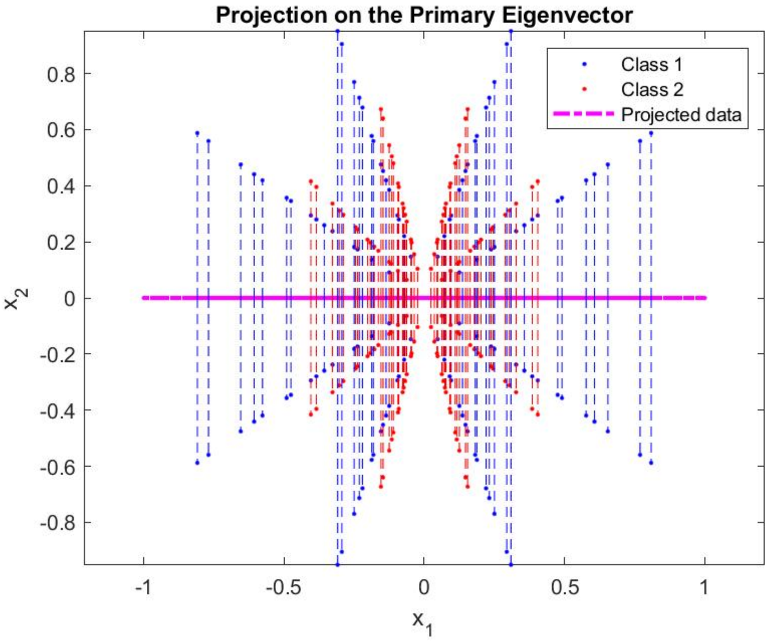

Figure 13: The projection of primary principle component on the non-linear data set.

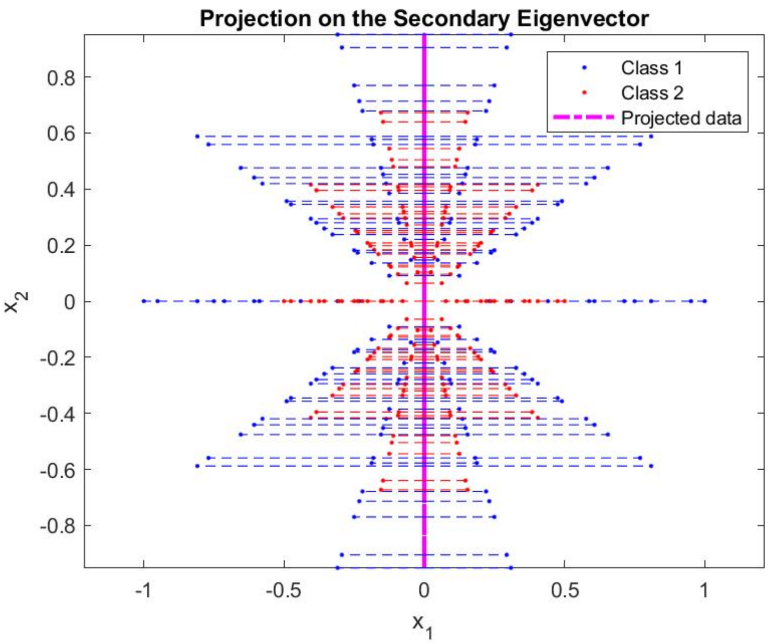

Figure 14: The projection of secondary principle component on the non-linear data set.

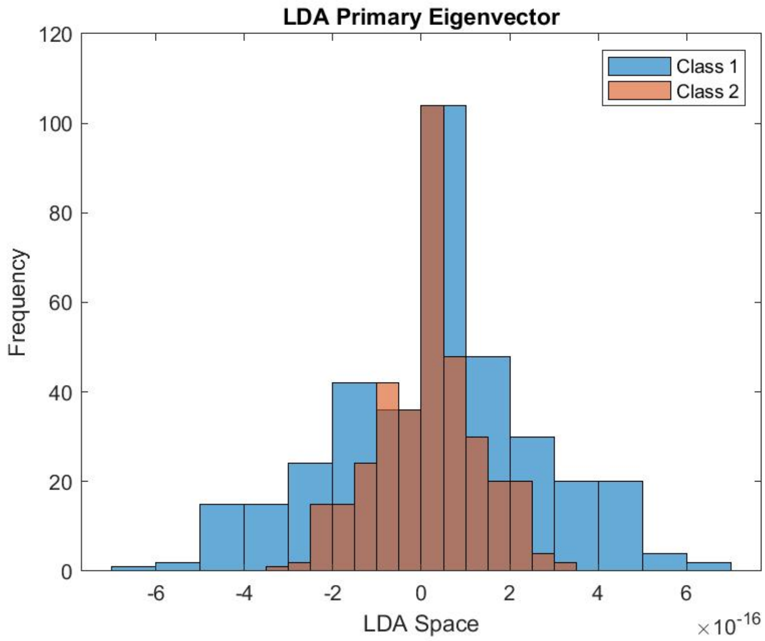

Figure 15: The histogram of the primary principle component's projected values on the non-linear data set.

