Introduction to Machine Learning (CS1390/PHY1390)
Monsoon 2021 - Quiz 1

Jenish Raj Bajracharya

Collaborators: None

---

1 Introduction

In this assignment, face recognition was implemented with Sklearn's Olivetti-faces data-set using Singular value Decomposition(SVD). A brief summary for facial recognition is as follows:

- A matrix of test images and training images were obtained.
- Both the matrices were normalized by subtracting the mean.
- The training images were decomposed using the SVD into eigenfaces(eignevectors).
- Adequate number of principle components(k) were determined amongst the largest eigenvalues.
- Each training image was then represented as a linear combination of the k-eigenfaces (eigenvectors).
- The testing images were then projected on the the eigenspace and the minimum distance of the testing image with a training image we obtained.
- The least distant image was recognized and compared with the correct label($f(x)$).
- If $h(x) = f(x)$, then a correct face was recognized.

2 Exploring the dataset

The Olivetti-faces data-set contains a set of facial images taken at AT&T Laboratories during the early 1900. The dataset consist of 10 images of 40 distinct people (figure:1) resulting in a total of 400 images. The 10 facial images have different facial expressions and details. For example, open or closed eyes , smiling or not smiling, glasses or no glasses (figure:2). All images are nose-centered, but some images are taken at a different angle while some have different lighting conditions. Each of the 40 subjects are labelled by an integer in range [0-39] .

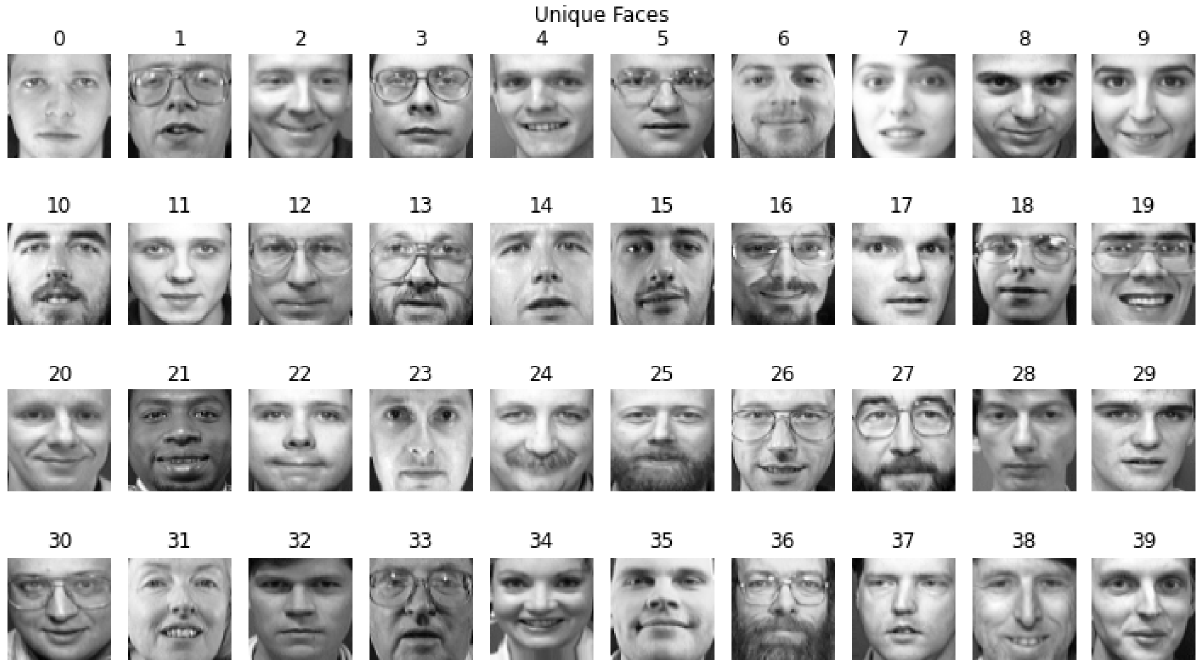

Figure 1: Faces of 40 subjects

Figure 2: Facial expressions and details of a subject

2.1 Cleaning up the dataset

Not much cleaning was necessary because all the images were of the same dimensions. The dataset was only split into training and testing set.

