Introduction To Machine Learning

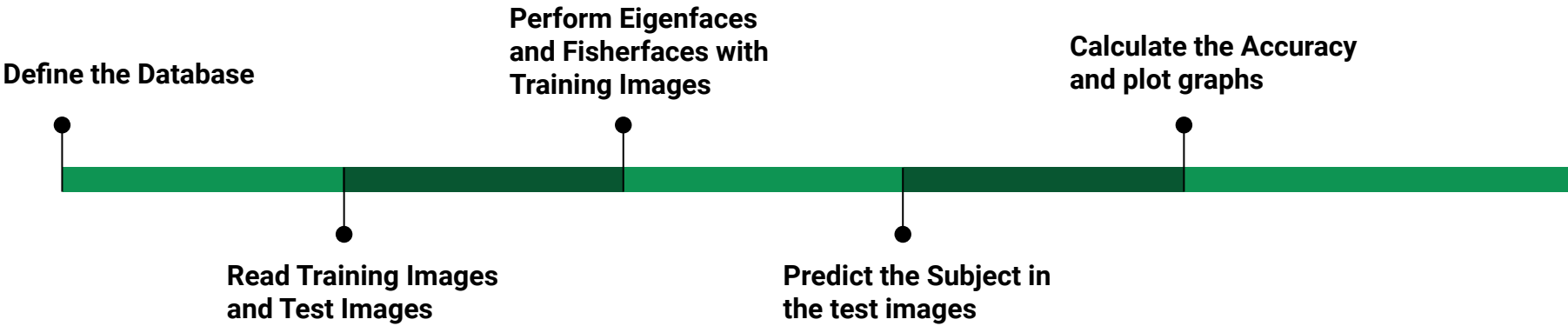
Assignment 2
Eigen faces and Fisher faces

Shivangi Bithel, Yukti Kaura, Vishnu

Index

- Assignment Task
- Dataset Explanation
- EigenFaces For 2 class
  - Training & Testing images
  - Mean and Normalized Images
  - EigenFaces and Test Results
  - Choice of Threshold
  - PCA, Choice of K, K-means Clustering plots
- EigenFaces for Multiple Classes
- FisherFaces for Multiple Classes
- Accuracy
- Future Work

Assignment Task:



Dataset: AT&T Face database

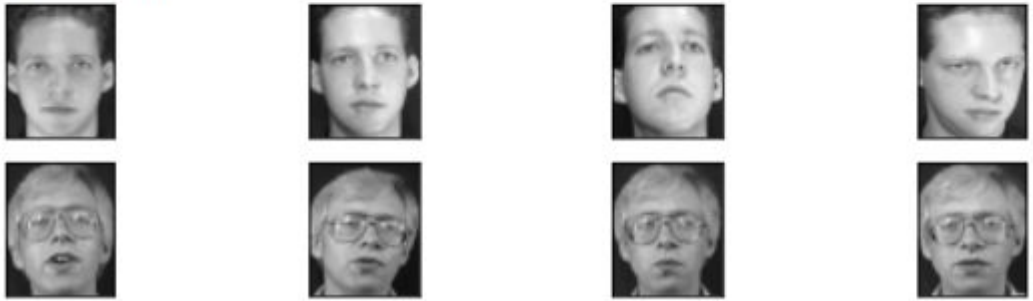
- This database contains a set of face images taken between April 1992 and April 1994 at the AT&T Laboratories Cambridge.
- There are 10 different images of each of 40 distinct subjects.
- the images were shot at different times with different lighting conditions, facial expressions and facial details (glasses/no glasses)
- The size of each image is 92x112 pixels, with 256 grey levels per pixel.
- Total number of features = 92*112 = 10,304

Eigenfaces for 2 Class

Training(6) & Testing(4) Images

Train Images:

Test Images:

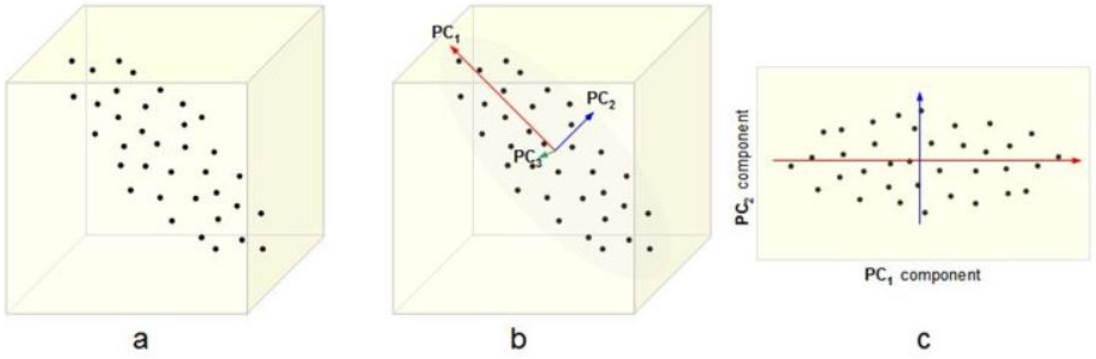


Eigen Faces

- Split the available data into train and test set.
- From the training set of face images, calculate the eigenfaces, which define the face space using principal component analysis.
- When a new face image is encountered, calculate a set of weights based on the input image and the M eigenfaces by projecting the input image onto each of the eigenfaces.
- Determine if the image is a face (whether known or unknown) by checking to see if the image is sufficiently close to “face space.”
- Classify the weight pattern as either a known person or as unknown.

