Nisaruj Rattanaaram 6031033521

Homework 3 Fisherface

Hello Soft Clustering (GMM)

T1. Using 3 mixtures, initialize your Gaussian with means (3,3), (2,2), and (-3,-3), and standard Covariance, I, the identity matrix. Use equal mixture weights as the initial weights. Repeat three iterations of EM. Write down $w_{n,j}$, m_j, μ_j, Σ_j for each EM iteration. (You may do the calculations by hand or write code to do so)

```
-------------------- ITER: 1 --------------------

wij

   0.119203     0.880797      1.81546e-09
   0.731059     0.268941      1.69571e-16
   0.268941     0.731059      1.01529e-11
   0.999983     1.67014e-05   2.03106e-42
   0.999089     0.000911051   5.37528e-32
   0.999877     0.000123395   3.30529e-37
   2.31952e-16  1.38879e-11             1
   2.31952e-16  1.38879e-11             1
   3.3057e-37   5.90009e-29             1

Weight (mj)

       Mixture 1 : 0.45757241940119386
       Mixture 2 : 0.20909424706571345
       Mixture 3 : 0.33333333353309275

Mean (μj)

       Mixture 1 : [5.78992692 5.81887265]
       Mixture 2 : [1.67718211 2.14523106]
       Mixture 3 : [-4.        -4.66666666]

Covariance Matrix (Σj)
       Mixture 1 :
              4.5362        0
              0             4.28701
       Mixture 2 :
              0.516457      0
              0             0.131527
       Mixture 3 :
              4.66667        0
              0              2.88889

-------------------- ITER: 2 --------------------
wij

 0.00316934      0.996825       5.96939e-06
   0.655097      0.344903       6.84249e-07
 0.00577506      0.994224       1.30003e-06
          1      9.15472e-73    4.53104e-19
          1      3.18387e-32    5.49966e-14
          1      1.60399e-50    1.66511e-16
4.73614e-08      1.9814e-52               1
3.08504e-08      1.36026e-67              1
5.39514e-16      1.09037e-168             1
```

Weight (m_j)

Mixture 1 : 0.407115699156724
Mixture 2 : 0.25955009245849986
Mixture 3 : 0.3333342083847761

Mean (μ_j)

Mixture 1 : [6.27176603 6.27263099]
Mixture 2 : [1.72091785 2.14764973]
Mixture 3 : [-3.99998589 -4.6666488]

Covariance Matrix (Σ_j)

Mixture 1 :

2.94672	0
0	2.93846

Mixture 2 :

0.496496	0
0	0.12585

Mixture 3 :

4.66673	0
0	2.889

-------------------- ITER: 3 --------------------

w_{ij}

9.82876e-05	0.999897	5.04309e-06
0.24595	0.754049	1.23089e-06
0.000318008	0.999681	9.94314e-07
1	9.45933e-76	3.14531e-19
1	1.86812e-33	4.19736e-14
1	1.37671e-52	1.08249e-16
5.61691e-13	6.99738e-55	1
3.64883e-13	1.02777e-70	1
1.03024e-25	1.73453e-176	1

Weight (m_j)

Mixture 1 : 0.3607073596545177
Mixture 2 : 0.3059584994238061
Mixture 3 : 0.33333414092167624

Mean (μ_j)

Mixture 1 : [6.6962821 6.69631238]
Mixture 2 : [1.91071852 2.27383846]
Mixture 3 : [-3.99998673 -4.6666501]

Covariance Matrix (Σ_j)

Mixture 1 :

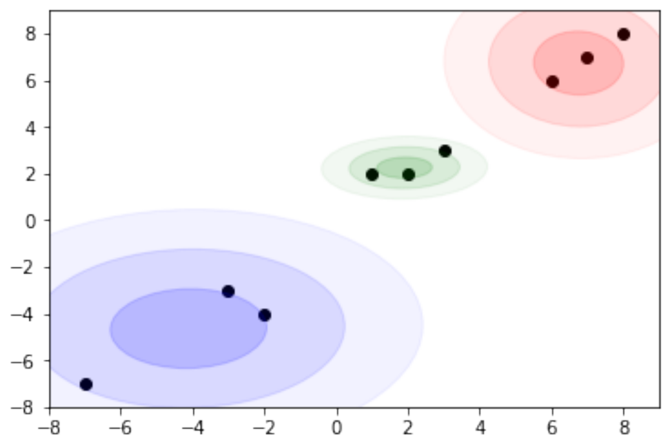
1.73955	0
0	1.73924

Mixture 2 :

