# SML Assignment 2 Report


## Dhruv Jaiswal - 2023200


## Accuracy of LDA

| | Class 0 | Class 1 | Class 2 | Overall Accuracy |
|---|---|---|---|---|
| FDA then LDA (train) | 0.94 | 0.94 | 0.91 | 0.93 |
| FDA then LDA (test) | 0.98 | 0.98 | 0.93 | 0.963 |
| PCA then LDA (train) | 1 | 0.99 | 0.98 | 0.99 |
| PCA then LDA (test) | 0.99 | 0.99 | 0.9 | 0.96 |
| PCA(90%) then LDA (train) | 1 | 0.99 | 0.97 | 0.986 |
| PCA(90%) then LDA (test) | 1 | 0.98 | 0.91 | 0.963 |
| PCA(2D) then LDA (train) | 0.87 | 0.85 | 0.74 | 0.82 |
| PCA(2D) then LDA (test) | 0.91 | 0.98 | 0.91 | 0.93 |

## Accuracy of QDA

| | Class 0 | Class 1 | Class 2 | Overall Accuracy |
|---|---|---|---|---|
| MLE (QDA, train) | 1 | 0.99 | 0.99 | 0.993 |
| FDA then QDA (train) | 0.95 | 0.93 | 0.9 | 0.926 |
| FDA then QDA (test) | 0.98 | 0.98 | 0.93 | 0.963 |

| | | | | |
|---|---|---|---|---|
| PCA then QDA (train) | 1 | 1 | 1 | 1 |
| PCA then QDA (test) | 0.99 | 0.74 | 0.99 | 0.906 |

## Effects of PCA

As seen in the accuracy data above, There are 3 kinds of PCA we apply

1. PCA (95% variance, approx. 82 dimensions)

   This transformation when applied on the dataset, keeps the accuracy relatively high on both the train set and test set however there is an unusual drop-off in some particular class of the trainset [Class 2 in LDA and Class 1 in QDA]. This might be because of discarding crucial dimensions needed to identify that class.

2. PCA (90% variance, approx. 50 dimensions)

   This transformation performs marginally worse than the 95% PCA but is more consistent as there is no one class that lags behind. This might be because the noise that might’ve come inside the data is removed completely here.

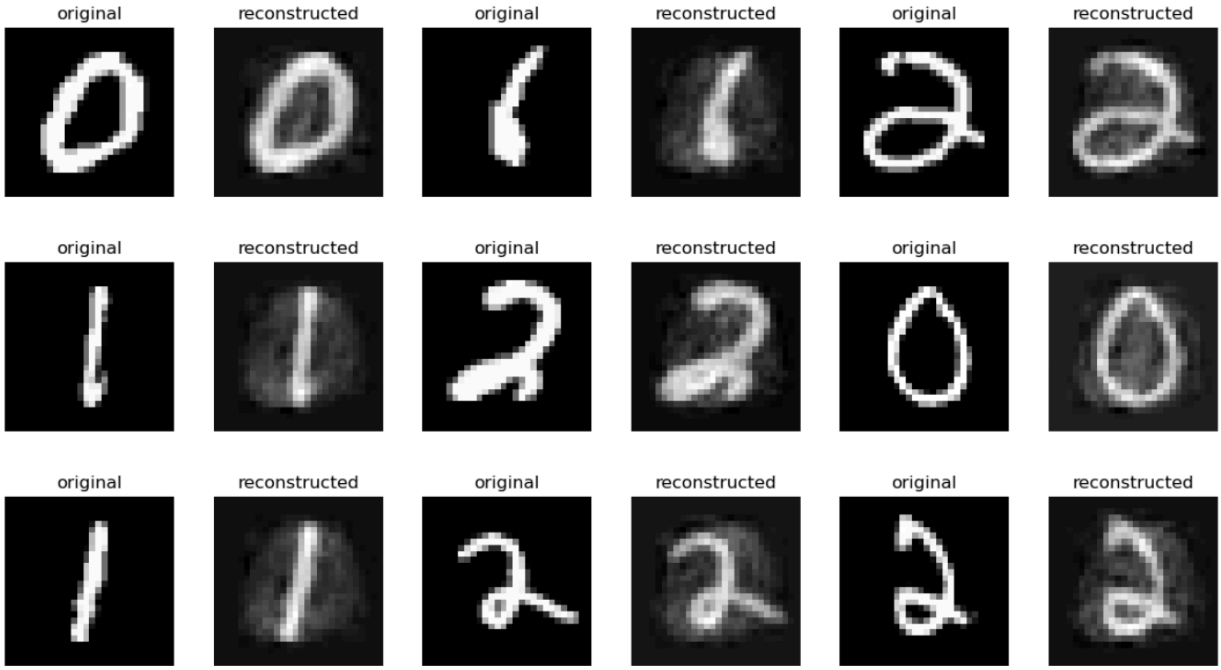
3. PCA (2 dimensions)

   This transformation gives bad results on the train set but mediocre (in comparison to other).