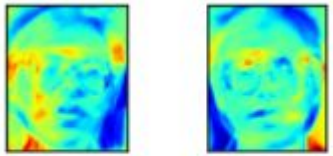
Principal Component Analysis

- Principal Component Analysis, is a dimensionality-reduction method that is used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

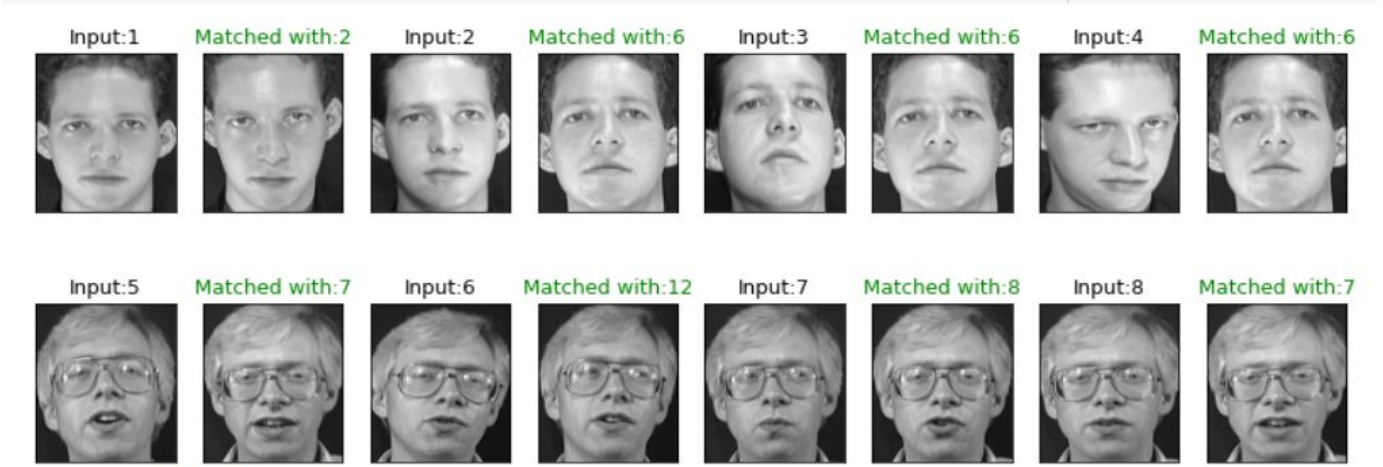


Mean Face & Normalized Training Images

Eigen faces (2 out of 10304) & Results

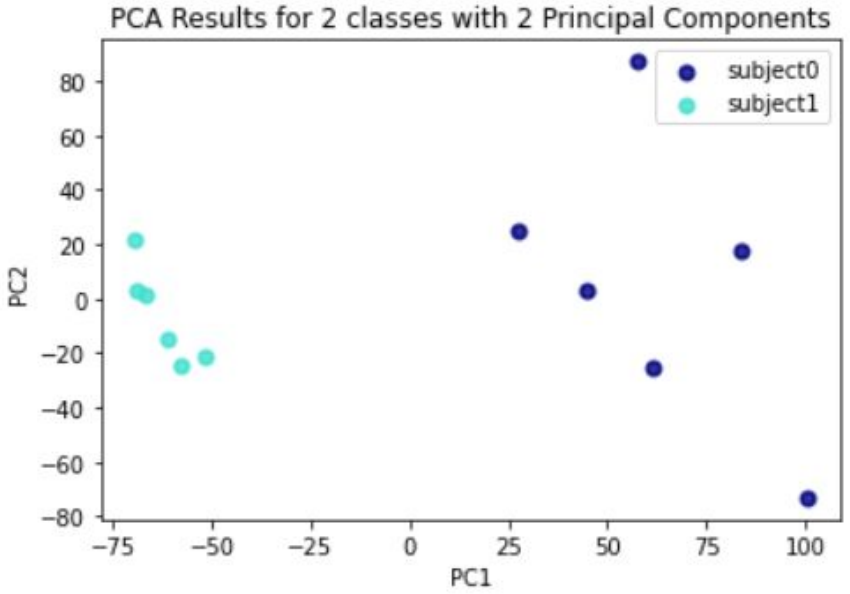


(2, 10304)



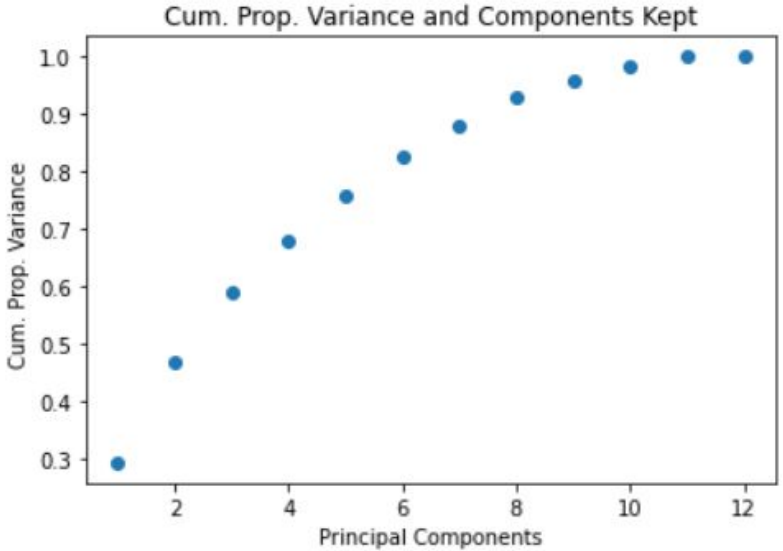
Accuracy: 8/8 = 100.0%

PCA Results

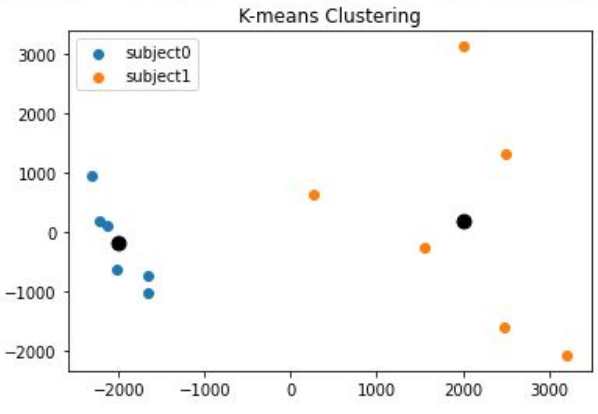

PCA Results for 2 classes with 2 Principal Components
subject0
subject1
PC2
PC1
80
60
40
20
0
−20
−40
−60
−80
−75
−50
−25
0
25
50
75
100