2.2 Effect on classification

Similar Faces
Some subjects in the dataset are quite similar to the other subjects. For example subjects 27, 36, 13, are quite similar because of the beard and glasses. Also subjects 20 and 4 are similar to each other. The similarity of subjects might result in mis-classification.

Image Angle
The angel at which the image is taken also plays a major role in mis-classification. For example, a right oriented image may be mis-classified with another similar right oriented image because, there exist white spaces in the image matrix.

Lighting Condition

Lighting conditions play a major role in classification. A dark image might result in a dense matrix i.e pixels with higher value. On the other hand bright images will be less dense. Since we are comparing the image vectors, there is a high chance of mis-classification.
Scale and background(For images in general)
The performance decreases quickly with changes to head size. For example, an image with background will be difficult to classify because a lot of information in the image matrix is insignificant. While a nose center face will have a lot of significant information.
For this dataset the scale effect will be less significant because 1)images do not have backgrounds and 2) images are nose centered.

2.3 Splitting the dataset

The 400 images were split at random into 70% training and 30% testing set. Thus, for each subject 7 images will be used for training while 3 images will be used for testing.

```
Total Training Images: 280                Total Testing Images: 120
Shape of Training matrix: (4096, 280)     Shape of Testing matrix: (4096, 120)
```

3 Facial Recognition and SVD

Facial recognition using the Singular Value Decomposition follows the following flowchart.

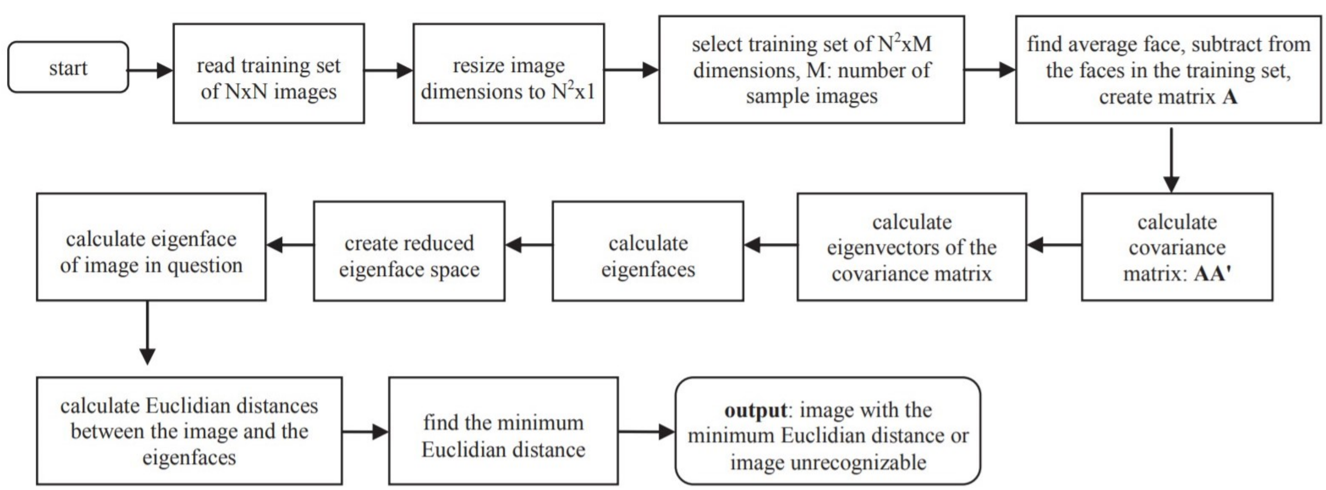

Figure 3: Facial recognition using SVD [1]

3.1 Reading training set and converting to a training matrix

3.1.1 Theoretical Foundation

Each image I_i is a $N \times N$ matrix. We first convert the $N \times N$ image matrix to a column vector of size $N^2 \times 1$ where,

$$I_i = \begin{bmatrix} i_i^1 \\ i_i^2 \\ \vdots \\ i_i^{n \times n} \end{bmatrix}$$

We repeat the same for all the images and obtain a training matrix (A) where,

$$A = [I_1, I_2, ..., I_M] \text{ for } 1 \leq i \leq M$$

Thus, size of A is $N^2 \times M$

3.1.2 Implementation

For the Olivetti dataset, each image I_i is represented as a 4096×1 column vector ($n^2 = 4096$). Our training set has a total of $m = 280$ images, so the training matrix A is of size 4096×280