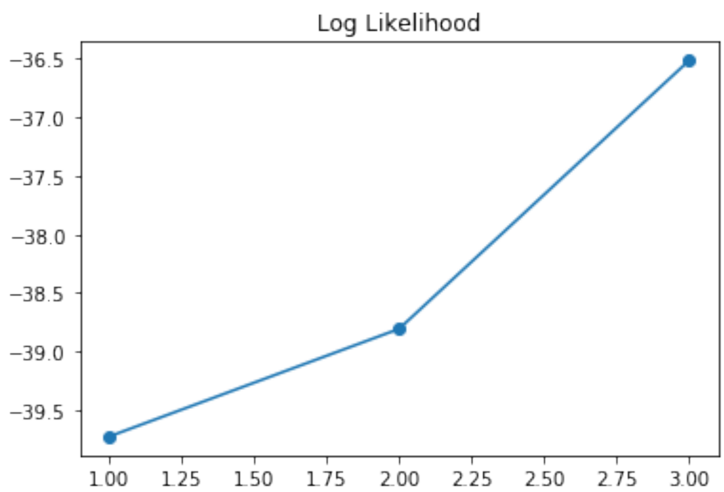
0.628988	0
0	0.198852

Mixture 3 :

4.66673	0
0	2.889



T2. Plot the log likelihood of the model given the data after each EM step. In other words, plot log $_n$ p(x_n|φ, μ,Σ). Does it goes up every iteration just as we learned in class?



Log Likelihood มีค่าเพิ่มขึ้นเรื่อย ๆ ในแต่ละ Iteration

T3. Using 2 mixtures, initialize your Gaussian with means (3,3) and (-3,-3), and standard Covariance, I, the identity matrix. Use equal mixture weights as the initial weights. Repeat three iterations of EM. Write down $w_{n,j}$,m_j,μ_j,Σ_j for each EM iteration.

```
-------------------- ITER: 1 --------------------

Wij

          1                     1.523e-08
          1                     2.31952e-16
          1                     3.77513e-11
          1                     2.03109e-42
          1                     5.38019e-32
          1                     3.3057e-37
          2.31952e-16           1
          2.31952e-16           1
          3.3057e-37            1
```

Weight (m_j)

    Mixture 1 : 0.6666666649702521
    Mixture 2 : 0.3333333350297478

Mean (μ_j)

    Mixture 1 : [4.50000001 4.66666667]
    Mixture 2 : [-3.99999997 -4.66666663]

Covariance Matrix (Σ_j)

    Mixture 1 :

6.91667	0
0	5.88889

    Mixture 2 :

4.66667	0
0	2.88889

-------------------- ITER: 2 --------------------

w_{ij}

0.999879	0.000120726
1	2.59403e-07
0.999976	2.40783e-05
1	9.39287e-19
1	7.41043e-14
1	2.98366e-16
0.000241448	0.999759
0.000152869	0.999847
5.22429e-09	1

Weight (m_j)

    Mixture 1 : 0.666694362106005
    Mixture 2 : 0.33330563789399464

Mean (μ_j)

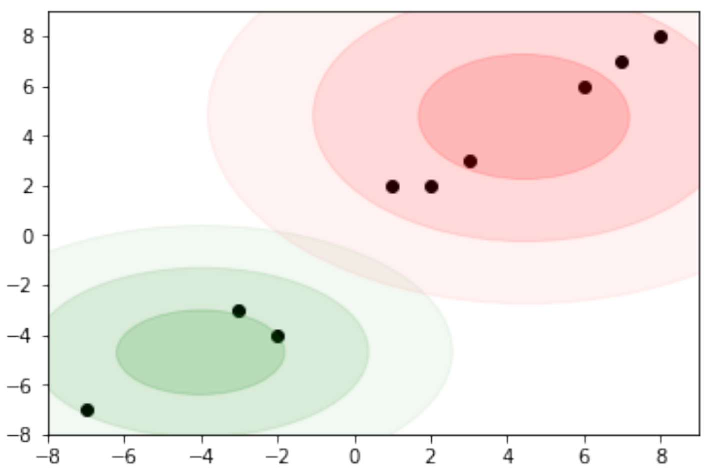
    Mixture 1 : [4.49961311 4.66620178]
    Mixture 2 : [-3.99993241 -4.66651231]

Covariance Matrix (Σ_j)