The choice of k

```
Cumulative proportion of variance explained vector: [0.29218049 0.4690388  0.58883799 0.6788435  0.75654836 0.8264829
 0.87872839 0.92771232 0.95649748 0.98432127 1.         1.        ]
```



K-Means Clustering

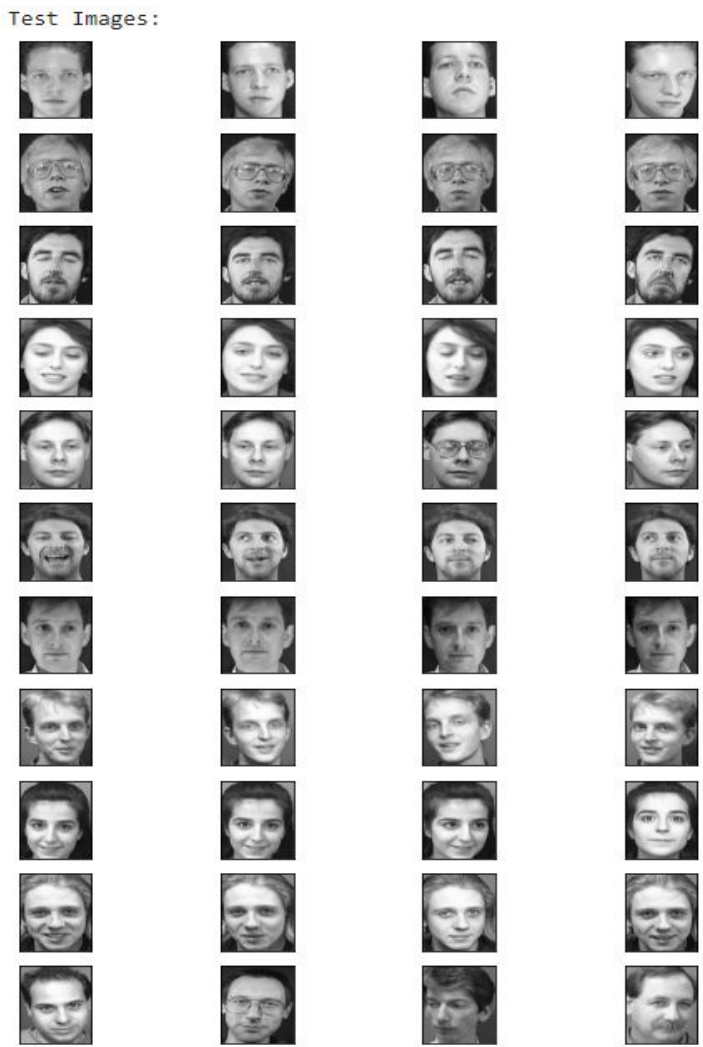



```
Distance of image  0  of subject 0 from its cluster centroid =  649.4748762540586
Distance of image  0  of subject 0 from other cluster centroid =  3770.3176418809685
Distance of image  1  of subject 0 from its cluster centroid =  444.01108229740043
Distance of image  1  of subject 0 from other cluster centroid =  4113.846658754157
Distance of image  2  of subject 0 from its cluster centroid =  1179.1547651708313
Distance of image  2  of subject 0 from other cluster centroid =  4380.787402981122
Distance of image  3  of subject 0 from its cluster centroid =  893.499916956754
Distance of image  3  of subject 0 from other cluster centroid =  3860.916471139779
Distance of image  4  of subject 0 from its cluster centroid =  321.14677925873997
Distance of image  4  of subject 0 from other cluster centroid =  4124.857187409977
Distance of image  5  of subject 0 from its cluster centroid =  444.00266373808694
Distance of image  5  of subject 0 from other cluster centroid =  4221.956588802311
--------------------------------------------------------------------------------
Distance of image  0  of subject 1 from its cluster centroid =  2947.525176728927
Distance of image  0  of subject 1 from other cluster centroid =  5202.412053679498
Distance of image  1  of subject 1 from its cluster centroid =  1792.5769485136504
Distance of image  1  of subject 1 from other cluster centroid =  2409.595162738013
Distance of image  2  of subject 1 from its cluster centroid =  2570.0205723297336
Distance of image  2  of subject 1 from other cluster centroid =  5544.233937766335
Distance of image  3  of subject 1 from its cluster centroid =  1228.3326688695104
Distance of image  3  of subject 1 from other cluster centroid =  4744.29614685127
Distance of image  4  of subject 1 from its cluster centroid =  1854.9111482864837
Distance of image  4  of subject 1 from other cluster centroid =  4696.9456114387385
Distance of image  5  of subject 1 from its cluster centroid =  636.3053132799802
Distance of image  5  of subject 1 from other cluster centroid =  3556.626539409943
```

Eigenfaces for Multiple Classes

Training Images in case of multiple classes (10 * 6 =60)

