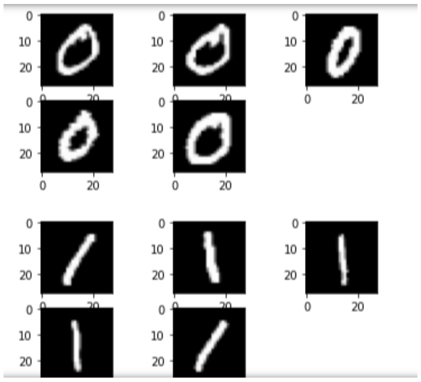
SML Assignment 2 Drishya Uniyal-MT21119

A. Coding.

[30 Marks] **You will explore PCA and FDA in this problem.**

Dataset: MNIST (Note: for every part, just consider class 0 and 1 of MNIST, both for training and testing)

(a) Download the dataset and visualize five samples from each class in images.



(b) Implement PCA from scratch.

(Note: PCA matrix will be calculated from the training data only)

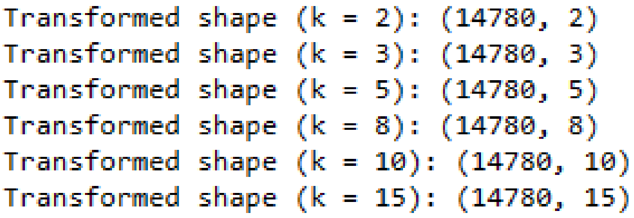
Sol:

1. Calculate the mean of the data
2. Center the data by subtracting the mean
3. Calculate the covariance matrix of the centered data using the formula Cov(X) = (X - mean)^T (X - mean) / (n - 1)
4. Calculate the eigenvectors and eigenvalues of the covariance matrix using the power iteration method
5. Sort the eigenvectors and eigenvalues in descending order of eigenvalue
6. Select the top k eigenvectors that explain the most variance
7. Transform the data into the new feature space

(c) Take n= 2,3,5,8,10,15 to project your input data (Note: here, n is the number of eigenvectors).

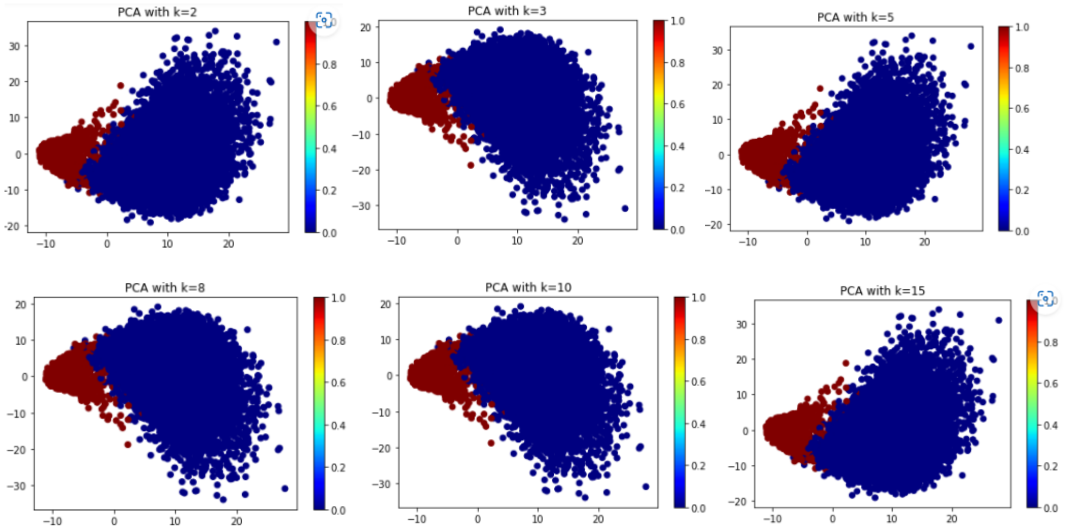
Sol:

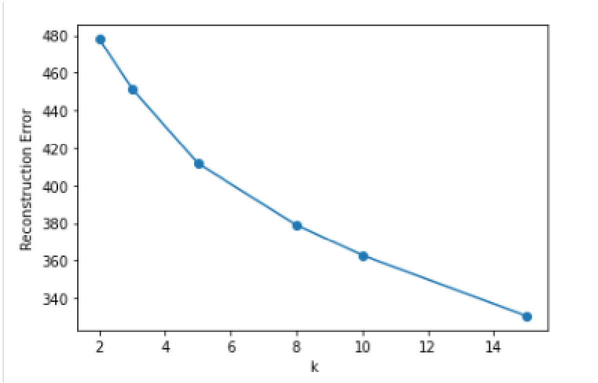
1. **Same apply PCA made in above**
2. **Take different values of n and check results.**



```
Transformed shape (k = 2): (14780, 2)
Transformed shape (k = 3): (14780, 3)
Transformed shape (k = 5): (14780, 5)
Transformed shape (k = 8): (14780, 8)
Transformed shape (k = 10): (14780, 10)
Transformed shape (k = 15): (14780, 15)
```

(d) Reconstruct the original data from the projected data you got in part c. Plot reconstruction error vs. n.





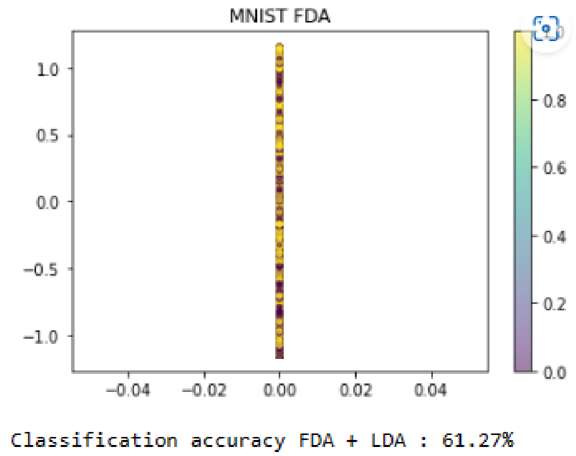
(e) Apply pca on Input data and use LDA to classify the testing samples. Perform this part
for all n given in part (c) and report accuracy on the test set for all cases.
(Note: you can use sklearn for LDA in all parts, parameters of LDA needs to be calculated using training data only)

N	Accuracy
2	99.05
3	99.19
5	99.39
8	99.26
10	99.63
15	**99.73**

(f) Implement FDA from scratch. (FDA vector will be calculated using training data only)
Sol:

1. Compute the mean vectors of each class.
2. Compute the within-class scatter matrix
3. Compute the between-class scatter matrix
4. Compute the eigenvectors and eigenvalues of (S_W)^(-1) S_B
5. Sort the eigenvectors by decreasing eigenvalues
6. Select the eigenvectors corresponding to the largest eigenvalues
7. Project the data onto the FDA subspace

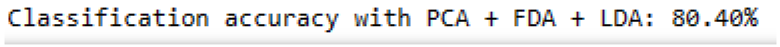
(g) Perform FDA on input data and use LDA to classify testing samples, and report the accuracy.
Sol: **The accuracy comes out to be FDA+LDA - 61.27**



```
Classification accuracy FDA + LDA : 61.27%
```

(h) Apply pca with the best n value from part (e) and then apply FDA, now, use LDA to classify
testing samples and report the accuracy.

Sol: N = 15 best from part (e), PCA + FDA + LDA — > 80.40



```
Classification accuracy with PCA + FDA + LDA: 80.40%
```