S1443483

INF2B-CW2 REPORT 1

Task 1

1.1

I wrote the code for calculating the eigenvectors and eigenvalues of a given N-by-D data matrix.

1.2

Here I am giving the demanded values:

Eigenvalue 1 = 14.503958799281799

Eigenvalue 2 = 11.260690043186028

E2(first five rows) =

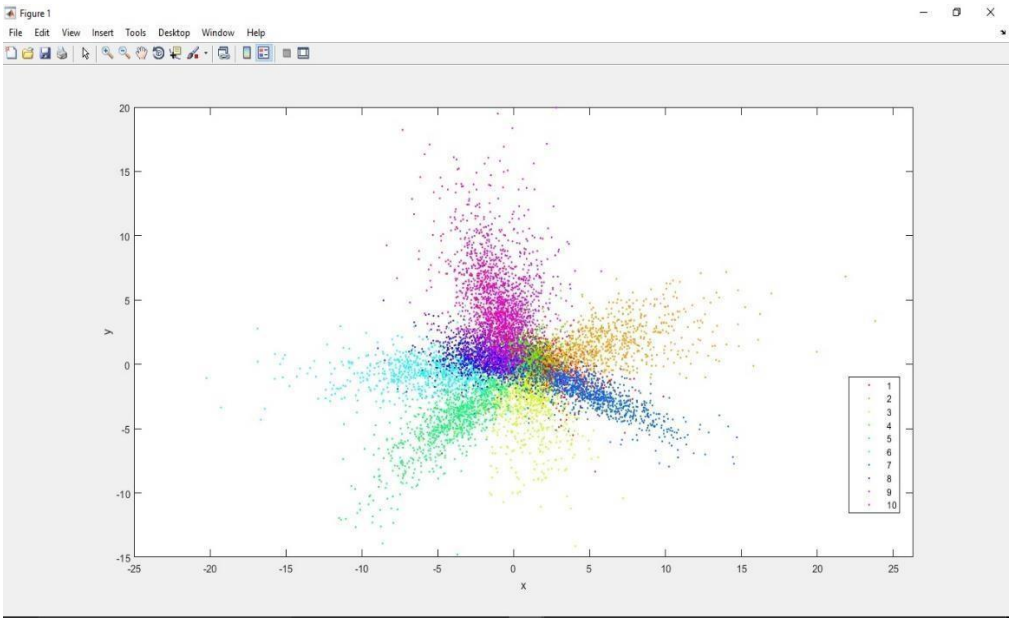
0.106333697496870	0.049928077999992
0.125239559042375	-0.033507882653368
-0.045401410027655	-0.018513207294231
0.126084328165988	0.079237068587066
-0.181511211826110	-0.002146601882309

Xpca(first five rows)=

0.508976013517021	-0.550463974816549
-2.244069805280685	2.619724275709841
0.673828578464231	1.019189408885882
-0.728746785515074	0.057888885822448
3.207099493617568	-1.148320147970131

1.3

I wrote the code for showing the scatter plot.



1.4

The total variance is 57.2432 so in order to explain at least 90% (51.5189) of it we need 10 components (cumulative sum = 518465). We notice that we only need 10% of the components to explain at least 90% of the total variance and that is because we have sorted the eigenvalues in a decreasing order and as we move down, the eigenvalues have very small values compared to the first 10 , which hardly make any difference in the cumulative variance. (10% of components explain 90% of variance, 90% of components explain the remaining 10%).

1.5

In this task, we took a matrix which holds the two eigenvectors for the largest two eigenvalues, resulting from the training data and we classified it depending on the two principal component axes. It relates to the digit classification in task 2 in the way that we classify the data, here using two of the eigenvectors (and thus two principal components). If we take two other eigenvectors, for smaller eigenvalues, the classification is different and we can see that in the scatter plot.