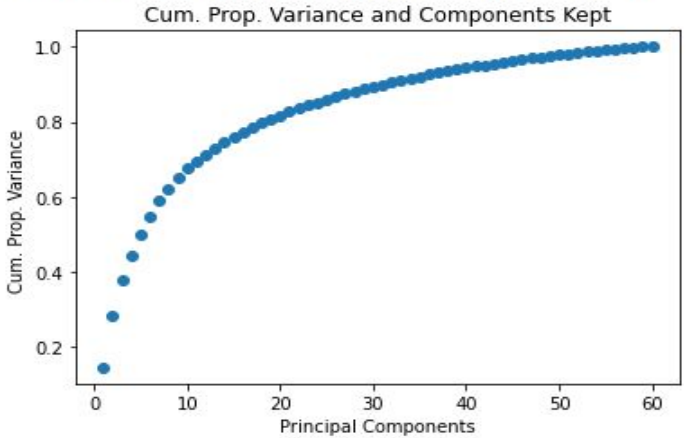
Testing Images in cases of multiple classes (10*4 = 40)



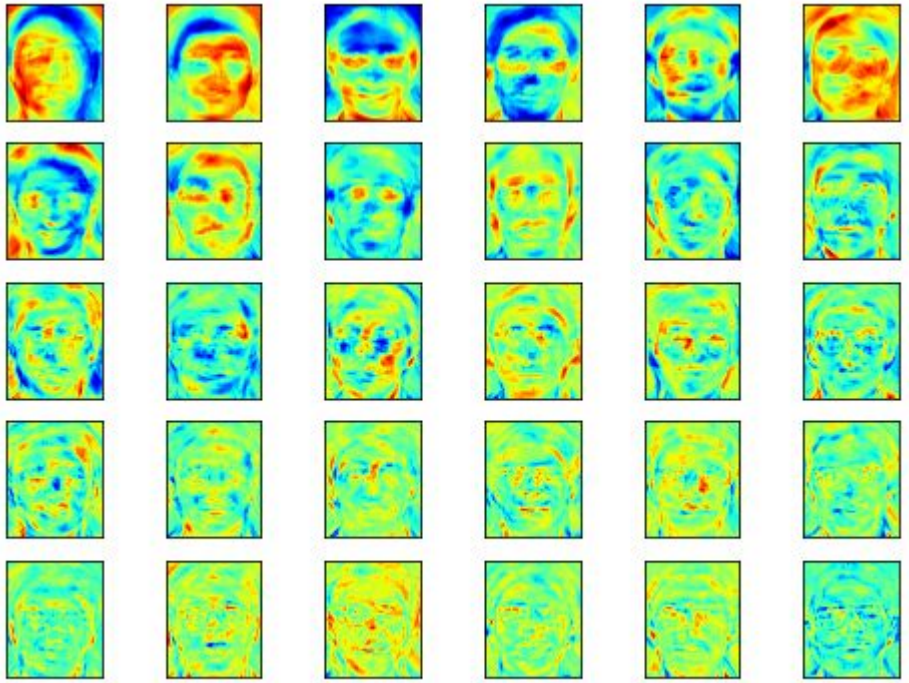
Mean and Normalized Training Images

Choice of K

```
Cumulative proportion of variance explained vector: [0.14621575 0.28249125 0.37746986 0.44446901 0.50050548 0.54955529
 0.59036522 0.62109044 0.64933827 0.67549917 0.69539551 0.71369034
 0.73030772 0.74607903 0.75955801 0.77238469 0.78480201 0.7960013
 0.80677497 0.81690991 0.82641362 0.8355828  0.84401968 0.85203057
 0.85961327 0.86682783 0.87386943 0.8807468  0.88700769 0.89293829
 0.89878879 0.90452133 0.91003453 0.91547771 0.92065132 0.92559223
 0.93042917 0.93489995 0.93916347 0.94324132 0.94728416 0.95122782
 0.95507255 0.95880403 0.96240433 0.96587795 0.96928831 0.9724665
 0.97561928 0.97866597 0.98161517 0.98449337 0.98717403 0.98973247
 0.99214526 0.99451831 0.99678496 0.9986109  1.         1.        ]
```