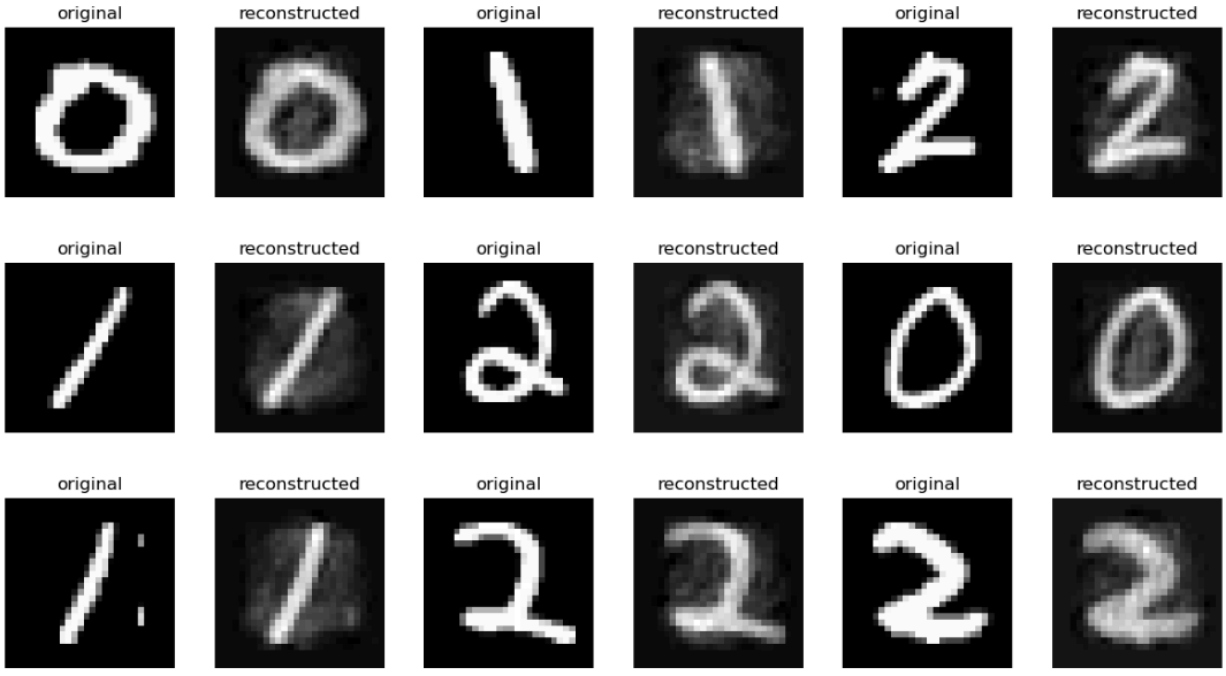
## Visualization

PCA [95%]

