PROJECT DESCRIPTION

This project requires to perform Data Clustering using K-means algorithm.

First ran K-means and then Hungarian Algorithm to get optimal matching. The confusion matrix is a bipartite graph. The Hungarian algorithm is used to permute columns of the confusion matrix to obtain the optimal matching. Reordering of the Column-Index is done and the sum of the diagonal elements/N gives the Accuracy.

TASK 1

Run k-means on **AT&T 100 images (10 class images), for k=10.** Obtain confusion matrix. Re-order the confusion matrix and obtain accuracy.

```
Enter 0 to specify number of classes else press 1 to use the whole data: 0

Enter the number of classes that you want to consider: 10

Enter the 1 class you want to consider: 1

Enter the 2 class you want to consider: 4

Enter the 3 class you want to consider: 6

Enter the 4 class you want to consider: 7

Enter the 5 class you want to consider: 8

Enter the 6 class you want to consider: 9

Enter the 7 class you want to consider: 2

Enter the 8 class you want to consider: 3

Enter the 9 class you want to consider: 5

Enter the 10 class you want to consider: 10

Enter the number of training elements: 7

Enter a value for k: 10
ACCURACY
0.9333333333333333
```

TASK 2

Run k-means on **AT&T 400 images (40 class images), for k=40.** Obtain confusion matrix. Re-order the confusion matrix and obtain accuracy.

```
Enter 0 to specify number of classes else press 1 to use the whole data: 1

Enter the number of training elements: 7

Enter a value for k: 40
ACCURACY
0.7083333333333334
```

TASK 3

Run k-means on **Hand-written-letters data, for k=26.** Obtain confusion matrix. Re-order the confusion matrix and obtain accuracy.

```
Enter 0 to specify number of classes else press 1 to use the whole data: 0

Enter the number of classes that you want to consider: 10
Enter the elements in order "Write in Uppercase only":
A
S
D
F
G
H
J
K
L
C

Enter the number of training elements: 20

Enter a value for k: 26
ACCURACY
0.4342105263157895
```

K-means on HandWrittenLetters for the complete dataset with K=26

```
Enter 0 to specify number of classes else press 1 to use the whole data: 1
Enter the number of training elements: 35
Enter a value for k: 26
ACCURACY
0.4375
```