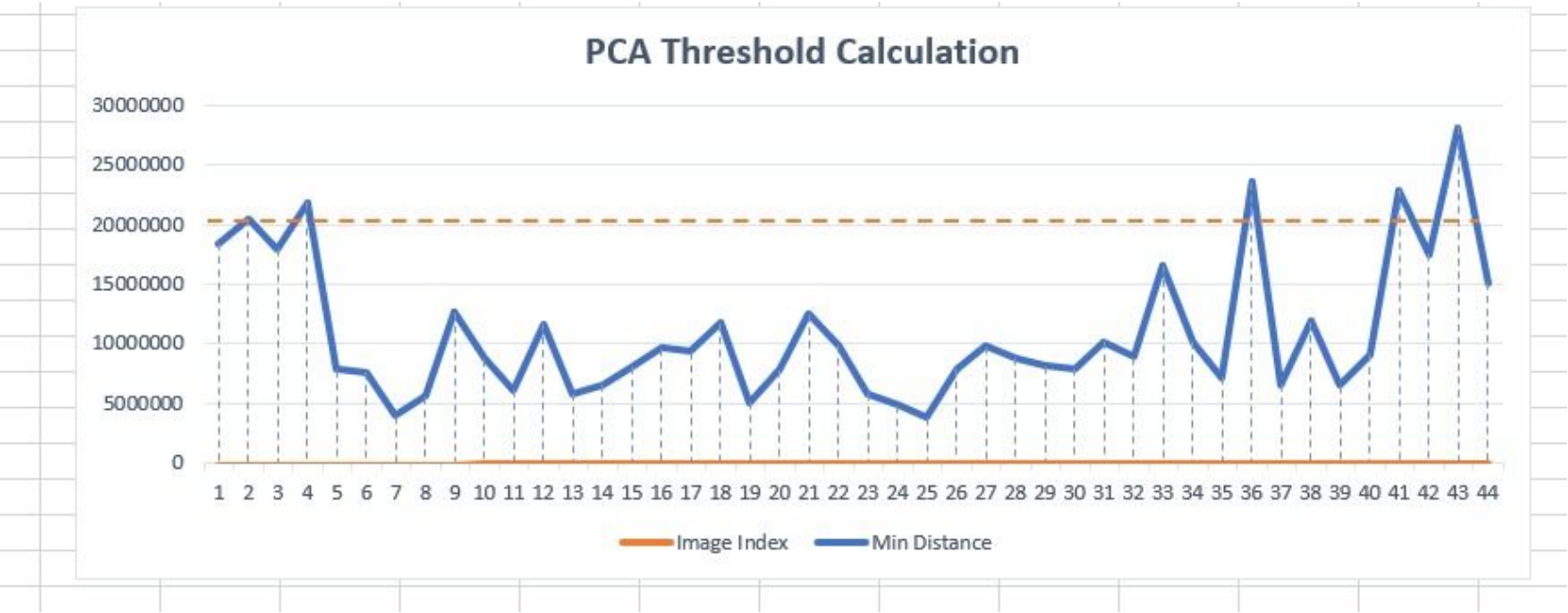


Top 30 Eigen Faces

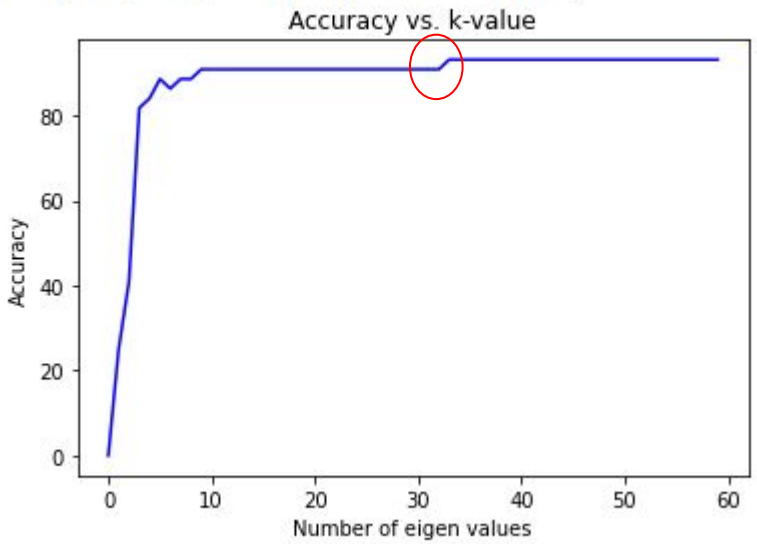


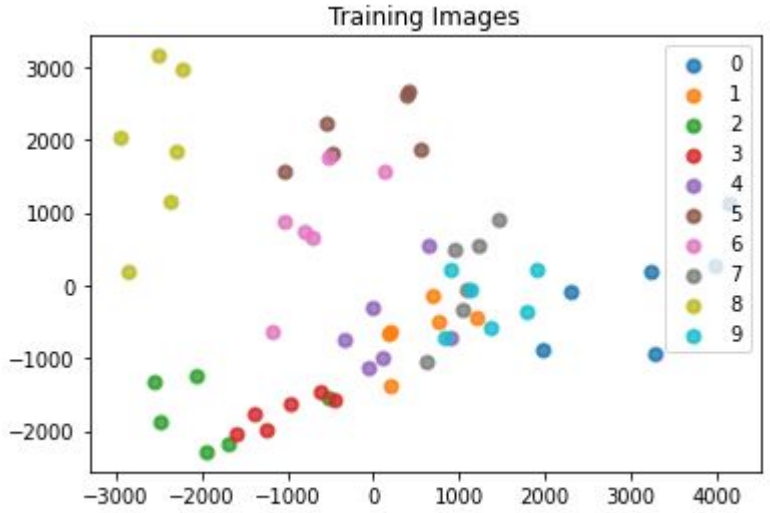
(30, 10304)

Choice of Threshold

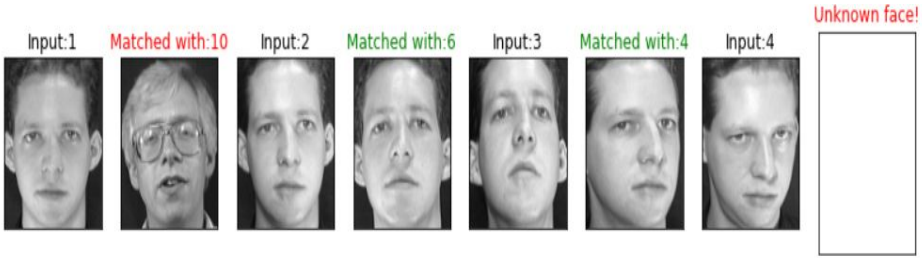


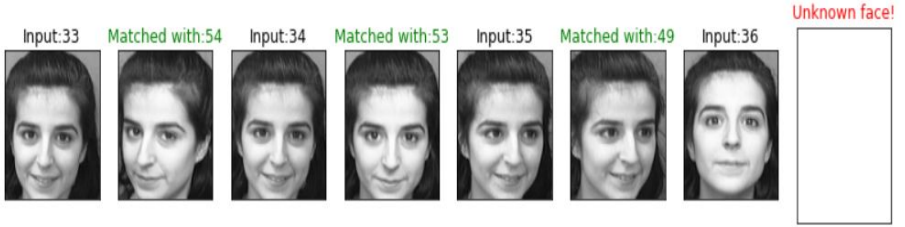
Accuracy vs K-value & K-Means CLustering