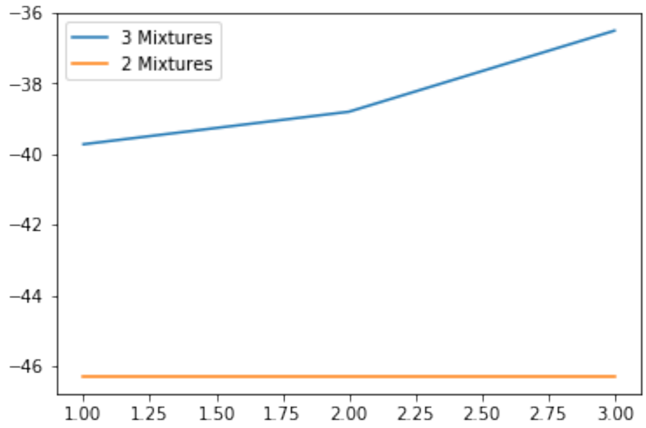
    Mixture 1 :

6.91945	0
0	5.89275

```
    Mixture 2 :

            4.66807        0
            0              2.89103

-------------------- ITER: 3 --------------------

wᵢⱼ

    0.999879            0.000121411
    1                   2.6177e-07
    0.999976            2.42296e-05
    1                   9.63496e-19
    1                   7.54839e-14
    1                   3.04964e-16
    0.000242793         0.999757
    0.000153838         0.999846
    5.28883e-09         1

Weight (mⱼ)

    Mixture 1 : 0.6666945259520648
    Mixture 2 : 0.3333054740479351

Mean (μⱼ)

    Mixture 1 : [4.49961084 4.66619903]
    Mixture 2 : [-3.99993206 -4.66651141]

Covariance Matrix (Σⱼ)

    Mixture 1 :

            6.91946        0
            0              5.89277

    Mixture 2 :

            4.66808        0
            0              2.89105
```



T4. Plot the log likelihood of the model given the data after each EM step. Compare the log likelihood between using two mixtures and three mixtures. Which one has the better likelihood?



3 Mixtures model มี log likelihood ดีกว่า 2 Mixtures model

The face database

The similarity matrix

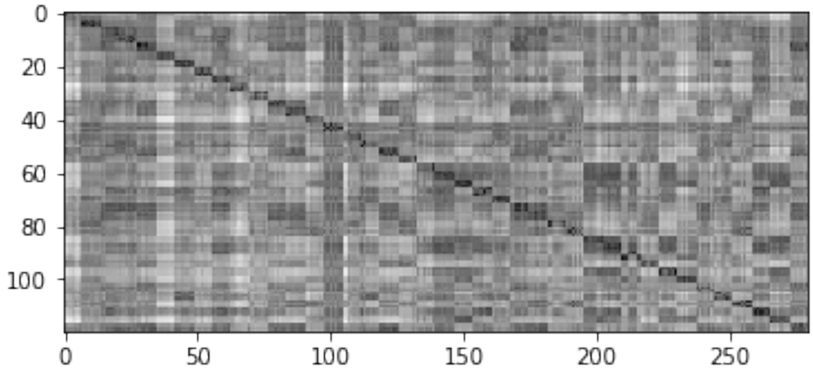
T5. What is the Euclidean distance between xf[0,0] and xf[0,1]? What is the Euclidean distance between xf[0,0] and xf[1,0]? Does the numbers make sense? Do you think these numbers will be useful for face verification?

|| xf[0,0] – xf[0,1] || = 10.037616294165492

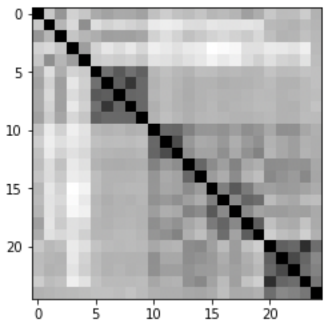
|| xf[0,0] – xf[1,0] || = 8.173295099737281

ตัวเลขยังไม่ make sense เนื่องจากภาพของคนเดียวกันควรจะมี Euclidean Distance น้อยกว่าภาพของต่างคนกัน แต่ตัวเลขเหล่านี้อาจจะนำไปใช้ verification ได้หากจัดการได้ดีพอ

T6. Write a function that takes in a set of feature vectors T and a set of feature vectors D, and then output the similarity matrix A. Show the matrix as an image. Use the feature vectors from the first 3 images from all 40 people for list T (in order x[0,0], x[0,1],x[0,2],x[1,0],x[1,1],...x[39,2]). Use the feature vectors from the remaining 7 images from all 40 people for list D (in order x[0,3],x[0,4],x[0,5],x[1,6],x[0,7],x[0,8],x[0,9],x[1,3],x[1,4]...x[39,9]). We will treat T as our training images and D as our testing images



T7. From the example similarity matrix above, what does the black square between [5:10,5:10] suggest about the pictures from person number 2? What do the patterns from person number 1 say about the images from person 1?



