Mehmet Mert Bezirgan - 68756

Engr 421 – homework 4

Problem: Linear Discriminant Analysis

Solution:

For this problem we have a dataset contains 3000 train and 3000 test data vectors which are 784 dimensional. Before classifying these vectors using KNN with k = 5, we reduce the dimension of data vectors to 2. First, we calculated within class scatter matrix by summing covariance matrices for all classes. Second, we calculated between class scatter matrix and using these two matrices we found eigenvectors, which corresponds to biggest eigenvalues, of product of inverse of within class scatter matrix and between class scatter matrix. When we project our both train and test data to two dimension we got following results.

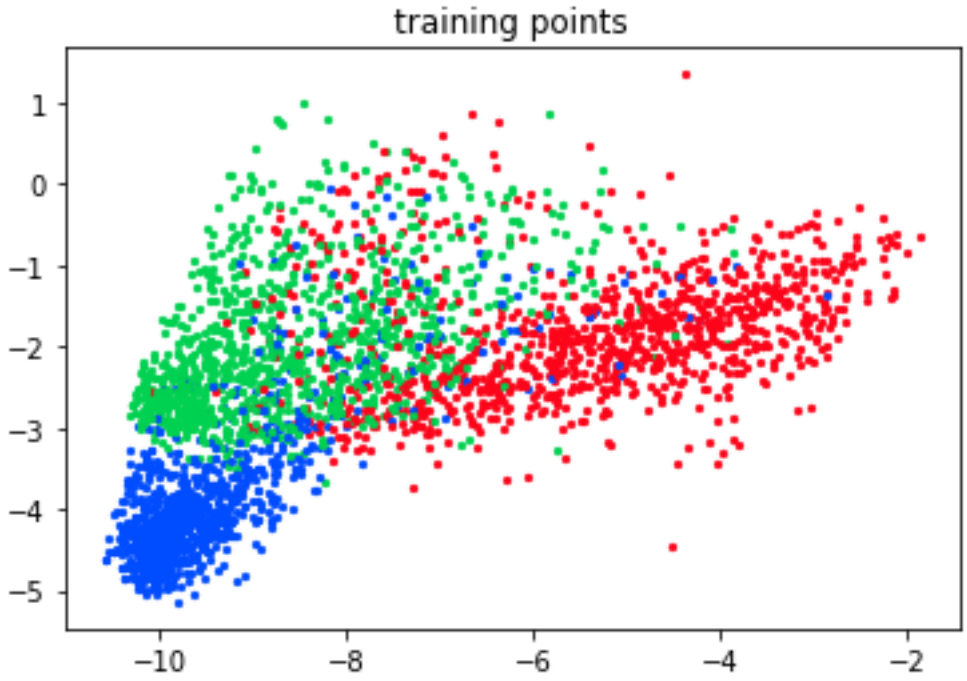

Figure 1 Projection of training points

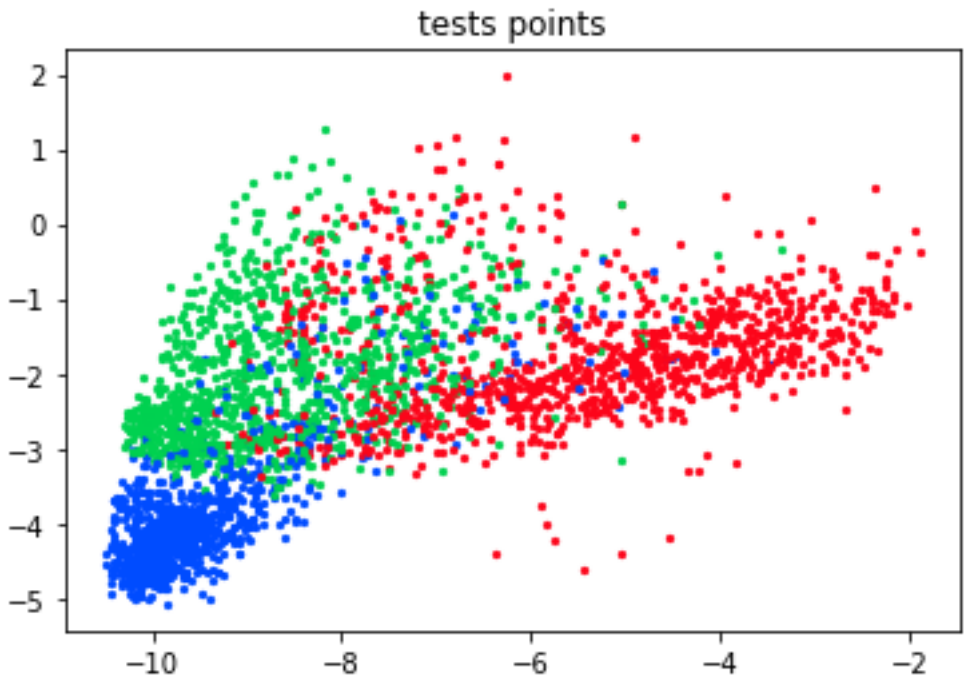

Figure 2 projection of test points

After projection step, we develop a KNN algorithm to classify our projected data points with parameter k = 5. After calculations we got confusion matrix of both training and test.

```
[[886  14 112]
 [ 40 845  89]
 [ 86  47 881]]
```

Figure 3 Training confusion matrix

```
[[835   8 157]
 [ 49 851 100]
 [134  61 805]]
```

Figure 4 Testing confusion matrix