Test Results





Fisherfaces for Multiple Classes

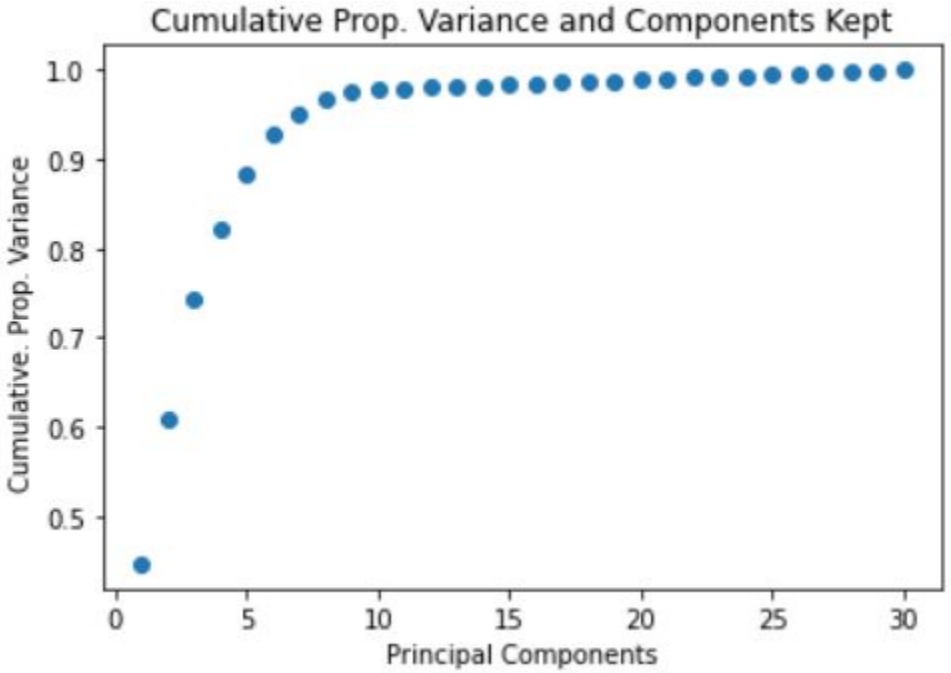
Fisher Faces

- Split the available data into train and test set.
- From the training set of face images, calculate the eigenvalues, which define the eigenspace using principal component analysis.
- Find the mean vectors of each class from the dataset and an overall mean vector of all the training images.
- Compute scatter matrices
  - Between class scatter matrix - S_B
  - Within class scatter matrix - S_W

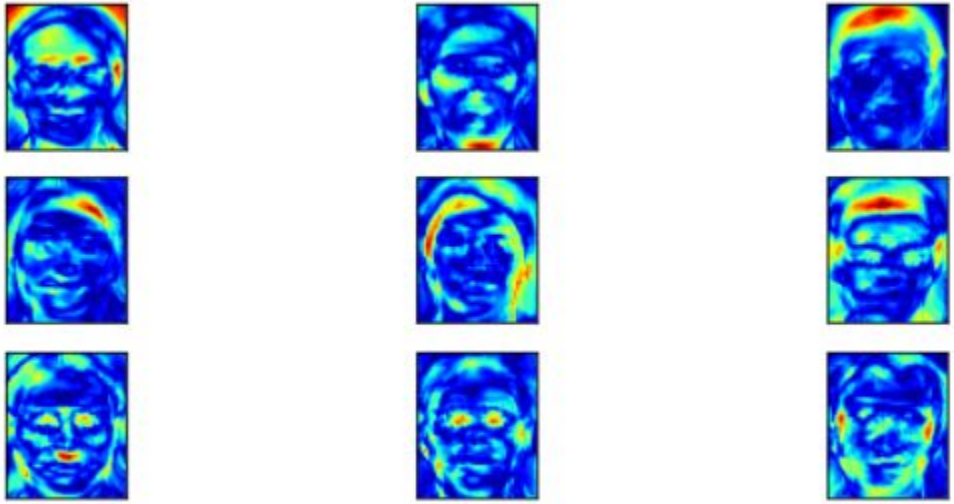
Fisher LDA

- $J = S_B \setminus S_W$
- Compute the eigenvectors and corresponding eigenvalues for the scatter matrices J
- Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors
- $W_{FS} = W_{PCA}$ * Reduced data of LDA
- When a new face image is encountered, calculate a set of weights based on the input image and the k fisherfaces by projecting the input image onto each of the fisherfaces.
- Determine if the image is a face (whether known or unknown) by checking to see if the image is sufficiently close to “face space.”
- Classify the weight pattern as either a known person or as unknown.

