Q2:

(1) Recognition accuracies by varying the M_pca and M_lda:

M_pca	M_lda	Recognition accuracy (%)
364	2	12.5
364	10	40.38
364	20	67.31
364	35	75.96
364	45	77.88
364	51	78.85
200	2	13.46
200	10	76.92
200	20	81.73
200	35	83.65
200	45	83.65
200	51	84.61
100	2	19.23
100	10	82.69
100	20	83.65
100	35	83.65
100	45	81.73
100	51	81.73
70	2	29.81
70	10	78.85
70	**20**	**88.46**
70	35	84.62
70	45	78.85
70	51	76.92
55	2	26.92
55	10	78.85
55	20	86.54
55	35	79.81
55	45	75.0
55	51	74.04

Discuss:

- In general, we can observe that M_pca = 364 give the worst recognition accuracy. This fact implies that keeping more information does not mean the recognition accuracy would be improved. We can explain this phenomenon as following: When we use M_pca = 364 to project our data to the 364-dimensional space, it would help to keep more information of our original data than when we project our data to lower dimension space such as M_pca = 100, M_pca = 70, M_pca = 55, however maybe many information are not necessary for the recognition task, and keeping them would lower the accuracy of our model. It maybe the reason why using M_pca = 364 could keep more information of our original data but would lead to lower recognition accuracy. At the same time,

when we use M_pca = 100, M_pca= 70 or M_pca=55, then only the most important information are kept, and they help to increase the recognition accuracy.
- Now, we consider the dependency of recognition accuracy on M_lda. From the above table, we can observe that this dependency would be different depending on our choice of M_pca. In fact, with M_pca = 364 or M_pca = 200, when we increase M_lda from 2 to 10, 10 to 20, 20 to 35, 35 to 45 and 45 to 51, the recognition accuracy would increase. However, when we choose M_pca = 100, or 70, or 55, then the recognition accuracy seems to be highest when M_lda = 20 and if we continue to increase the value of M_pca from 20 to 35, 45 and 51, the recognition accuracy would decrease.
- When we increase the value M_lda from 2 to 10, there would be a big increase in the recognition accuracy, but when we increase M_lda from 20 to 35, to 45, to 51, the change in accuracy is small. Thus, it implies that the 10 largest eigenvectors obtained from the LDA process are super important for the face recognition task.
- The best recognition accuracy obtained from these above statistics is 88.46%, when we choose M_pca = 70, M_lda = 20.

(2) Now, we compare the recognition of PCA and PCA-LDA for some choices of M_pca:

	Recognition accuracy (%)	
	PCA	PCA-LDA
M_pca = 364	67.31	78.85
M_pca = 200	67.31	84.61
M_pca = 100	67.31	83.65
M_pca = 70	66.35	88.46
M_pca = 55	65.38	86.54

In each row, considering the recognition accuracy of PCA-LDA, I choose the highest accuracy when varying M_lda, for example, for the row M_pca = 364, the recognition accuracy of PCA-LDA is 78.85 when M_lda = 51, and for the row M_pca = 70, the accuracy of PCA-LDA is 88.46 when M_lda = 20.