ข้อมูลของรูปคนที่สองสามารถจำแนกได้ชัดเจนกว่าภาพอื่น ๆ เนื่องจาก มี Euclidean Distance ระหว่างภาพของคนที่ 2 น้อย

ในขณะที่รูปของคนที่หนึ่งมีความต่างกันของรูปภาพมาก เนื่องจากมี Euclidean Distance สูง จึงทำให้โปรแกรมจำแนกได้ยาก

A simple face verification system

T8. Write a function that takes in the similarity matrix created from the previous part, and a threshold t as inputs. The outputs of the function are the true positive rate and the false alarm rate of the face verification task (280 Test images, tested on 40 people, a total of 11200 testing per threshold). What is the true positive rate and the false alarm rate for t = 10?

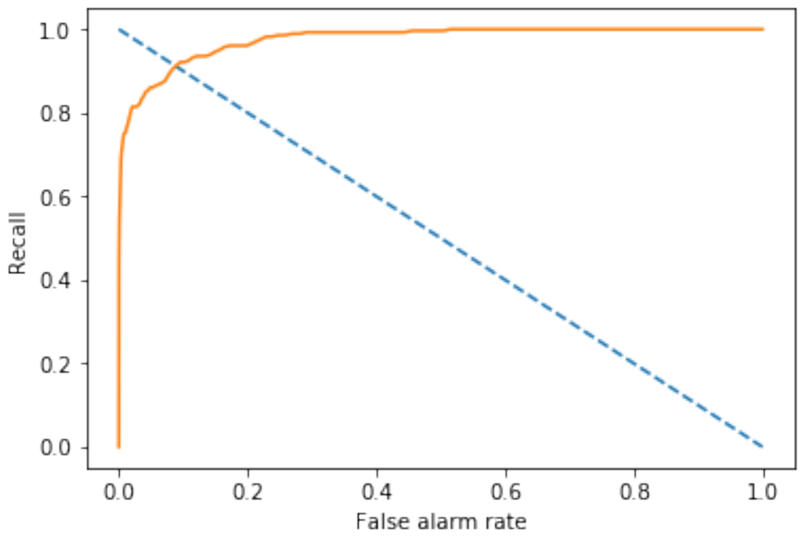
Recall = 0.966

False Positive Rate = 0.456

T9. Plot the RoC curve for this simple verification system. What should be the minimum threshold to generate the RoC curve? What should be the maximum threshold? Your RoC should be generated from at least 1000 threshold levels equally spaced between the minimum and the maximum. (You should write a function for this).

Minimum threshold อยู่ที่ 0 เนื่องจาก Euclidean distance เป็น จำนวนจริงที่ไม่ติดลบ

Maximum threshold อยู่ที่ $\max(similarity\_matrix)$ เพราะ threshold ไม่มีทางเกินค่าสูงสุดใน similarity matrix

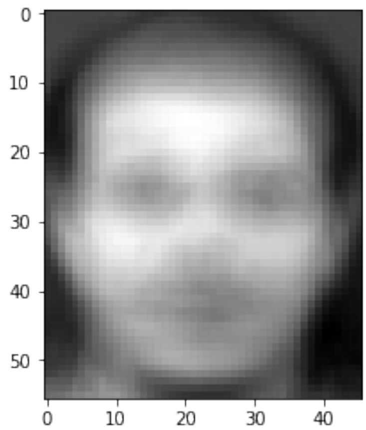


T10. What is the EER (Equal Error Rate)? What is the recall rate at 0.1% false alarm rate? (Write this in the same function as the previous question)

EER = 0.088

ที่ 0.001 FAR จะมี Recall อยู่ที่ประมาณ 0.543

T11. Compute the mean vector from the training images. Show the vector as an image (use numpy.reshape()). This is typically called the meanface (or meanvoice for speech signals).



T12. What is the size of the covariance matrix? What is the rank of the covariance matrix?

d x d เมื่อ d คือมิติของเวกเตอร์ จึงได้ covariance matrix ขนาด 2576 x 2576, Rank 119

T13. What is the size of the Gram matrix? What is the rank of Gram matrix? If we compute the eigenvalues from the Gram matrix, how many non- zero eigenvalues do we expect to get?