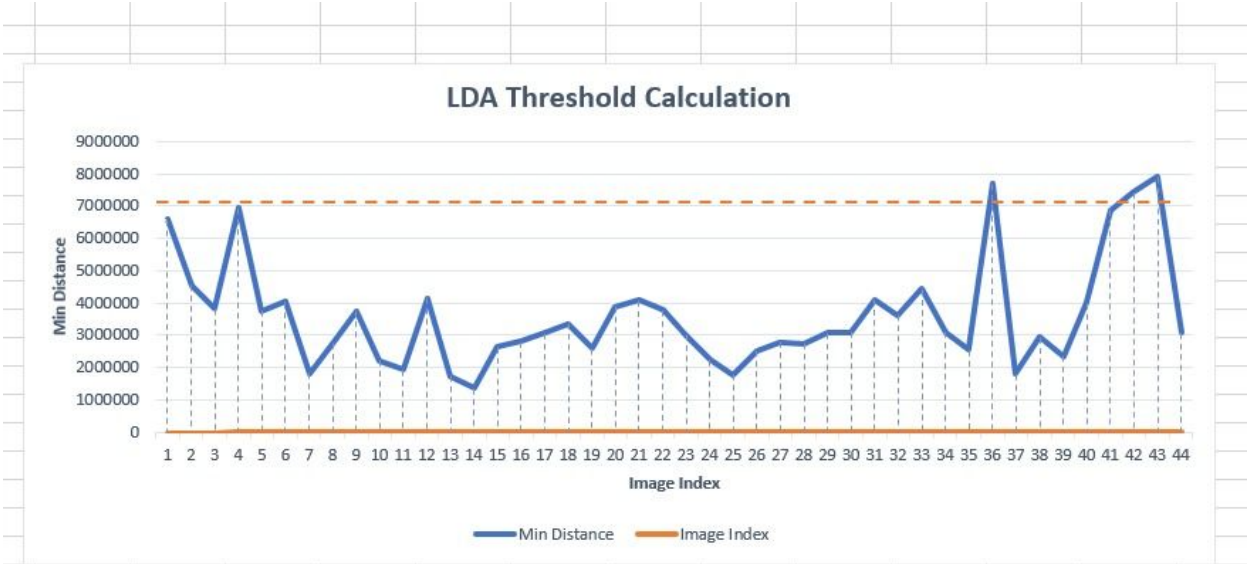
Choice of K


Cumulative Prop. Variance and Components Kept
Cumulative. Prop. Variance
1.0
0.9
0.8
0.7
0.6
0.5
0
5
10
15
20
25
30
Principal Components