3.2 Average Face and Normalization

We compute the average face vector as a sum of each image vector (I_i) divided by M i.e.

$$\mu = \frac{1}{m} \sum_{i-1}^{m} I_i$$

Since, μ is the average of all the faces, reshaping the column vector results in a mean face (fig:4).
We normalize each of the training and testing set by subtracting the mean vector from each image vector such that,

$$\text{normal\_matrix} = I_i - \mu$$

where, $I_i \in$ [training_matrix, testing_matrix]. The size of the normal matrix is same as the size of its parent matrix.

4 Co-variance Matrix and SVD

4.1 Theoretical Foundation

Definition: Co-variance matrix
For a matrix $X \in \mathbb{R}^{m \times n}$, co-variance matrix ($C$) is defined as

$$C = XX^T = \frac{1}{M} \sum_{n=1}^{M} \phi_n \phi_n^T$$

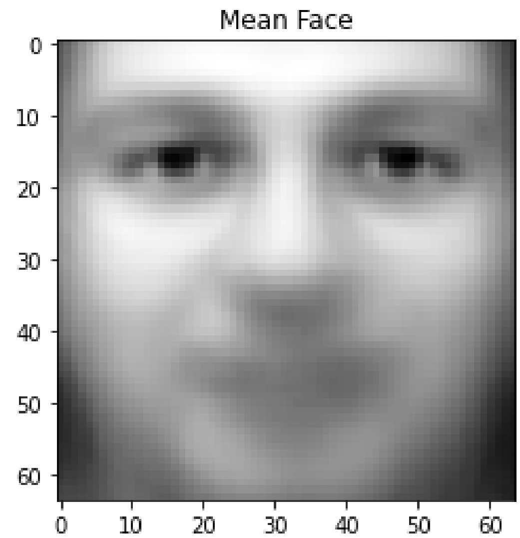

Figure 4: Mean Face

where $\phi_n \in X$ and $C \in \mathbb{R}^{m \times m}$.
Note: A co-variance matrix is symmetric.

Theorem: Diagonalization of co-variance matrix
A co-variance matrix $C \in \mathbb{R}^{m \times m}$ is always diagonalizable such that

$$C = EDE^{-1}$$

where, D is the diagonal matrix of eigenvalues and E is the eigenvectors of C.

Theorem: Singular Value Decomposition (SVD)
For any matrix $A \in \mathbb{R}^{m \times n}$, A can be uniquely represented such that:

$$A = U\Sigma V^T$$

where, $U \in \mathbb{R}^{m \times m}$ is a orthonormal matrix,
$V \in \mathbb{R}^{n \times n}$ is a orthonormal matrix
$\sum =$ diag $(\sigma_1 \dots \sigma_n) \in \mathbb{R}^{m \times n}$ is a diagonal matrix such that $\sigma_1 \geq \dots \geq \sigma_n \geq 0$

Lemma
For any matrix $A \in \mathbb{R}^{m \times n}$,

$$AA^T = U\Sigma^2 U^T$$

where U, Σ comes from SVD.
Proof: Multiplying the SVD expression by A^T proves the lemma.

Combination of SVD and diagonalization
For a co-variance matrix $C \in \mathbb{R}^{m \times m}$,