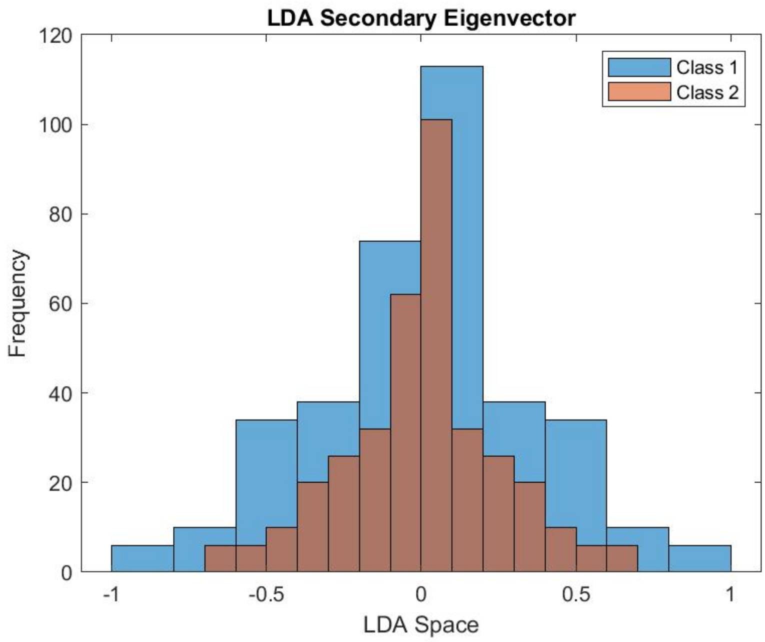

Figure 16: The histogram of the secondary principle component's projected values on the non-linear data set.

1.4 Discussions

The overall objective of both PCA and LDA is to optimize a feature extraction mapping for dimensionality reduction. As stated above, there a two notable criteria to accomplish said goal: signal representation (PCA) and classification (LDA).

On the linearly separable data set, LDA does better than PCA as is to be expected. LDA maximizes the amount of separation between classes. PCA serves more as a representation of the signal. It so happens that the data worked in the favor of the largest eigenvector of the PCA and aligned properly enough to create such a distinction in classes. The contrast of this behavior is When the projection demonstrated on PC2, because it is not aligned in such a manner that separates the classes. This is good example of when PCA would perform well due to the direction of greatest variance also aligning with good classification.

On the non-linear and slightly overlapping data set, both PCA and LDA do not work very well. Particularly for the PC1 of PCA, everything is seen as overlapping. This is a time where the greatest variance does not serve as a good represenation of the data. The PC 2 direction of PCA, however, does better at representation despite it being in the direction of lower variance. In the other figures, the distributions of the projected values are appropriately placed, however, this does not really distinguish between classes. Another important feature to note for LDA is that the distributions have similar means. This is critical as LDA fails when the means are close and variance serves as the discriminatory feature. This applies to both eigenvectors of LDA. Despite there being slight overlapping, the non-separable property of the data set leads to failures of both PCA and LDA.

2 Problem 2: ICA

2.1 Background

Independent Component Analysis (ICA) is a method that assumes that a source can be decomposed into linear components when both mixture signals and source signals are zero-mean as well as the source following a non-Gaussian distribution and the mixing matrix being square.

2.2 Method

Used the code provided by Dr.Raman to test several examples of the application of ICA on the MNIST data set. Two source signals – digits – were mixed accordingly then processed through the ICA algorithm in attempt to separate the mixed signals. The ICA algorithm works by centering and whitening the data, then determining the independent components while attempting to maximize kurtosis. During this process the outputs are uncorrelated via Gram-Schmidt method. To gain a better understanding of when ICA would succeed and fail, the digit 9 was compared against digits that looked similar to it such as digits 0 and 6 and digits that looked dissimilar to digit 9 such as 2 and 7.