Fisher Faces



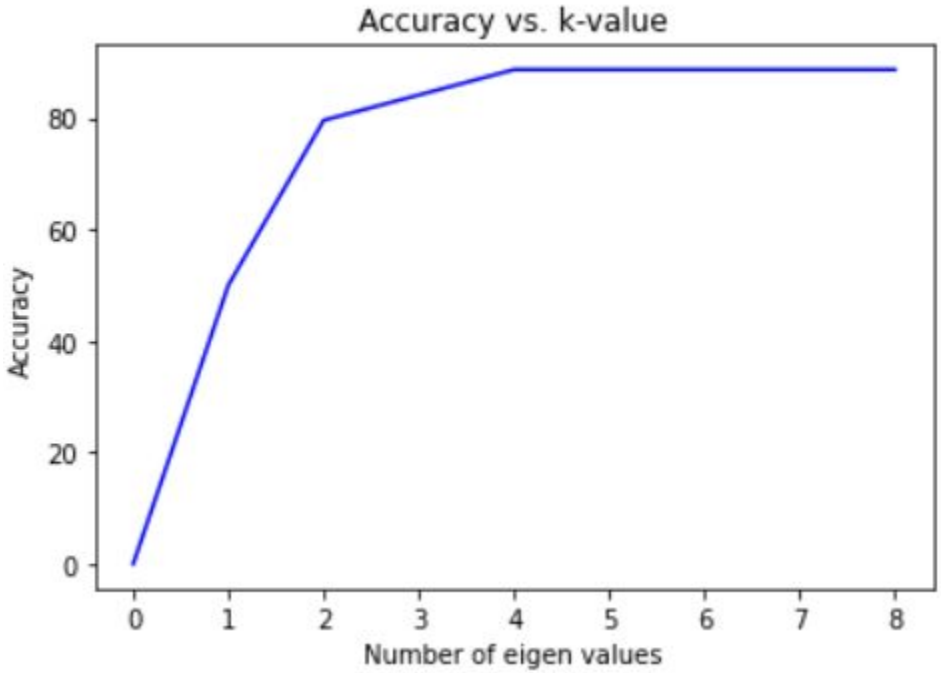
Mean Face & Threshold Calculation



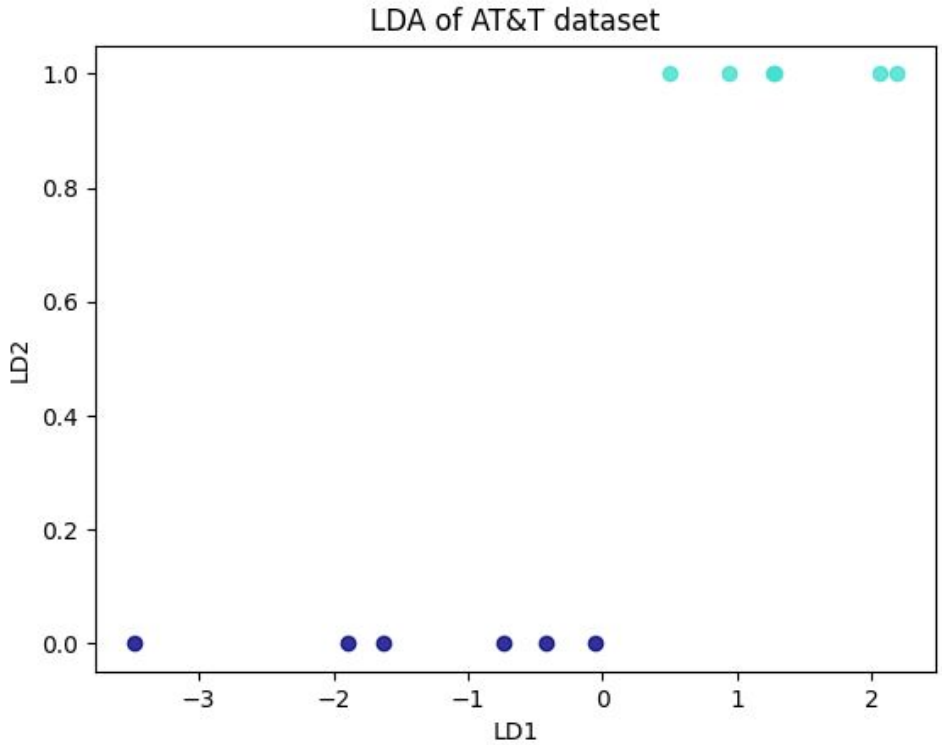
Results


Input:1
Matched with:2
Input:2
Matched with:4
Input:3
Matched with:3
Input:4
Matched with:2
Input:5
Matched with:7
Input:6
Matched with:11
Input:7
Matched with:11
Input:8
Matched with:7
Input:9
Matched with:14
Input:10
Matched with:17
Input:11
Matched with:14
Input:12
Matched with:16

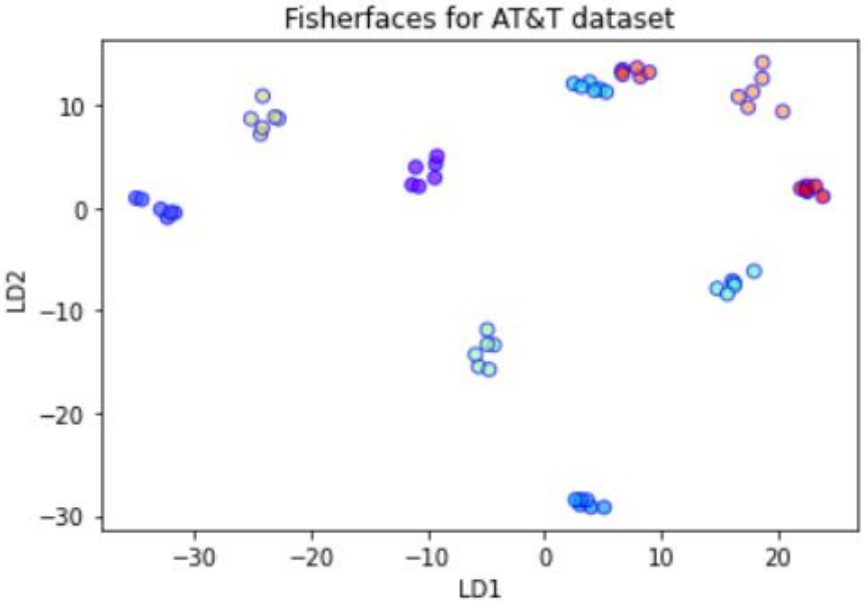
Accuracy vs k-value

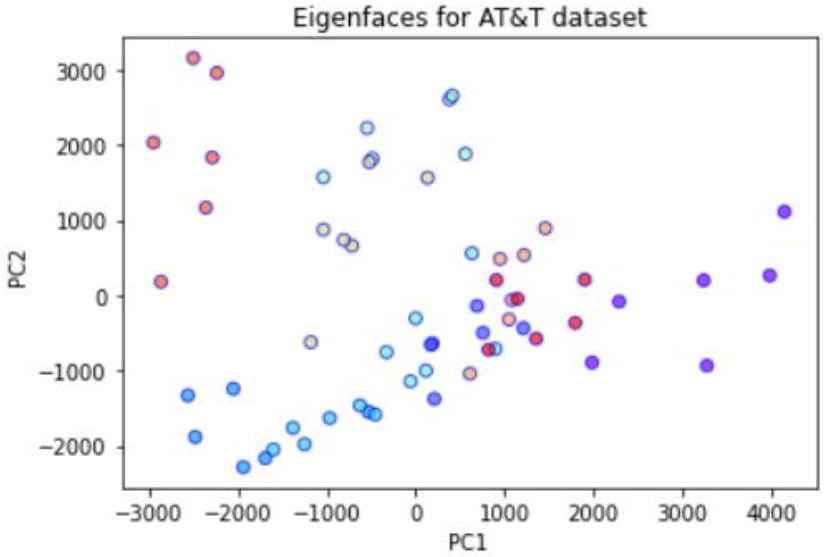


LDA Results for 2 class problem



FisherFaces and EigenFaces Projection on Multiclass





Final Accuracy

```
python 2020CSY7657_2020BSZ8013_2020EEY7625_Fisher_Faces.py
k: 9
Correct predictions: 41/44 = 93.18181818181817%
(base) PS C:\Users\SHIVANGI BITHEL\Desktop\ELL784\new\assignment2-main\assignment2-main\2020CSY7657_2020BSZ8013_2020EEY7625>
python 2020CSY7657_2020BSZ8013_2020EEY7625_Eigen_Faces_Multi_Class.py
k: 30
Correct predictions: 38/44 = 86.36363636363636%
(base) PS C:\Users\SHIVANGI BITHEL\Desktop\ELL784\new\assignment2-main\assignment2-main\2020CSY7657_2020BSZ8013_2020EEY7625>
python 2020CSY7657_2020BSZ8013_2020EEY7625_Eigen_Faces_Two_Class.py
k: 2
Correct predictions: 8/8 = 100.0%
(base) PS C:\Users\SHIVANGI BITHEL\Desktop\ELL784\new\assignment2-main\assignment2-main\2020CSY7657_2020BSZ8013_2020EEY7625>
```

Future Work

- This work can be extended by doing more number of experiments like building a classifier for glass detection
- Further we can include a real time training, incase an image which was not initially available in training and is occurring again and again, weights for such image can be learned and included in the training data.
- Face Recognition in videos

THANK YOU :)