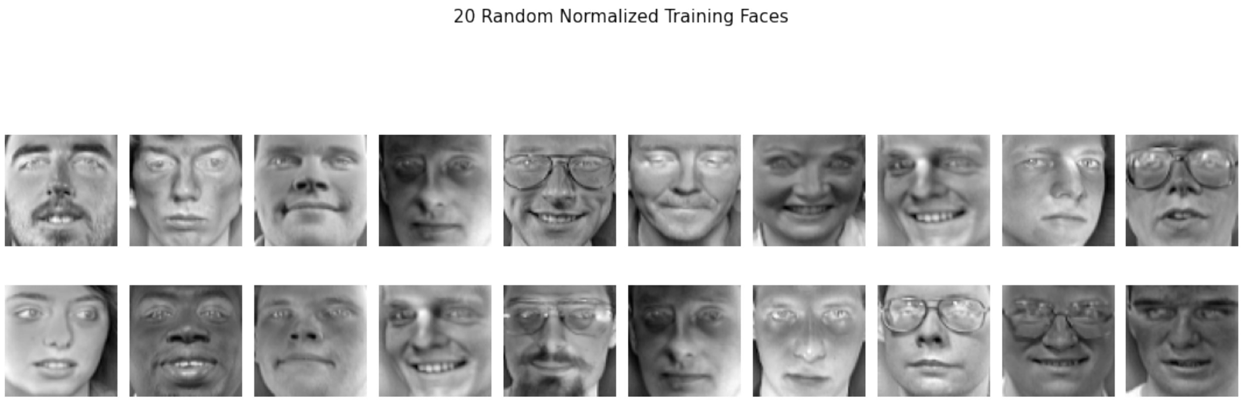

Figure 5: Training Set

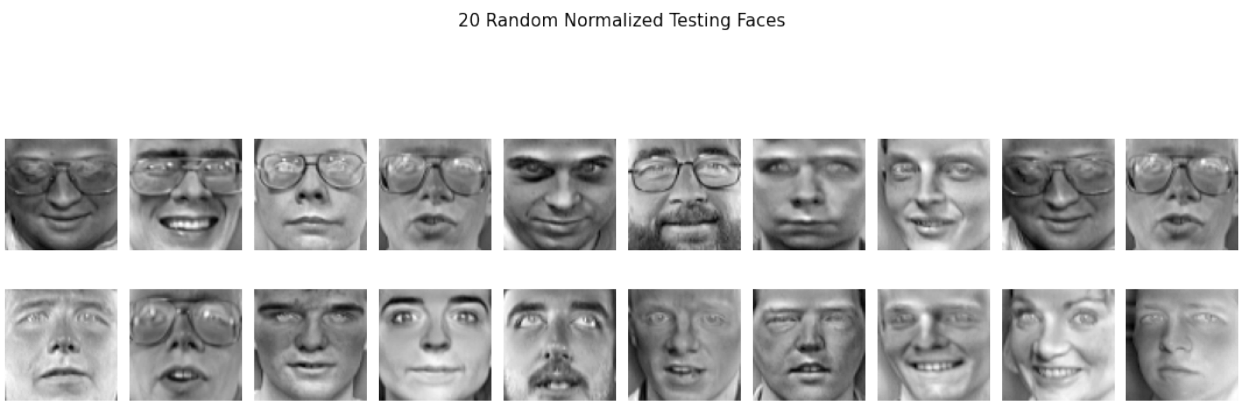

Figure 6: Testing Set

Figure 7: 20 Random Normalized faces

$$C = EDE^T = U\Sigma^2U^T$$

where, $D = \Sigma^2$ and $E = U$

Note: The eignevectors of the co-variance matrix represents the eigenfaces of the training matrix and are the principle components of the eigenspace.

4.2 Computing the eigenvectors of co-variance matrix

The matrix AA^T is a $N^2 \times N^2$ matrix, hence computing its eigenvectors are impractical. So we consider $A^TA(M \times M)$ matrix resulting in M eigenvalues and eigenvectors.
Let v_i be the eigenvectors of A^TA then, the eigenvectors(u_i) of AA^T is related by the equation:

$$u_i = Av_i$$

Thus, we calculate the M eigenpairs of A^TA and then compute u_i from the above relation.
Proof

Let C be the co-variance matrix such that $C = AA^T$.
Let there exist eigenvectors v_i of $A^T A$, then by definition:

$$
\begin{aligned}
&A^T A v_i = \sigma_i v_i \\
&\implies AA^T A v_i = \sigma_i A v_i \\
&\implies C A v_i = \sigma_i A v_i \\
&\implies C u_i = \sigma_i u_i \text{ where } , u_i = A v_i
\end{aligned}
$$

Thus, $u_i = Av_i$.
Computing the co-variance matrix in our image dataset results in the following heat-map of $A^T A$ (fig:11). Note that the diagonal of the heat-map are the eigenvalues. Note: From the figure we observe that, there is very less co variance between the data points as off-diagonal entries are close to 0.