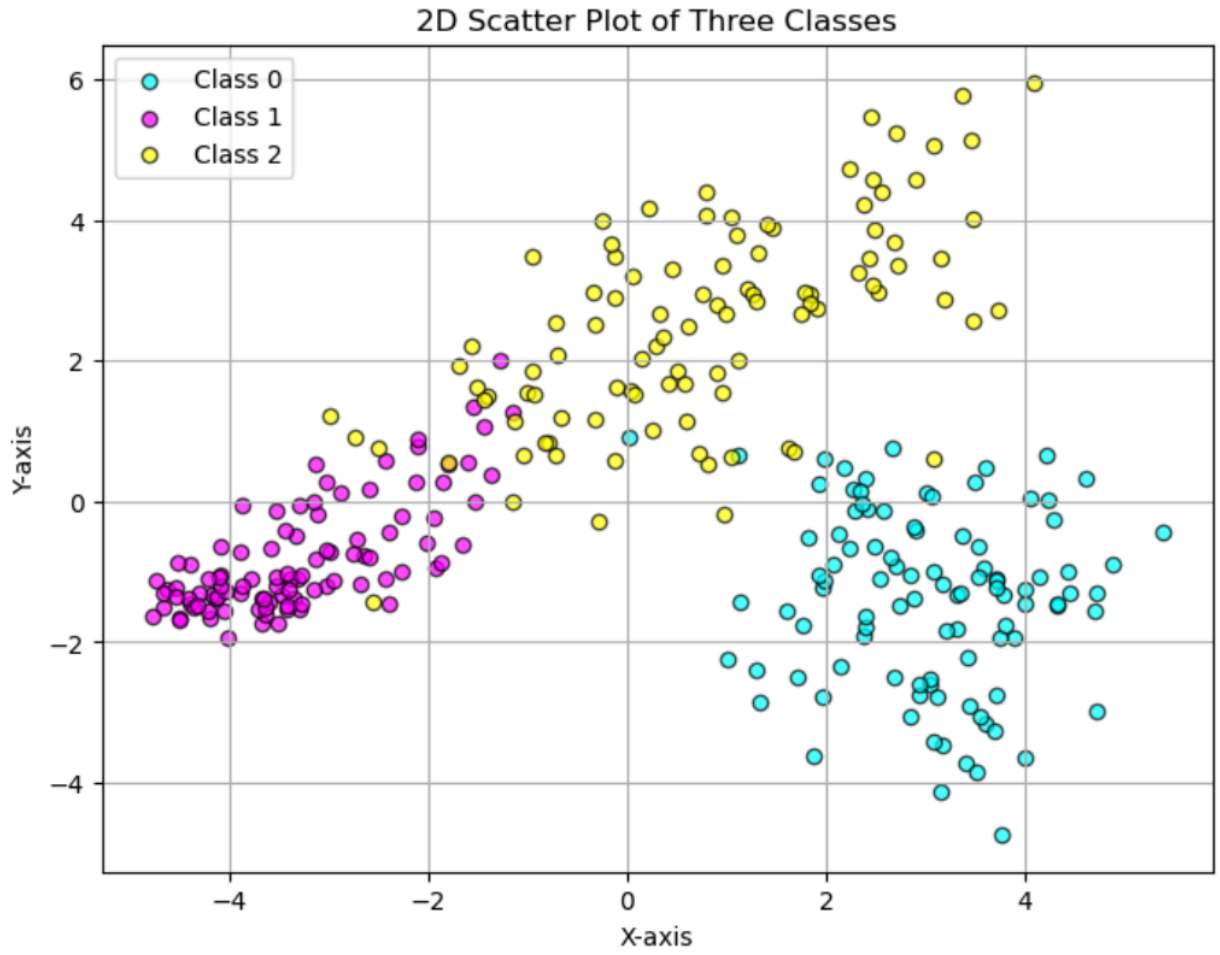
p = 82

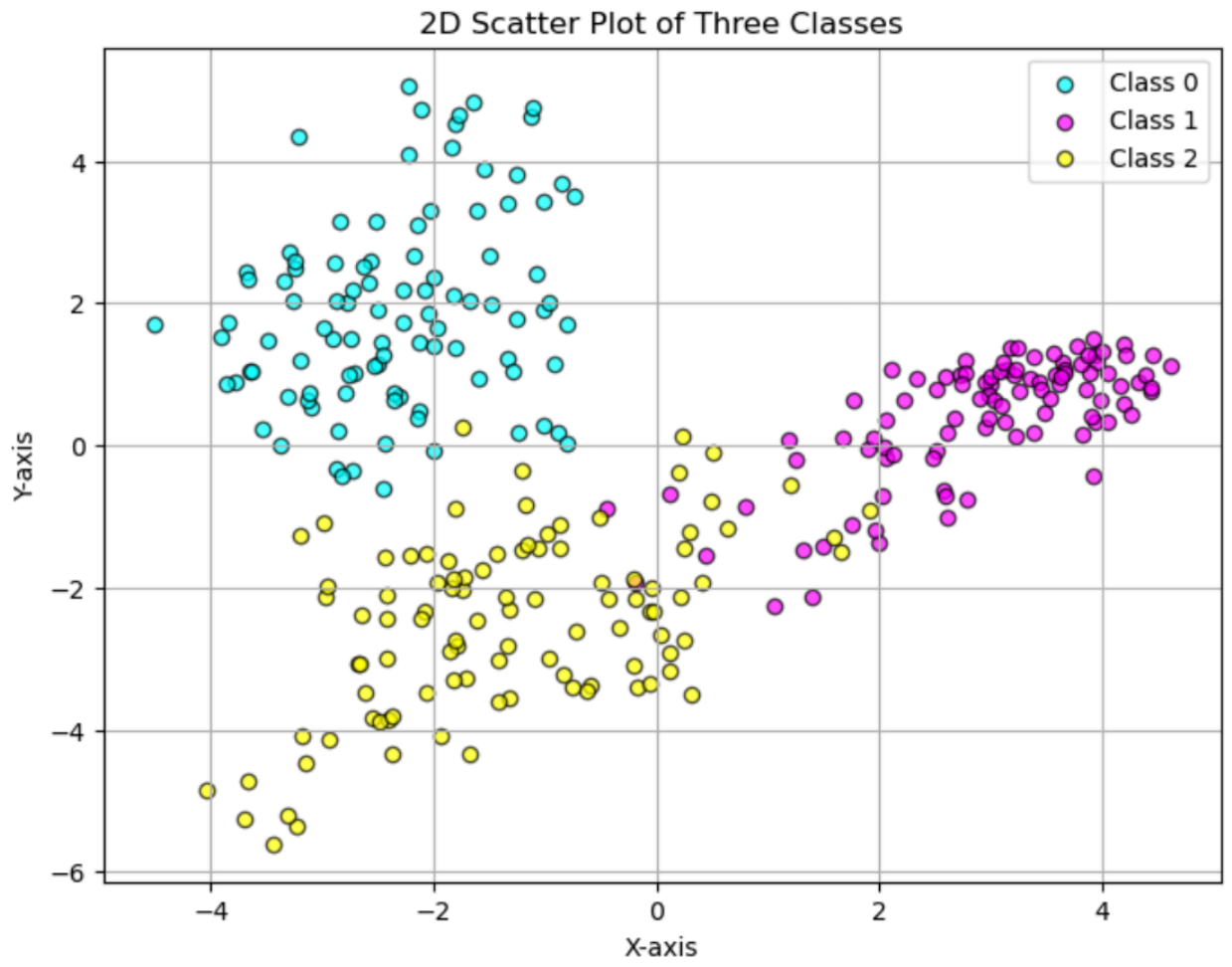

original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed

p = 81