N x N เมื่อ N คือจำนวน data sample จึงได้ขนาด 120 x 120, Rank 119

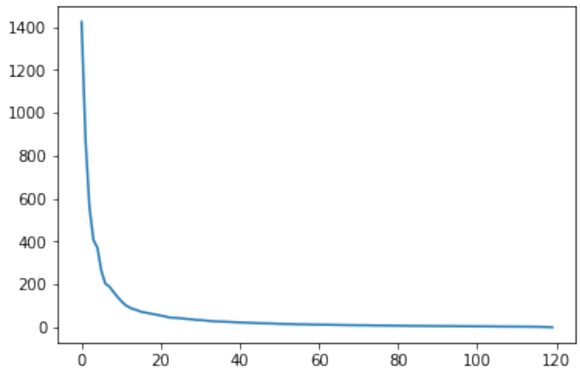
T14. Is the Gram matrix also symmetric? Why?

Symmetric เนื่องจาก Gram matrix คือ $X^T X$ เมื่อใช้สมบัติ $(AB)^T = B^T A^T$ จะได้เป็น $(X^T X)^T = X^T X$ ซึ่งหมายถึง matrix นี้สมมาตร

T15. Compute the eigenvectors and eigenvalues of the Gram matrix, v' and λ. Sort the eigenvalues and eigenvectors in descending order so that the first eigenvalue is the highest, and the first eigenvector corresponds to the best direction. How many non-zero eigenvalues are there? If you see a very small value, it is just numerical error and should be treated as zero.

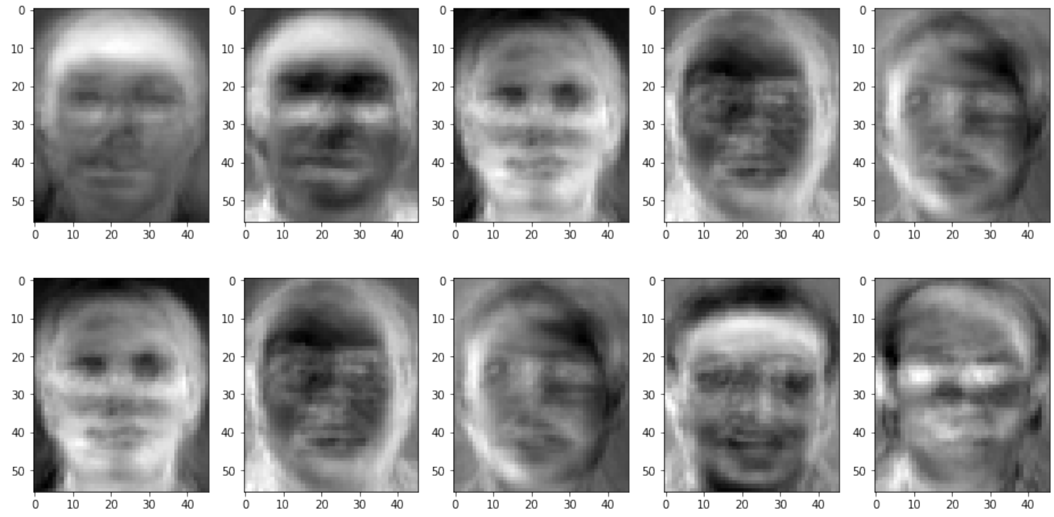
มี Non-zero eigen value อยู่ 119 ตัว

T16. Plot the eigenvalues. Observe how fast the eigenvalues decrease. In class, we learned that the eigenvalues is the size of the variance for each eigenvector direction. If I want to keep 95% of the variance in the data, how many eigenvectors should I use?

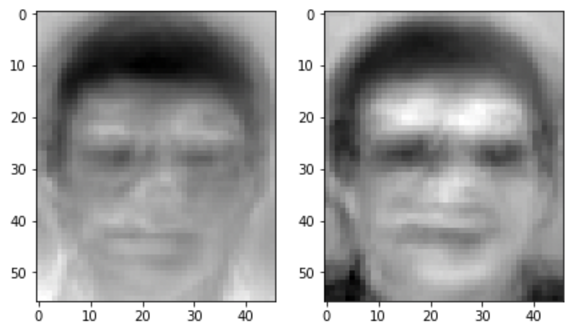


เราจะใช้ eigenvector 63 ตัวเพื่อให้ครอบคลุม 95% ของ variance

T17. Compute v. Don't forget to renormalize so that the norm of each vector is 1 (you can use numpy.linalg.norm). Show the first 10 eigenvectors as images.



T18. From the image, what do you think the first eigenvector captures? What about the second eigenvector? Look at the original images, do you think biggest variance are capture in these two eigenvectors?