2.3 Results

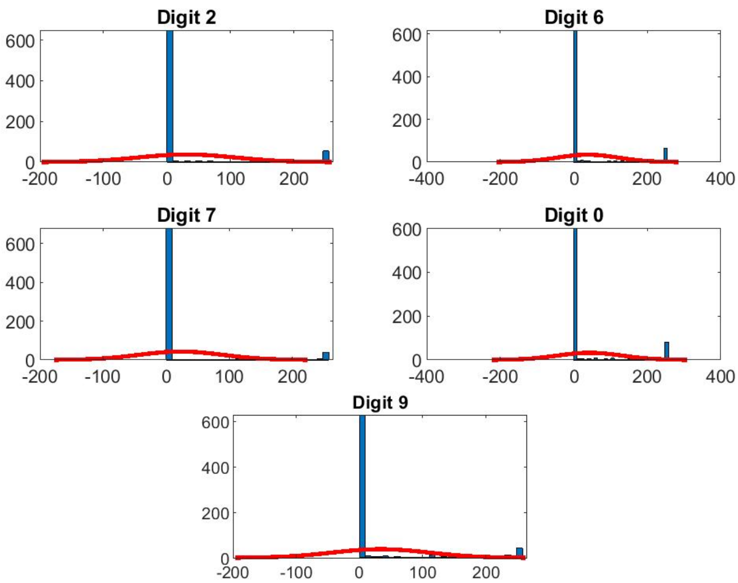

Figure 17: Histograms of each MNIST digit used.

2.3.1 Successful Examples

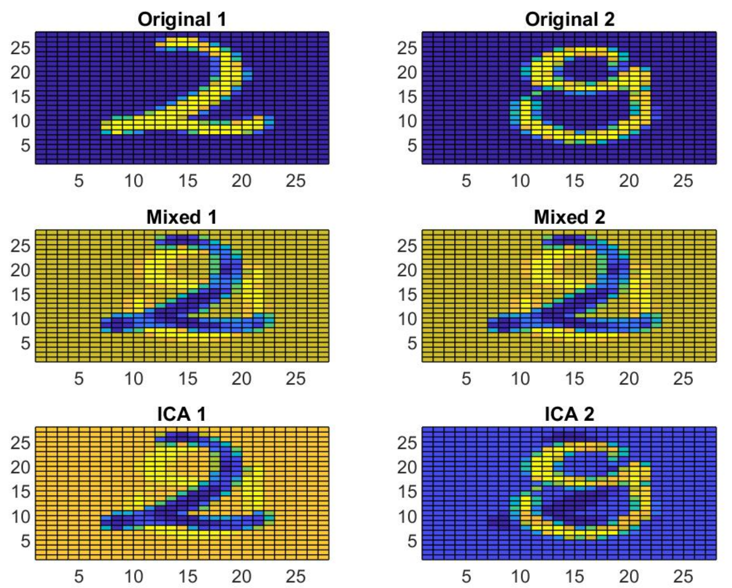

Figure 18: Applying ICA to digits 2 & 9 from the MNIST data set.

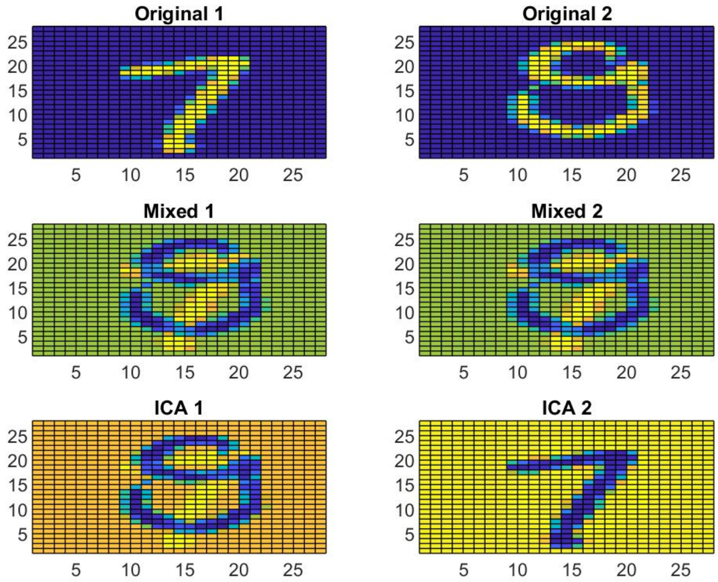

Figure 19: Applying ICA to digits 7 & 9 from the MNIST data set.