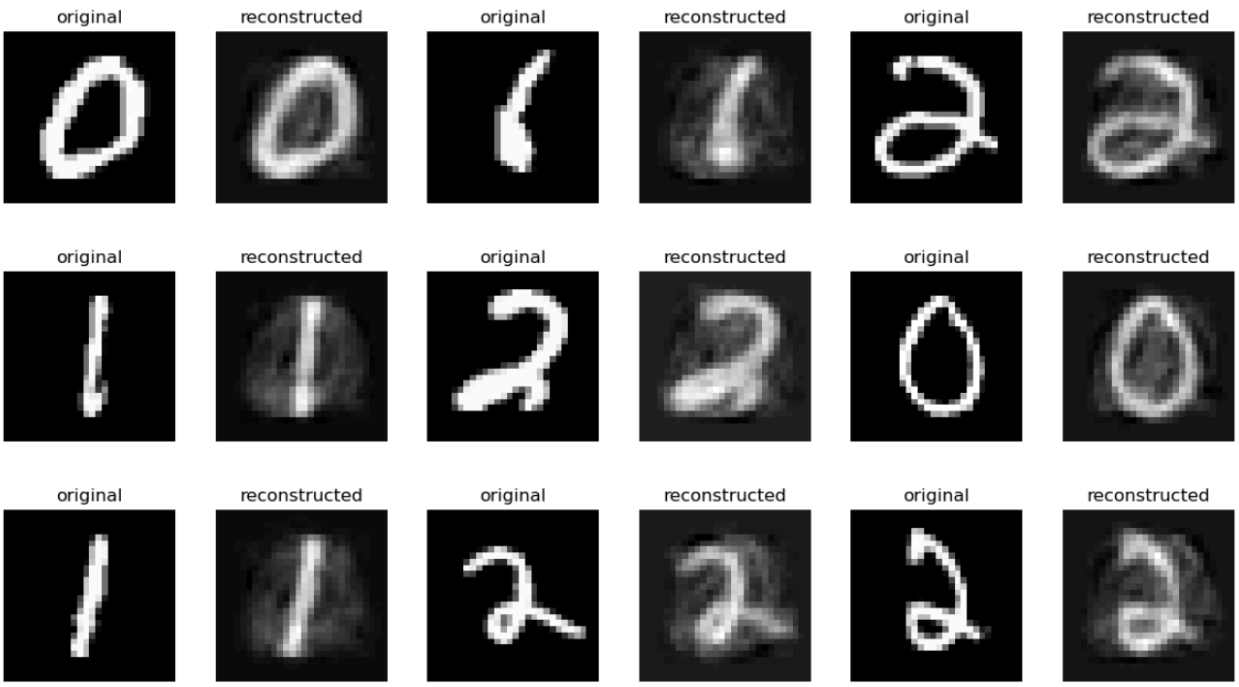

original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed

FDA

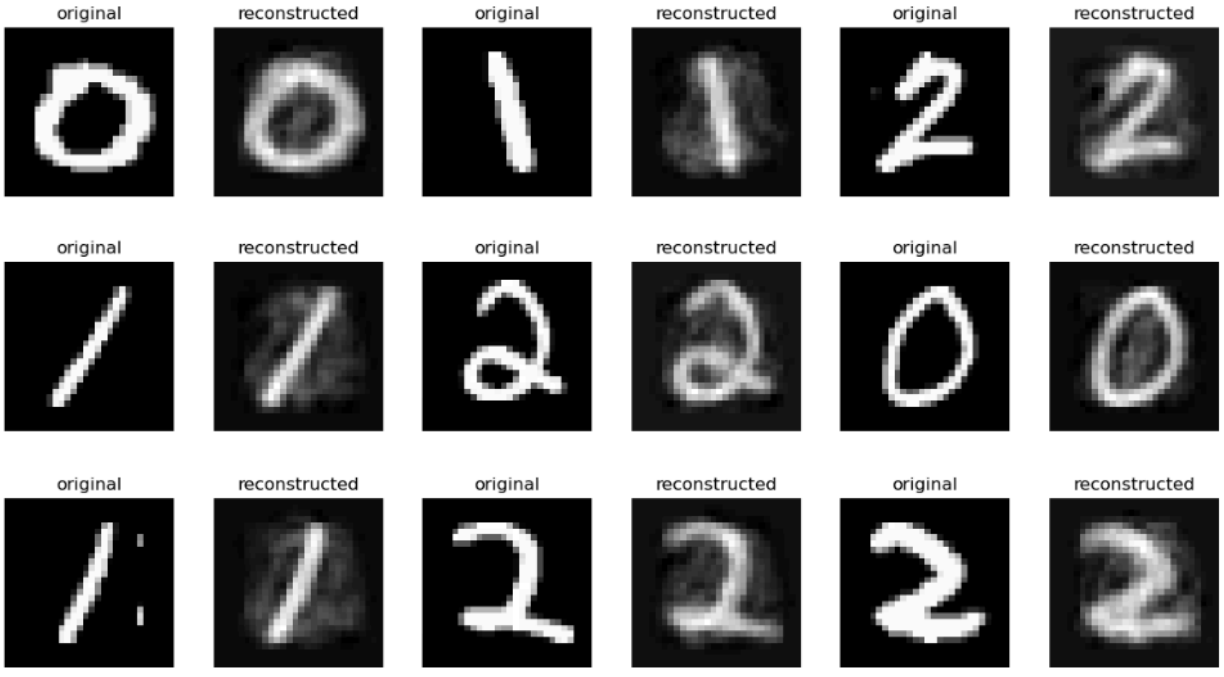

2D Scatter Plot of Three Classes
Class 0
Class 1
Class 2
Y-axis
X-axis
6
4
2
0
−2
−4
−4
−2
0
2
4


2D Scatter Plot of Three Classes
Class 0
Class 1
Class 2
Y-axis
X-axis
4
2
0
−2
−4
−6
−4
−2
0
2
4