รูปแรกจะ capture ในส่วนบริเวณคิ้ว และจมูก ส่วนรูปที่สอง capture ส่วนของใบหน้า ซึ่งมีความชัดเจนกว่าภาพแรก

T19. Find the projection values of all images. Keep the first k = 10 projection values. Repeat the simple face verification system we did earlier using these projected values. What is the EER and the recall rate at 0.1% FAR?

At FAR = 0.001, Recall = 0.514

EER = 0.0787

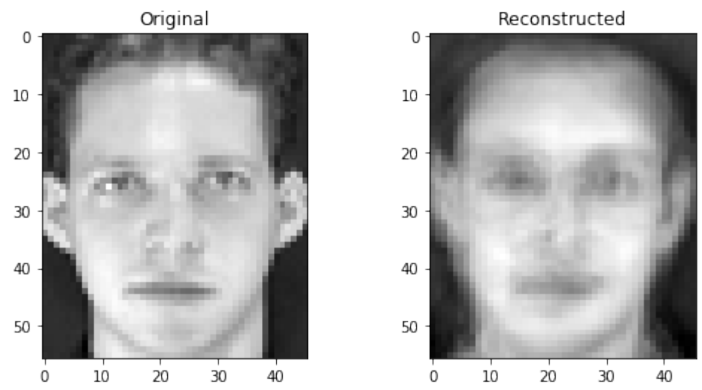
T20. What is the k that gives the best EER? Try k = 5,6,7,8,9,10,11,12,13,14.

K = 11 ได้ EER น้อยที่สุด (ดีที่สุด) ที่ 0.0780

K	Equal Error Rate
5	0.1068
6	0.0935
7	0.0926
8	0.0855
9	0.0811
10	0.0787
11	0.0780
12	0.0853
13	0.0823
14	0.0819

(Optional) PCA reconstruction

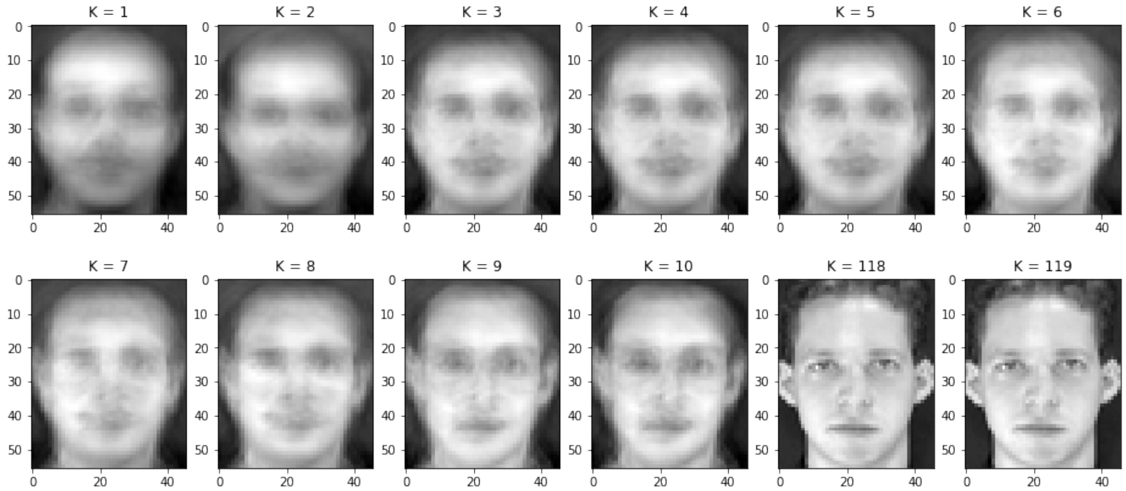
OT1. Reconstruct the first image using this procedure. Use k = 10, what is the MSE?

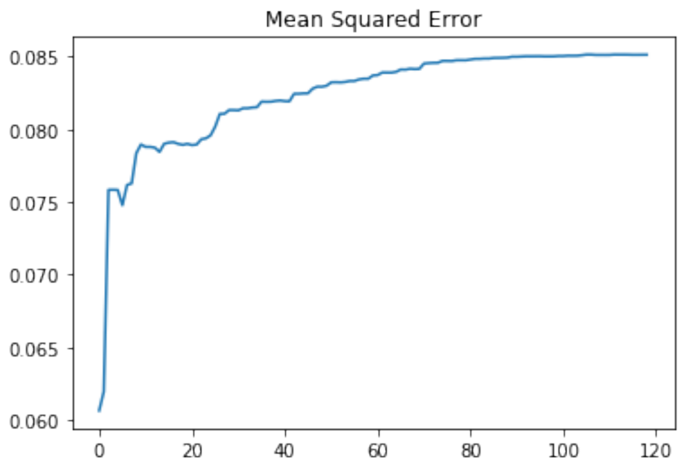


Original Reconstructed (K = 10)

MSE = 0.07896

OT2. For k values of 1,2, 3, ..., 10, 119, show the reconstructed images. Plot the MSE values.





