EE 573 Pattern Recognition Project 1

Oğuzhan Sevim

I. INTRODUCTION

In this project, we first perform dimension reduction on a given dataset by using Principal Component Analysis (PCA). Then, on these approximated samples, Fischer Linear Discriminant (FLD) method is performed in order to separate the classes properly. The given dataset consists of $n = 1315$ samples where each sample has $d = 500$ features. The given dataset is classified into $c = 8$ different classes.

II. PRINCIPAL COMPONENT ANALYSIS

A. Selecting the number of directions ($d^{'}$)

Instead of picking up a random value for the maximum error for each class, I plotted the total error vs. the all possible $d^{'}$ values. The resulting graph is shown in Figure 1.

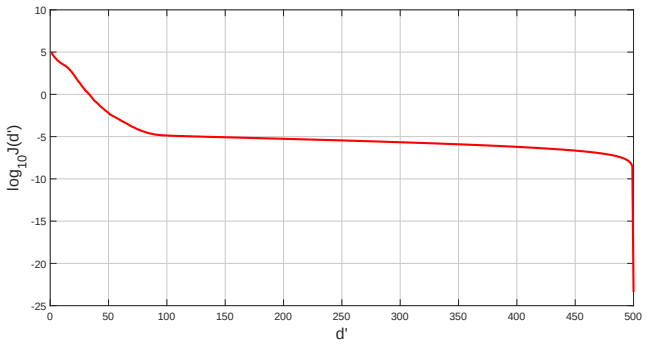

Fig. 1. Total error vs. number of directions ($d^{'}$).

In Figure 1, it can be seen that the decrease in the total error becomes too small when $d^{'} > 100$. Also, a steep decrease in the error is observed for $d^{'} < 50$. Therefore, by taking the trade of between $d^{'}$ and $J(d^{'})$ into account, a choice between $50 \leq d^{'} \leq 100$ seems to be reasonable.

The corresponding code for this graph can be found in Q1_d_prime_selection.m file.

B. Errors of different classes

For $d^{'} = 50$, the average of the approximation errors of different classes are calculated as follows:

Class	Average Error
1	2.095e-06
2	1.728e-06
3	3.497e-06
4	2.229e-06
5	2.077e-06
6	8.154e-06
7	1.196e-05
8	1.383e-05

The corresponding code for this error calculations can be found in Q1_errors_for_classes.m file.

III. FISCHER LINEAR DISCRIMINANT

After reducing the dimension of the samples from $d = 500$ to $d^{'} = 50$, FLD is performed on the data, so that the classes are assured to be separated well. After applying FLD for each class by $1 - to - rest$ method, the corresponding pairwise center distances are found as follows:

$$10^{-3} \begin{bmatrix} 0.00 & 0.59 & 0.66 & 0.67 & 0.62 & 0.67 & 0.66 & 0.68 \\ 0.59 & 0.00 & 0.26 & 0.18 & 0.13 & 0.30 & 0.29 & 0.30 \\ 0.66 & 0.26 & 0.00 & 0.21 & 0.21 & 0.21 & 0.22 & 0.21 \\ 0.67 & 0.18 & 0.21 & 0.00 & 0.13 & 0.28 & 0.28 & 0.27 \\ 0.62 & 0.13 & 0.21 & 0.13 & 0.00 & 1.44 & 1.51 & 1.54 \\ 0.67 & 0.30 & 0.21 & 0.28 & 1.44 & 0.00 & 0.09 & 0.09 \\ 0.66 & 0.29 & 0.22 & 0.28 & 1.51 & 0.09 & 0.00 & 0.02 \\ 0.68 & 0.30 & 0.21 & 0.27 & 1.54 & 0.09 & 0.02 & 0.00 \end{bmatrix}.$$

Here, the element with index (i, j) corresponds to the distance between the means of class i and class j when the transformation vector is found by considering class i versus the rest approach. Also, pairwise overlap (as percentage) matrix is calculated as

$$\begin{bmatrix} 100 & 24 & 3 & 3 & 4 & 4 & 4 & 4 \\ 24 & 100 & 5 & 50 & 60 & 9 & 30 & 3 \\ 3 & 5 & 100 & 2 & 3 & 3 & 3 & 3 \\ 3 & 50 & 2 & 100 & 42 & – & – & – \\ 4 & 60 & 3 & 42 & 100 & 4 & – & – \\ 4 & 9 & 3 & – & 4 & 100 & 4 & 4 \\ 4 & 30 & 3 & – & – & 4 & 100 & 12 \\ 4 & 3 & 3 & – & – & 4 & 12 & 100 \end{bmatrix}.$$

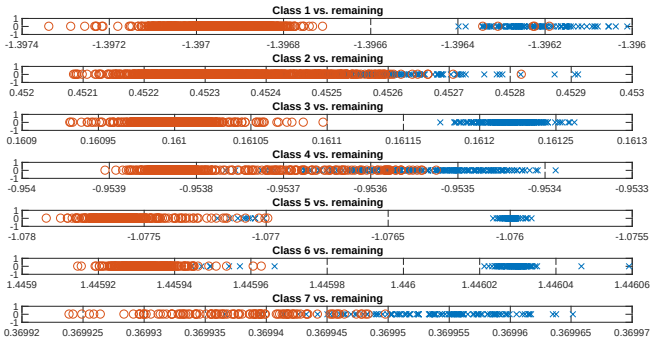

Fig. 2. 1D projection of data points.

For the better understanding of separation of the classes, the 1 dimensional projection of the data points are shown in Figure 2. On the projection line i, where $i = 1, 2, \ldots, (c-1)$, samples from class i are marked by blue x, and the union of the j indexed classes, where $j = (i+1), (i+2), \ldots, c$, are marked by red o.

The MATLAB code for FLD part can be found in Q2.m.