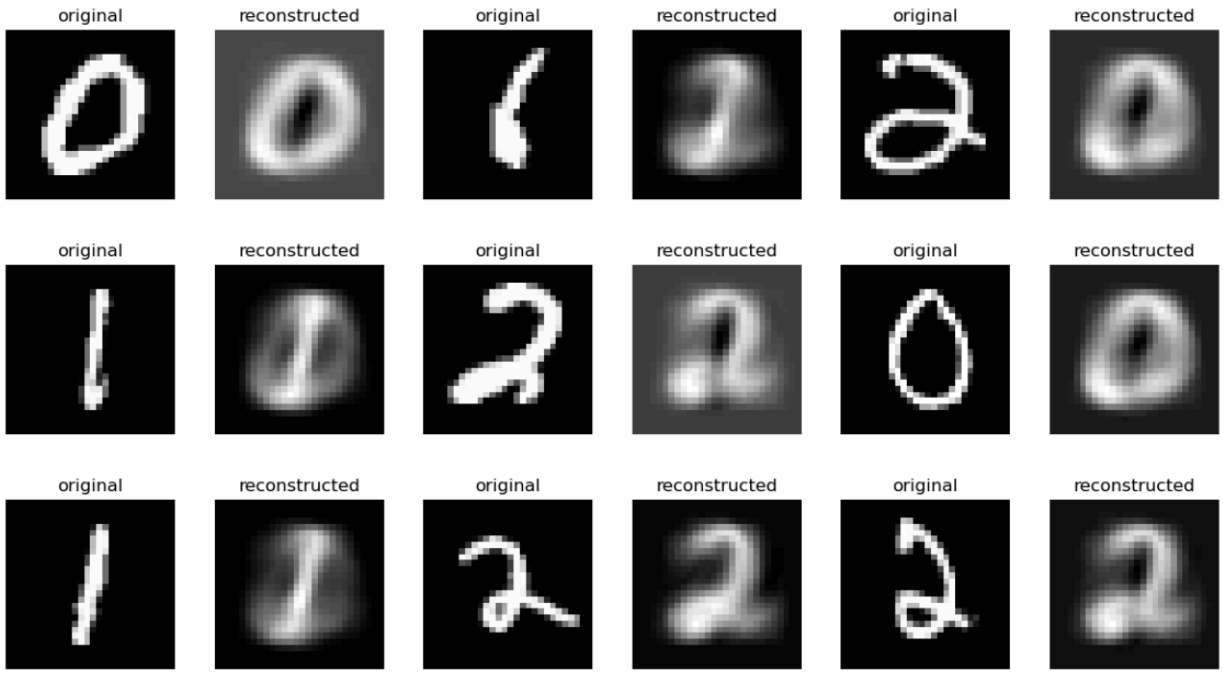
PCA [90%]

