Assignment 2 | Coding Part A

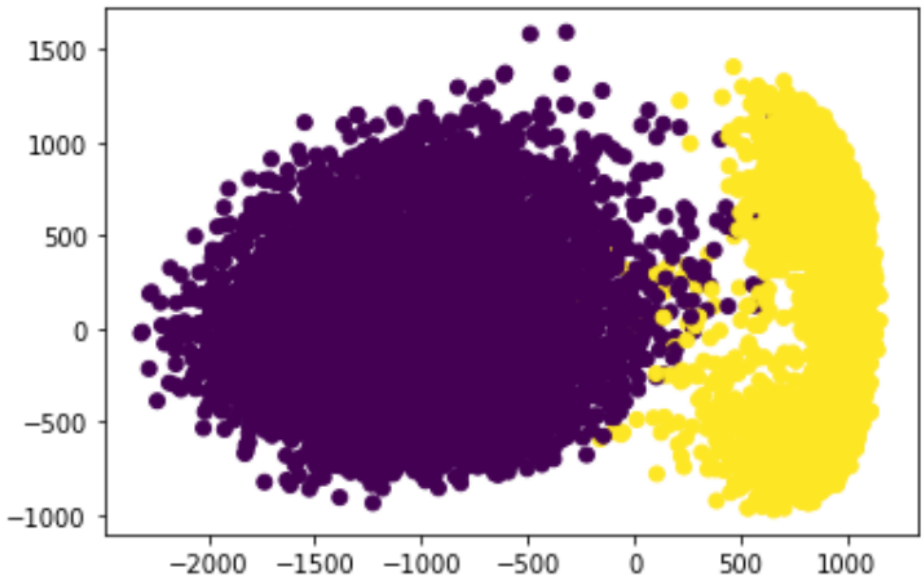
Dataset : we imported mnist dataset from keras.datasets and loaded it in splitted manner. The training set contains 60000 samples and testing contains 10000 samples. Size of each image is 28x28.

Preprocessing :
We need to preprocess the train and test data in such a way that only class 0 and class 1 images remain in the train and test dataset.

Implementation :
Firstly we did visualization of 5 samples from class 0 and class 1. Then we define a PCA function for applying pca to our input data. Projecting the input data for n is 2 and 3 where n is the number of eigenvectors. Then Reconstruct the original data to revert the PCA applied earlier and plot a graph reconstruction error vs n. Again we apply pca on input data followed by LDA to do classification on test data and do this process for all values of n and reported accuracy.

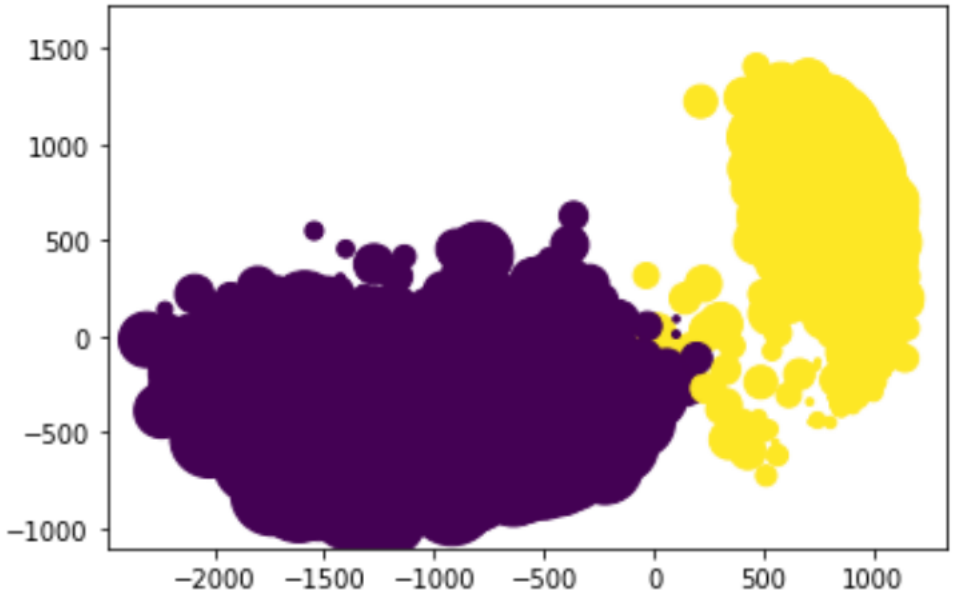
Projection for n = 2 (no of components)

```
(12665, 2)
```

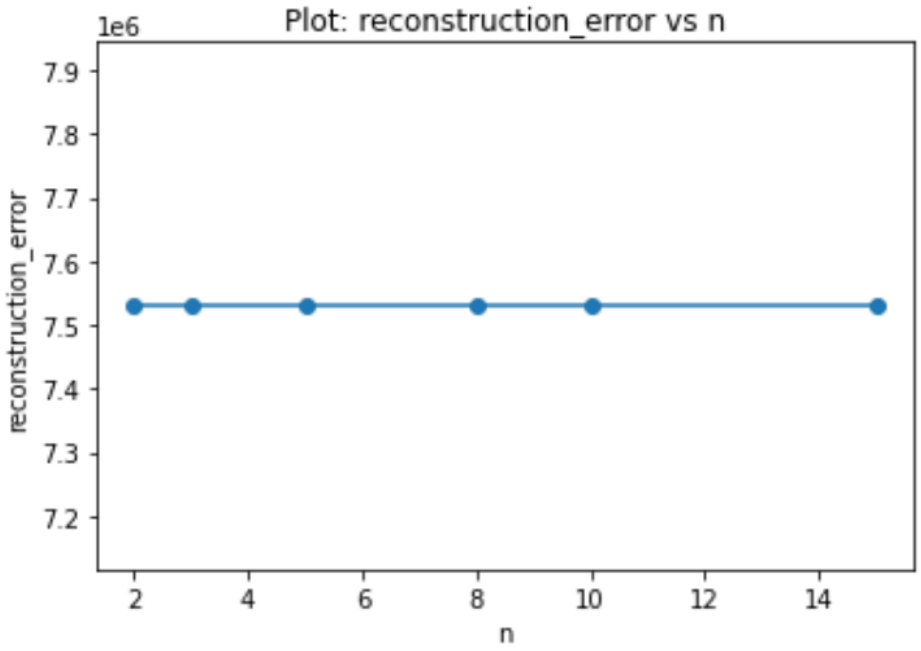


Projection for n = 3 (no of components)

```
(12665, 3)
/usr/local/lib/python3.8/dist-packages/matplotlib/collections
  scale = np.sqrt(self._sizes) * dpi / 72.0 * self._factor
```



Plot : Reconstruction error vs n



PCA followed by LDA (logistic regression used as classifier):

When n = 2 then
accuracy = `0.9966903073286052`

When n = 3 then
accuracy = `0.9971631205673759`

When n = 5 then
accuracy = `0.9957446808510638`

When n = 8 then
accuracy = `0.991016548463357`

When n = 10 then
accuracy = `0.9929078014184397`

When n = 15 then
accuracy = `0.9905437352245863`