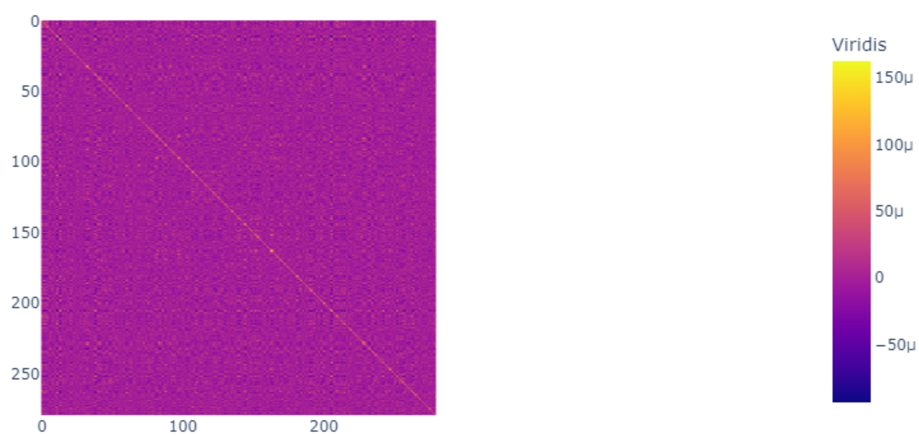

Figure 8: Heat-map of Co-variance Matrix

5 Principle Component Analysis (Choosing k)

Now that we have decompose our matrix into linearly independent eigenvectors. The eigenvectors can span the entire space, thus the eigenvectors are the principle components.
Note: The eigenvalues of the principle component i.e the diagonal entries become smaller and smaller, thus the variation decreases. We can choose eignevectors with high eigenvalues to represent our data. But, what number of eigenvalue-vector pair(k) contribute to the PC? We plot the cumulative sum of eigenvalues. The graph then gives us a good idea about the value of k to be chosen.
From the graph (fig: 9) $k = 135$ was chosen.

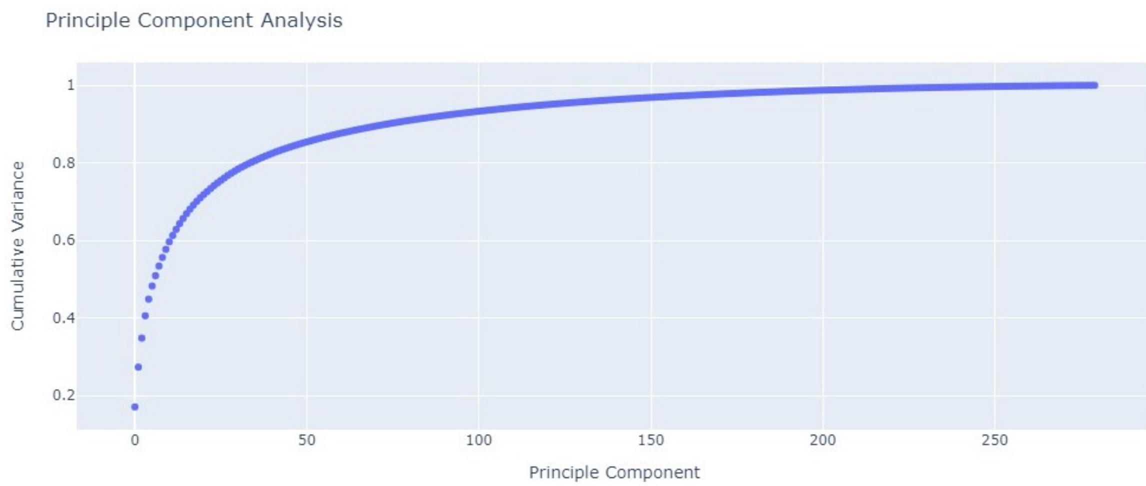

Figure 9: Finding k

6 Eigenfaces

Given the value of k, we compute the eigenfaces $u_i = Av_i$

Figure 10: 20 Random Eignefaces

7 Representing the normalized training faces onto the basis

The normal training faces can be represented as a linear combination of the k eigenvectors such that:

$$\text{normal\_training\_matrix} = \sum_{i=1}^{K} w_i u_i$$

where, $w_i = u_i^T(I_i - \mu)$.
Each normalized faces can be represented by the following basis.

$$W = \begin{bmatrix} w_1^i \\ w_2^i \\ \vdots \\ w_k^i \end{bmatrix}$$

where, $i \in [1, M]$.