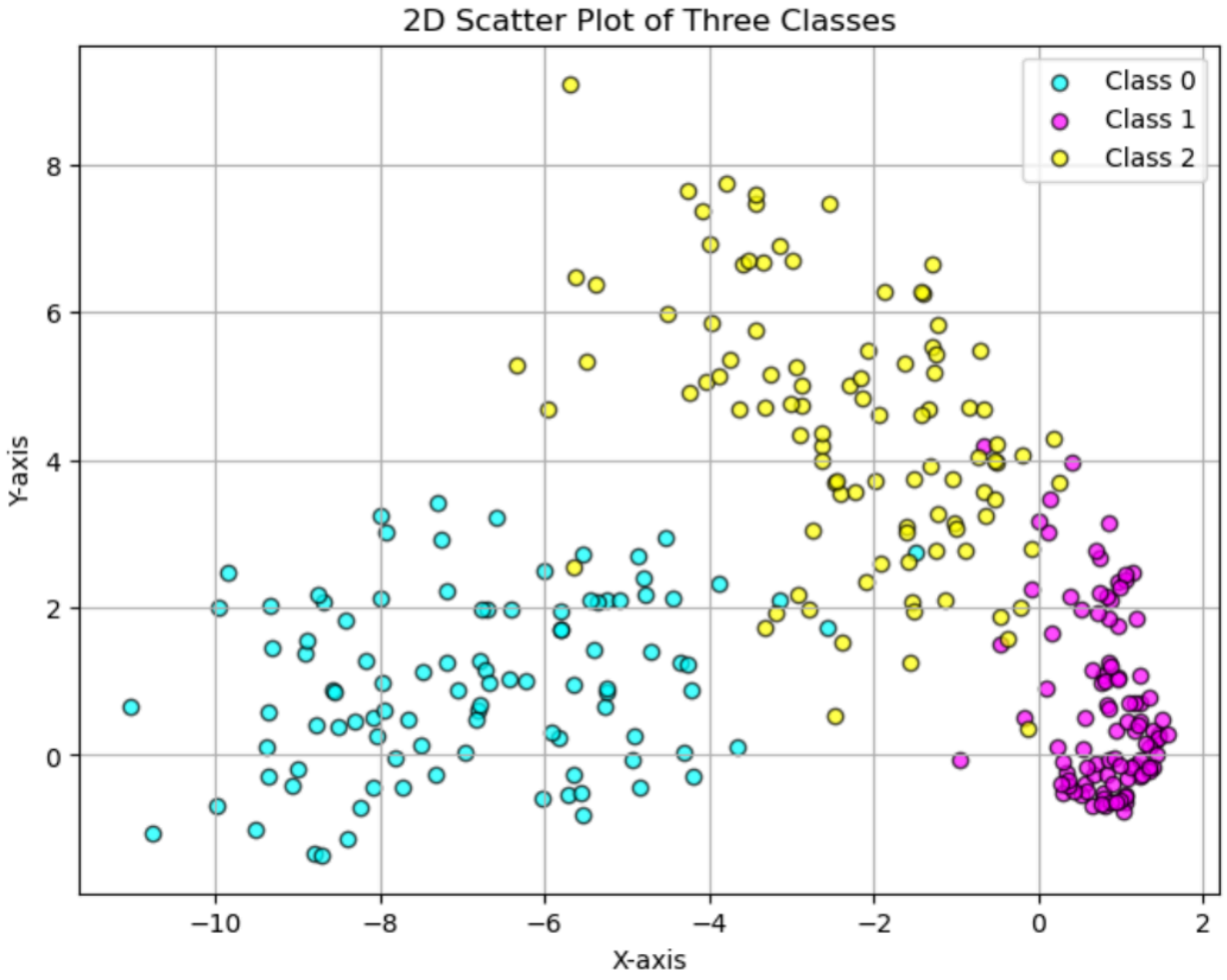

original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed

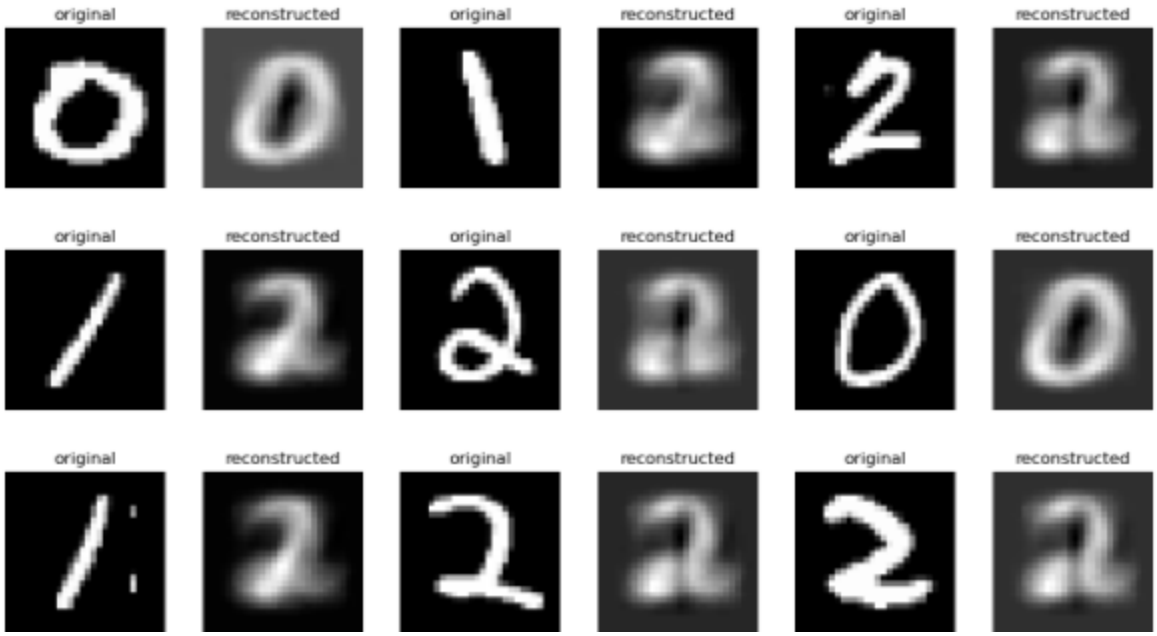
p = 50

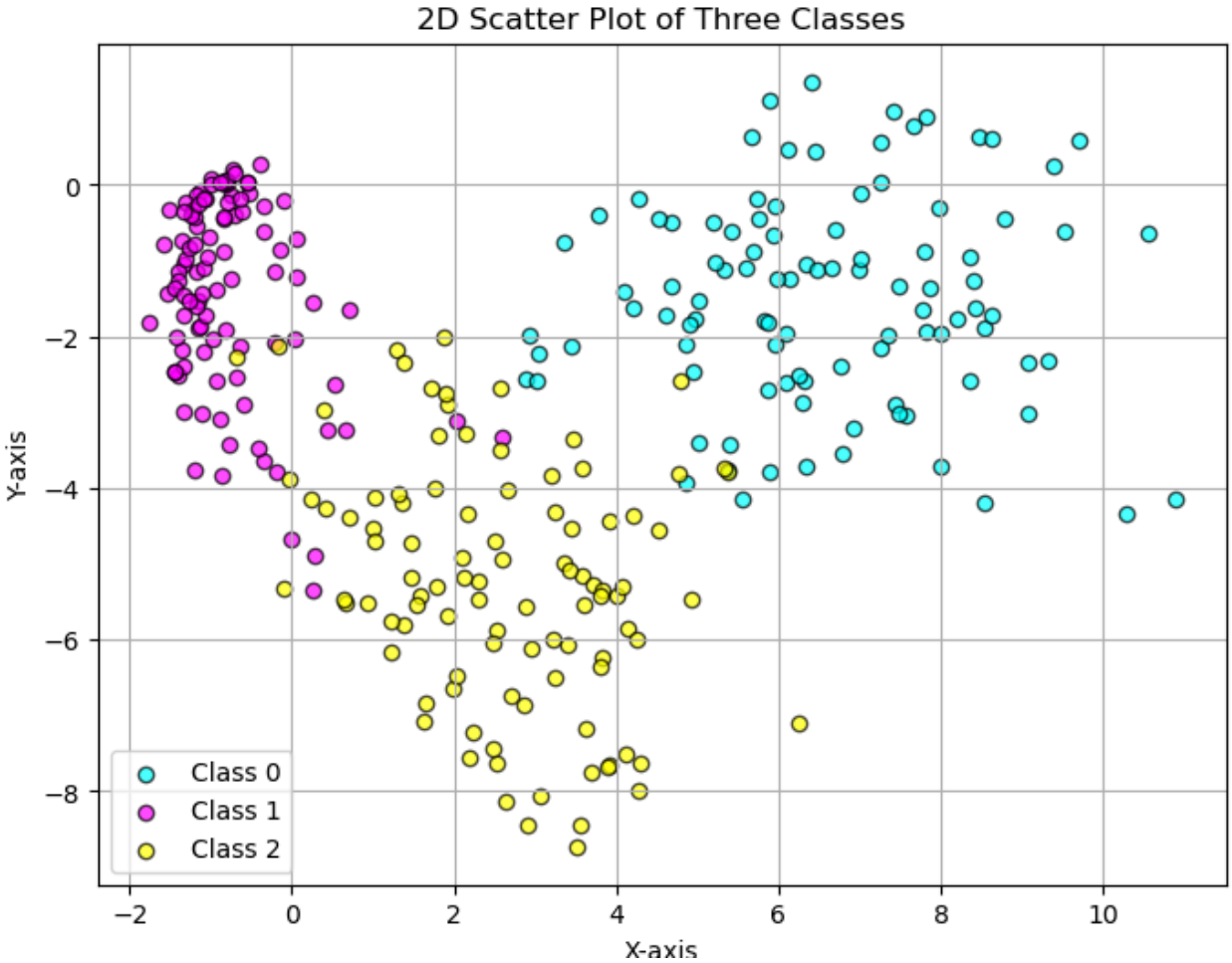

original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed

PCA [2D]


original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed


2D Scatter Plot of Three Classes
Class 0
Class 1
Class 2
Y-axis
X-axis
8
6
4
2
0
−10
−8
−6
−4
−2
0
2


original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed
original
reconstructed


2D Scatter Plot of Three Classes
Y-axis
X-axis
Class 0
Class 1
Class 2
0
−2
−4
−6
−8
−2
0
2
4
6
8
10