OT3. Consider if we want to store 1,000,000 images of this type. How much space do we need? If we would like to compress the database by using the first 10 eigenvalues, how much space do we need? (Assume we keep the projection values, the eigenfaces, and the meanface as 32bit floats)

กำหนดให้รูปภาพมี $N = 10^6$ รูป แต่ละรูปขนาด $h \times w = 2576$ pixel และเราเลือก Eigenvalue มา $\mathrm{k} = 10$ ตัว และให้แต่ละค่าเก็บ $x = 4$ bytes

<u>Original</u>

จะต้องเก็บทั้งหมด $Nwhx = 10^6 \times 2576 \times 4 \approx 1.03 \times 10^{10}\ bytes \approx 9.6\ GB$

<u>Compressed</u>

เราต้องเก็บ Projection values $\vec{p}$ ทั้งหมด $Nk = 10^6 \times 10 = 10^7$ ค่า

Eigenfaces V ทั้งหมด $whk = 2576 \times 10 = 25760$ ค่า

Meanface $\overrightarrow{\mu_x}$ ทั้งหมด $wh = 2576$ ค่า

รวมแล้วต้องเก็บทั้งหมด $(10^7 + 25760 + 2576) \times 4 \approx 4.01 \times 10^7\ bytes \approx 38.3\ MB$

Linear Discriminant Analysis (LDA)

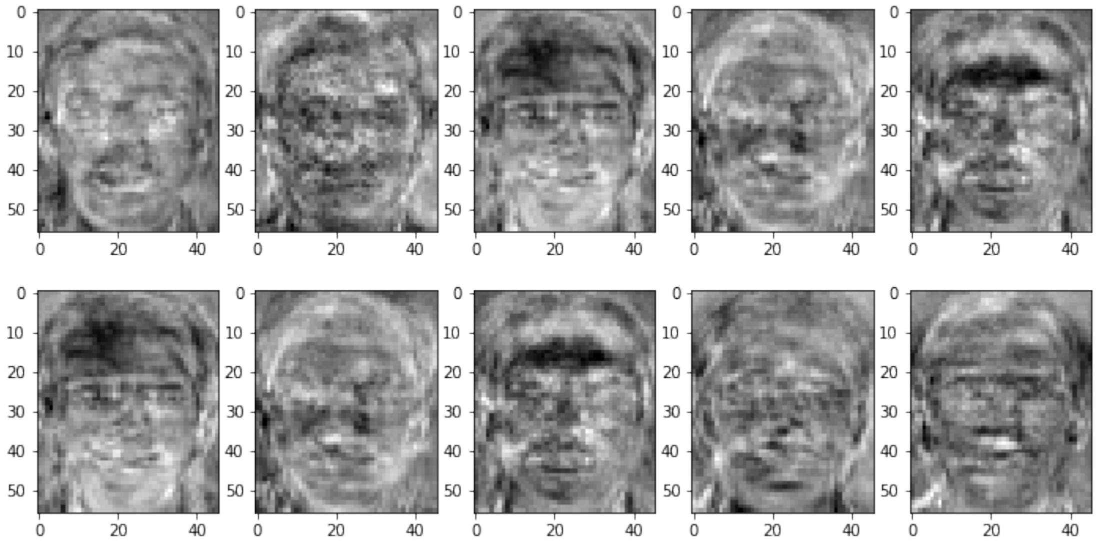
T21. In order to assure that S_W is invertible we need to make sure that S_W is full rank. How many PCA dimensions do we need to keep in order for S_W to be full rank? (Hint: How many dimensions does S_W have? In order to be of full rank, you need to have the same number of linearly independent factors)

จำนวน dimension จะต้องน้อยกว่าจำนวน data point จึงจะทำให้ S_w invertible

T22. Using the answer to the previous question, project the original input to the PCA subspace. Find the LDA projections. To find the inverse, use numpy.linalg.inv. Is $S_w^{-1}S_B$ symmetric? Can we still use numpy.linalg.eigh? How many non-zero eigenvalues are there?

$S_w^{-1}S_B$ ไม่สมมาตร แต่ยังสามารถใช้ numpy.linalg.eigh ได้โดยไม่ Error และพบว่าผลลัพธ์มี 120 non-zero eigenvalues

T23. Plot the first 10 LDA eigenvectors as images (the 10 best projections). Note that in this setup, you need to convert back to the original image space by using the PCA projection. The LDA eigenvectors can be considered as a linear combination of eigenfaces. Compare the LDA projections with the PCA projections.