8 Recognizing Unknown faces

For each unknown faces(U_i) we convert, the image vector into a normal vector by subtracting the mean. We then project the normal image into our eigenspace using the following operation,

$$W_u = u_k^T(U_i - \mu)$$

. We then compare the unknown weight with each $w_i \in W$ bu computing the Euclidean distance such that

$$e = min||W_u - W||$$

If e is within the tolerance threshold, we index I_i as the recognized face from the training matrix.

Note: The notion of training and testing error
The training error is always 0 because, each training image can be expressed as the linear combination of the basis vector so the distance is always 0.

9 Results

With $k = 135$, the SVD image classifier resulted in a accuracy score of 81.67%. The confusion matrix below show the classification of our testing images with the training images. The diagonal entries of the confusion matrix depicts correct classification, while the off diagonal entries show mis-classification. From the diagonal matrix, 21 images out of 120 were mis-classified resulting in 81%accuracy.

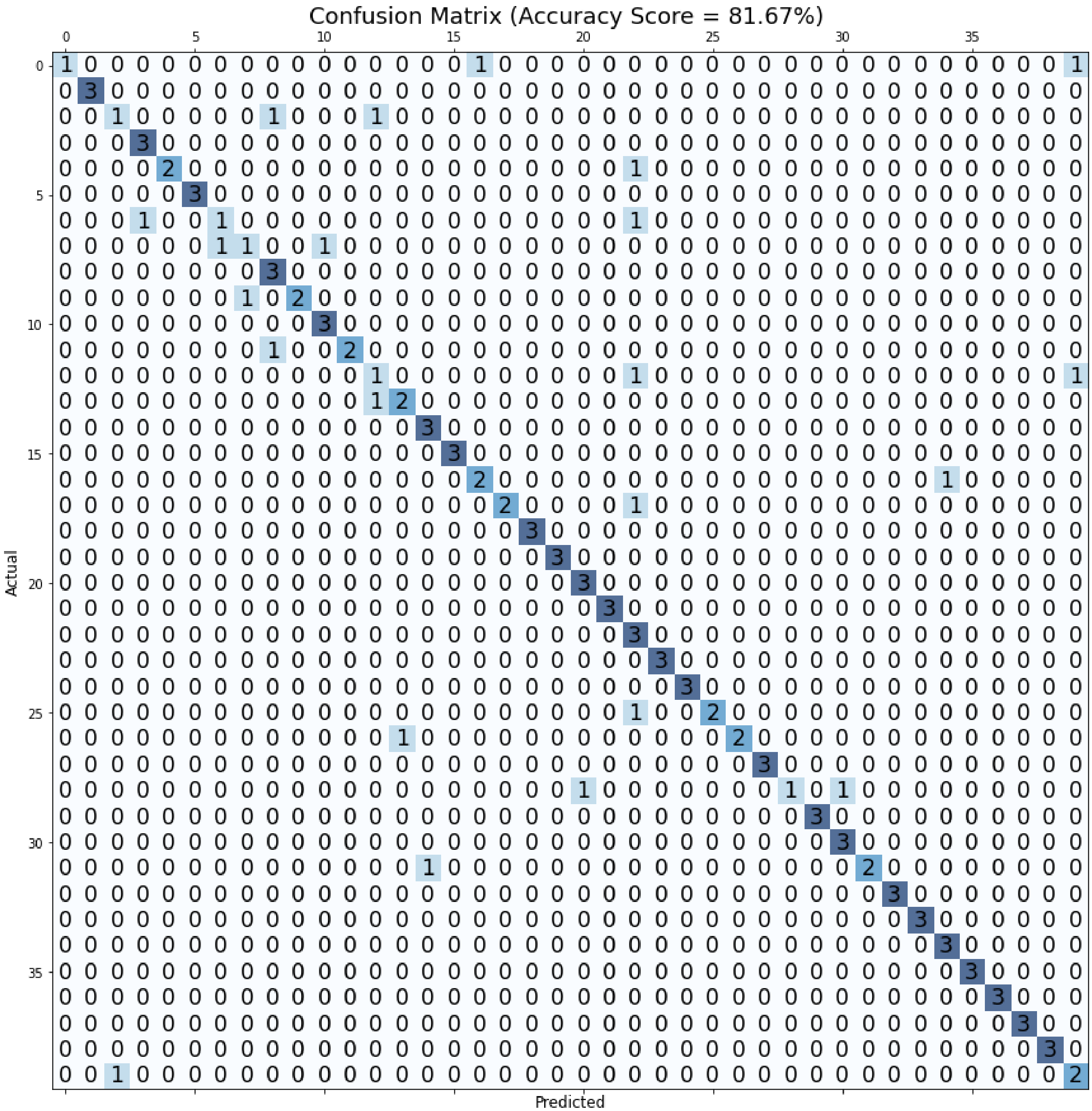

Figure 11: Confusion Matrix

9.1 Correct Classification

Figure 12

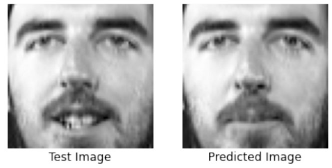

Figure 13

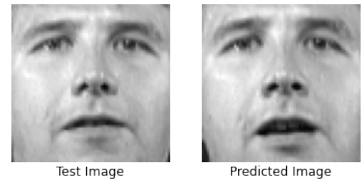

Figure 14

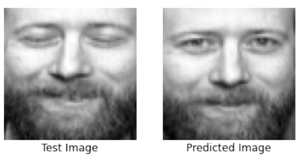

Figure 15

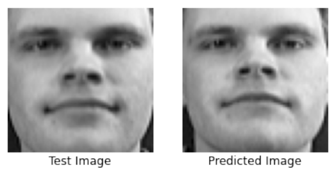

Figure 16

Figure 17

Figure 18

Figure 19

Figure 20