2.3.2 Failed Examples

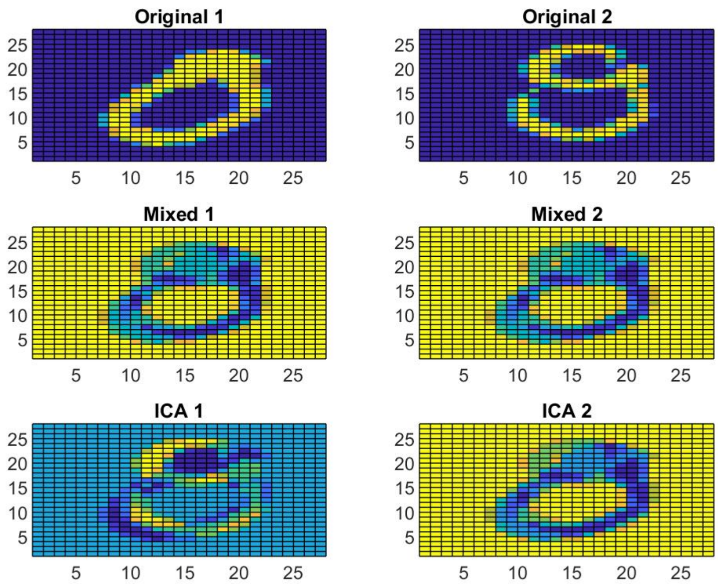

Figure 20: Applying ICA to digits 0 & 9 from the MNIST data set.

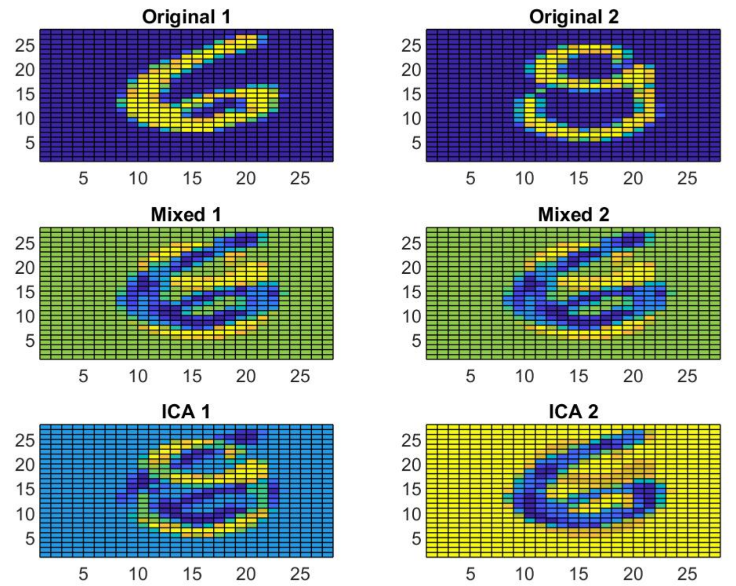

Figure 21: Applying ICA to digits 6 & 9 from the MNIST data set.

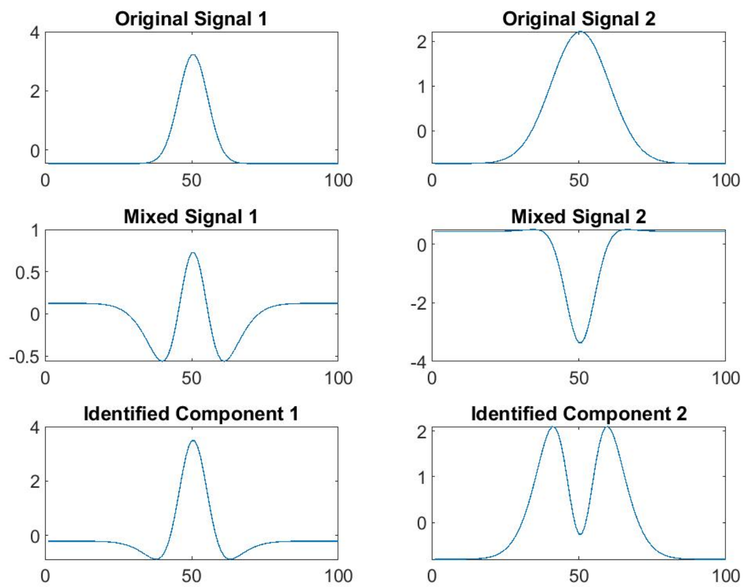

Figure 22: Gaussian source signal with zero-mean and unit standard deviation and another Gaussian source signal with zero-mean and standard deviation of 2.

2.4 Discussions

A critical limitation of ICA is its requirement for sources to be non-Gaussian due to the symmetries that arise from a such as distribution as well as its inability to determine variances of the sources: This can be seen in Fig. 22 with Gaussian source signals and different variances. ICA does overall well on the MNIST data set, particularly for numbers distinct from the digit 9. This could be due to the fact each of the chosen digits do not follow a Gaussian distribution and have a unit variance and zero-mean. However, ICA stil has trouble with numbers that look similar. This could be potentially attributed to the idea of these digits being somewhat dependent upon each other when there is overlap, so ICA has a hard time distinguishing between the two.