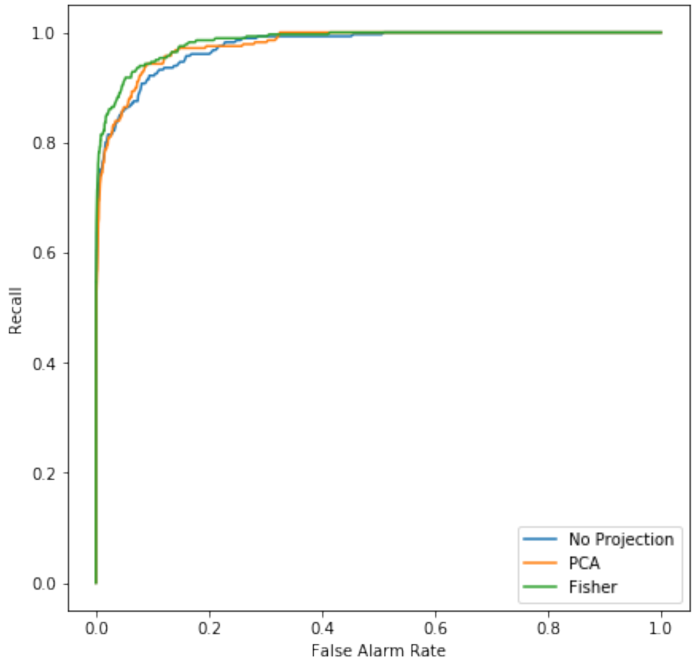


T24. The combined PCA+LDA projection procedure is called fisherface. Calculate the fisherfaces projection of all images. Do the simple face verification experiment using fisherfaces. What is the EER and recall rate at 0.1% FAR?

At FAR = 0.001, Recall = 0.643

EER = 0.07083

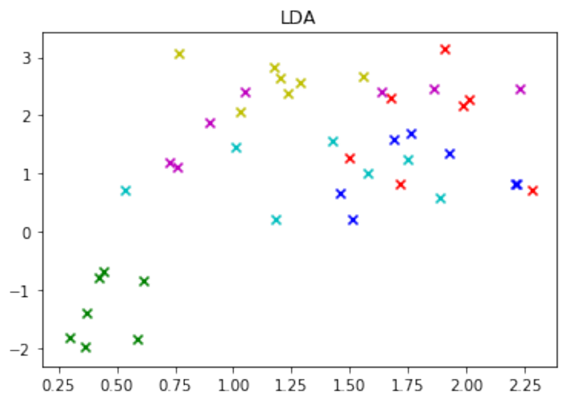
T25. Plot the RoC of all three experiments (No projection, PCA, and Fisher) on the same axes. Compare and contrast the three results. Submit your writeup and code on MyCourseVille.

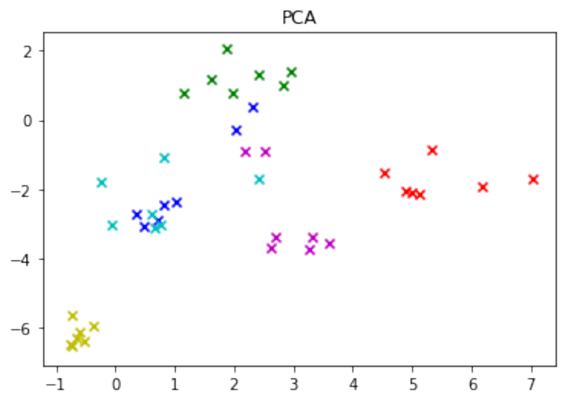


ทั้งสามวิธีให้ผลลัพธ์ไม่ต่างกันมาก แต่ Fisherface method ให้ RoC ดีที่สุดในทั้งสามวิธีเนื่องจากเป็นการผสมระหว่าง PCA และ LDA รองลงมาคือ PCA และที่แย่ที่สุดคือ No Projection และยังประมวลผลช้าอีกด้วย

https://colab.research.google.com/drive/16vzq5MXcswaRc50e4EXzCo_aLA4v_NEF

OT4. Plot the first two LDA dimensions of the test images from different people (6 people 7 images each). Use a different color for each person. Observe the clustering of between each person. Repeat the same steps for the PCA projections. Does it come out as expected?





LDA ไม่ได้แยกแต่ละ class ได้ดีตามที่คาดไว้