From the above classification, we can observe that our classifier does a good job to identify faces that have similar facial expressions and similar lighting conditions i.e not very bright.

9.2 Mis-classification

Figure 21

Figure 22

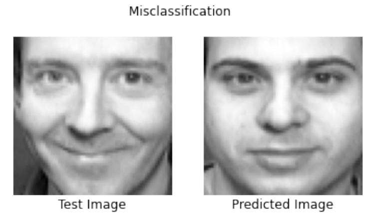

Figure 23

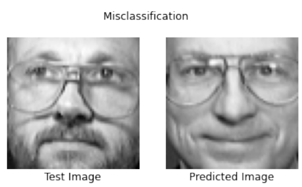

Figure 24

Figure 25

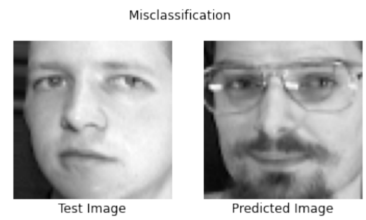

Figure 26

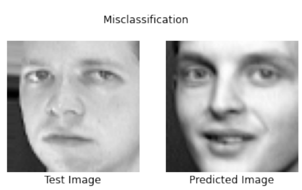

Figure 27

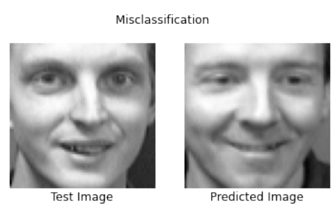

Figure 28

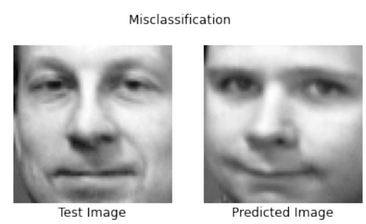

Figure 29

- Figure 22, 23, 26,27, 28 are mis-classified because of the orientation issue as described previously.
- Figures 25 has been mis-classified possibly because of the bright lighting condition.
- Figures 24 has been mis-classified because the two subjects are quite similar to each other.

9.3 Improvement

Images have been mis-classified because of the orientation, lighting difference and similarity of faces. Thus, we can improve the dataset by only taking face centered zoomed in picture of the subjects.

9.4 Conclusion

In conclusion, the facial recognition system using the SVD has many limitations because of the orientation of images, lighting. The only way it can correctly classify is having nose centered images. Since, in real world images comes in all shape and sized, the SVD may not be a good hypothesis class.

References

[1] Çarıkçı, M., amp; Özen, F. (2012). A face recognition system based on eigenfaces method. Procedia Technology, 1, 118–123. https://doi.org/10.1016/j.protcy.2012.02.023