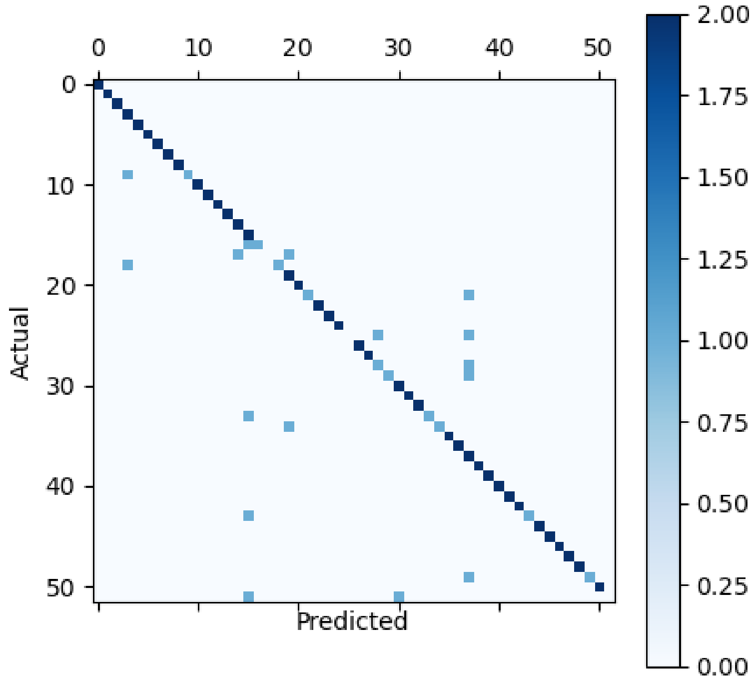
Discuss:
From this table, we can see that by combining PCA and LDA, we can improve the recognition accuracy by about 11-12% than using PCA alone.

(3) Rank of scatter matrices:
In PCA-LDA method, from the formula for PCA-LDA, the rank of within-class scatter matrix S_W and between-class scatter matrix S_B only depend on M_pca. By implementation, I can observe that :
+ With M_pca >= 51, then rank(S_B) = 51, rank(S_W) = M_pca
+ With M_pca <= 50, then rank(S_B) = M_pca and rank(S_W) = M_pca

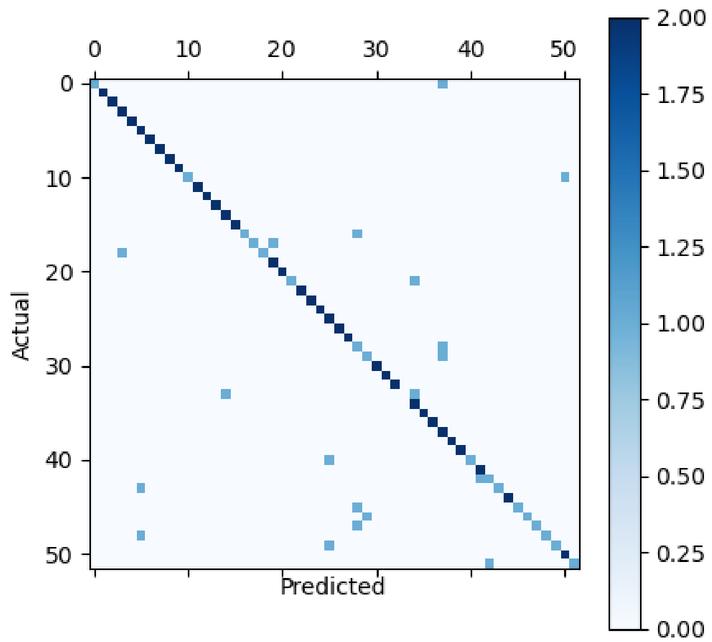
(4) Confusion matrix:

PCA-LDA:
Some example confusion matrix:

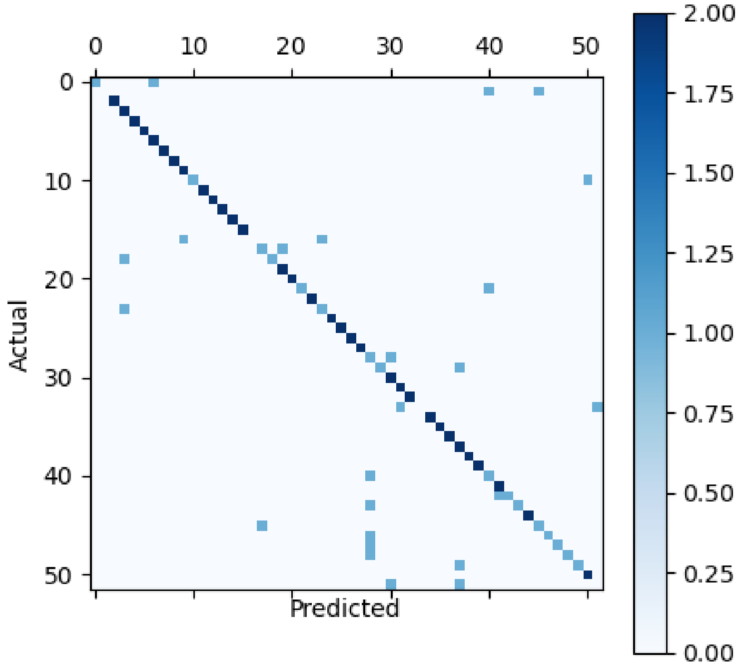
M_pca = 200, M_lda = 51:



M_lda = 100, M_lda = 51 :



M_lda = 70, M_lda = 51 :



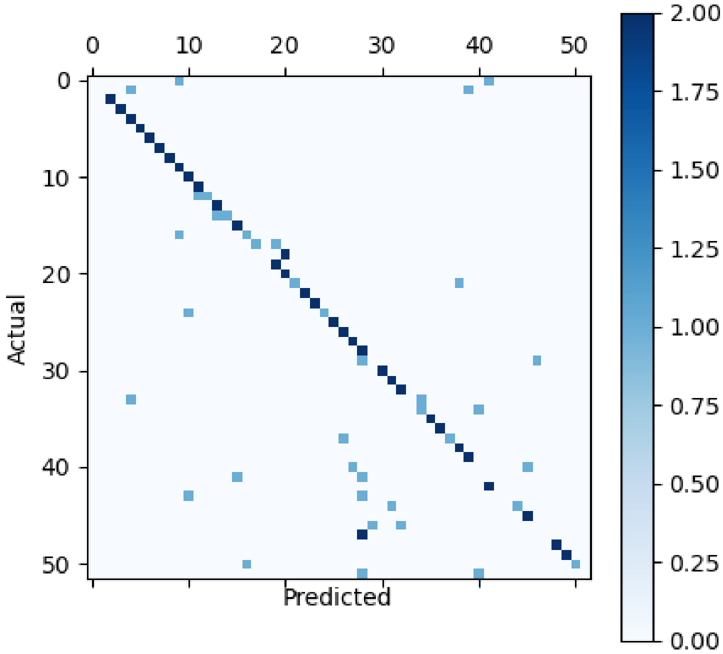
- **Example for success and failure cases:**

We consider when M_lda = 70 and M_lda = 51. We have faces belonging to 52 classes : 0,1,2,… 51 and in the test set, there are 2 faces for each class.
Example for success case of prediction : 2 faces with of class 12, 15, 20,... are predicted correctly
Example for partial success of prediction: among 2 faces of label 18, one face is predicted correctly and the remaining face is predict wrongly.
Example for failure case of prediction: both 2 faces of label 1, 16, 33 are predicted wrongly.

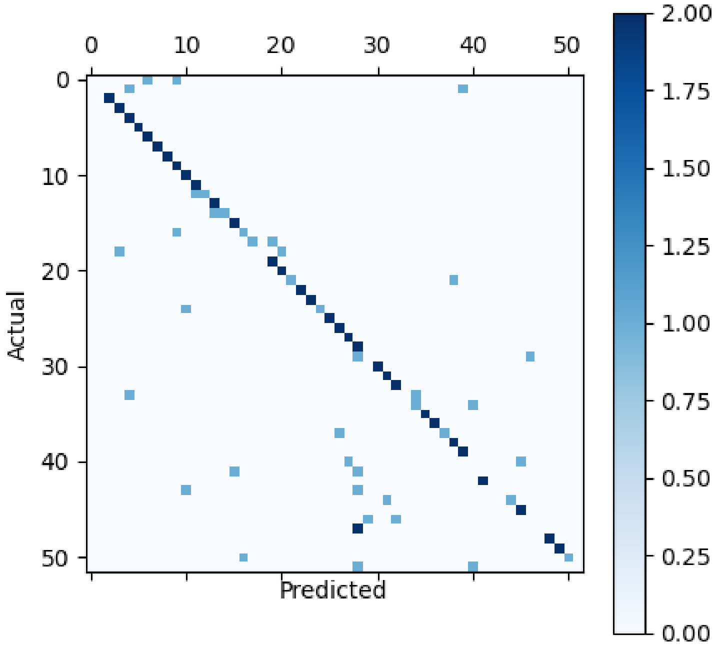
PCA:

Some example confusion matrix:

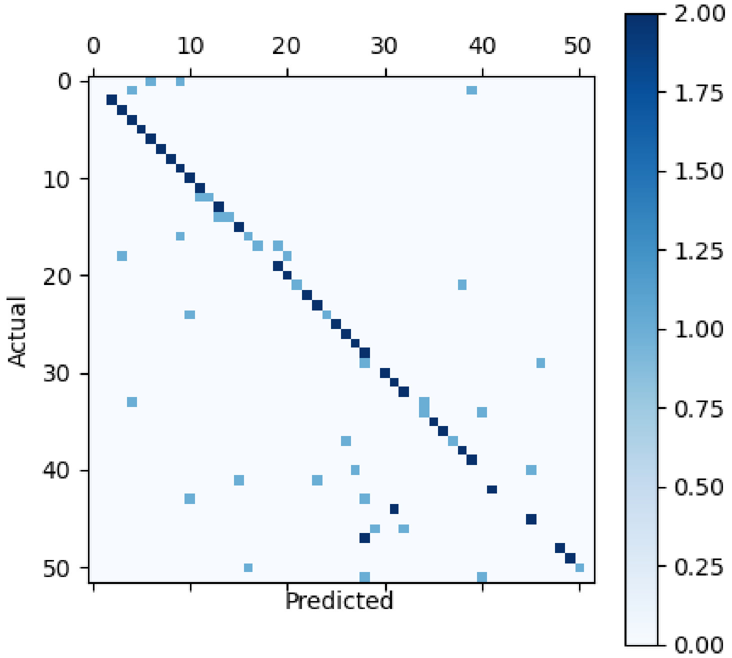
M_pca = 200:



M_pca = 100:



M_pca = 70:



- **Example for success and failure cases:**

We consider when M_lda = 70. We have faces belonging to 52 classes : 0,1,2,… 51 , and in the test set, there are 2 faces for each class.
Example for success case of prediction : 2 faces with of class 6,8,11,... are predicted correctly
Example for partial success of prediction: among 2 faces of label 24 , one face is predicted correctly and the other face is predict wrongly
Example for failure case of prediction: both 2 faces of label 18, 40 are predicted wrongly.

(5) Time/ Memory:
With PCA method, for M_pca = 70, the whole process of computing eigenvalues, eigenvectors of covariance matrix S cost 19.3758 seconds and process of predicting the test dataset cost total 0.6782 seconds. In total, it costs 20.054 seconds

With PCA-LDA method, for M_pca = 70, and M_lda = 51, the process of computing eigenvalues and eigenvectors of covariance matrix S for PCA cost 19.686 seconds, the process of computing eigenvalues and eigenvectors of scatter matrix S_B,S_W for LDA cost 0.0296 seconds and the time to predict the whole test set costs 0.676 seconds. In total it costs 20.3916 seconds.

Discuss:
From these above statistics, we can see that in both method, almost of the time is cost by the process of computing eigenvalues and eigenvectors of covariance matrix for PCA. The only difference between the PCA-LDA method and the PCA method is that after using PCA to project data to lower dimension, the PCA-LDA method also apply LDA to the obtained data from previous step. However, as we can see above, the process of computing eigenvalues and eigenvectors of scatter matrix S_B,S_W for LDA cost only a very small amount of time, it’s the reason why the time to execute PCA-LDA method is just a slightly larger than time to execute PCA method (from the above example, time to execute PCA-LDA is 20.3916 seconds and time to execute PCA is

20.054 seconds). This is really meaningful because PCA-LDA just cost a slightly more amount of time to run but it gives a much better recognition accuracy than using PCA alone.