Project 1: Classifiers

Saniya Naphade

February 23, 2020

Abstract

The aim of this project is to classify multidimensional data using Least Squares and Fisher Discriminant Analysis

Contents

1 Introduction

Pattern is defined as composite of features that are characteristic of an individual. In classification, a pattern is a pair of variables x,w where x is a collection of observations or features (feature vector) and w is the concept behind the observation (label). The quality of a feature vector is related to its ability to discriminate examples from different classes. Examples from the same class should have similar feature values and while examples from different classes having different feature values. The goal of a classifier is to partition feature space into class-labeled decision regions. Borders between decision regions are called decision boundaries. In this project, decision boundaries are constructed using Linear Discriminant Analysis(LDA). Discriminant function is used to help classify data to the appropriate classes.

2 Approach

In discriminant function, each feature vector is directly assigned to a specific class. The algorithms implemented in this project to learn the parameters of the linear discriminant function are-

1. Least Squares for classification
2. Fisher's Linear Discriminant

2.1 Data

2.1.1 Wine

These data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines.

This dataset comprises of 13 dimensions/variables with 3 classes. The training data comprises of 90 samples and the test data of 88 samples.

2.1.2 Wallpaper

A wallpaper group (or plane symmetry group or plane crystallographic group) is a mathematical classification of a two-dimensional repetitive pattern, based on the symmetries in the pattern. Such patterns occur frequently in architecture and decorative art, especially in textiles and tiles as well as wallpaper.

This dataset has 500 dimensions with 17 classes. The training data comprises of 1700 samples and the test data of 1700 samples

2.1.3 Taiji

This is a dataset of the joint angles (in quaternions) of 35 sequences from 4 people performing Taiji in LPAC motion capture lab.

This dataset comprises of 64 dimensions with 8 classes. The training data comprises of 11361 samples and the test data of 3924 samples.

3 Methods

We are going to use two methods for linear regression. They are,

1. Least Squares Discriminant
2. Fisher Discriminant

Linear Discriminat Analysis focuses on maximizing the separability between the known classes.

3.1 Least Squares Discriminant

The linear discriminant function is defined by taking a linear function of the input vector such that,

$$y(x) = w^T x + w_0 \tag{1}$$

For a total of K classes. Each class C_k is described by a linear model specific to it of the form:

$$y_k(x) = w_k^T x + w_{k0} \tag{2}$$

where k = 1, . . . , K. We can conveniently group these together using vector notation so that,

$$y(x) = \widetilde{W}^T \widetilde{x} \tag{3}$$

The value of $\widetilde{W}$ can be determined by minimizing the sum-of-squares error by taking the derivative of the error with respect to $\widetilde{W}$ and setting it to zero which gives,

$$\widetilde{W} = (\widetilde{X}^T \widetilde{X})^{-1} \widetilde{X} T \tag{4}$$

We substitute $\widetilde{W}$ back in eq(3) which gives us the discriminant function.

3.1.1 Fisher Linear Discriminant

Consider a case with only 2 classes:
Let C_1 and C_2 denote the two classes. T For a D-dimensional input vector X projecting it down to 1-D using the linear function becomes:

$$y(x) = w^T x \tag{5}$$

Let n_1 samples belong to class C_1 and n_2 samples belong to class C_2. To get the projection of the input vector we need to compute the intra covariance matrix and inter covariance matrix.

Computing the class means. i.e, means of individual classs.
Let μ_1 be the mean of C_1 and μ_2 be the mean of C_2.

$$\mu_1 = \frac{1}{n_1} \sum_{n \in C_1} x_n \tag{6}$$

$$\mu_2 = \frac{1}{n_2} \sum_{n \in C_2} x_n \tag{7}$$

The intra covariance matrix is given as,

$$S_w = \sum_{x_i \in C_1} (x_i - \mu_1)(x_i - \mu_1)^T + \sum_{x_i \in C_2} (x_i - \mu_2)(x_i - \mu_2)^T \tag{8}$$

In order to find the inter covariance matrix, we first determine the total covariance matrix whihc is given as,

$$S_T = \sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T \tag{9}$$

where μ is the mean of the total data set

$$\mu = \frac{1}{N} \sum_{n=1}^{N} x_n \tag{10}$$

$$= \frac{1}{N} \sum_{k=1}^{K} N_k \mu_k \tag{11}$$

where, $N = \sum_k N_k$ is the total number of points. The total covariance matrix can be decomposed into the sum of the intra covariance matrix, plus inter covariance matrix.

$$S_T = S_W + S_B \tag{12}$$

where,

$$S_B = \sum_{k=1}^{K} N_k (\mu_K - \mu)(\mu_k - \mu)^T \tag{13}$$

The objective of FLDA is to have inter covariance as large as possible and have intra covariance as small as possible. Giving a scalar which satisfies the requirements as,

$$J(W) = Tr\left\{ s_W^{-1} s_B \right\} \tag{14}$$

References: [1], [2]

4 Results

4.1 Data set:WINE

4.1.1 Least Squares Classifier

Visual boundaries

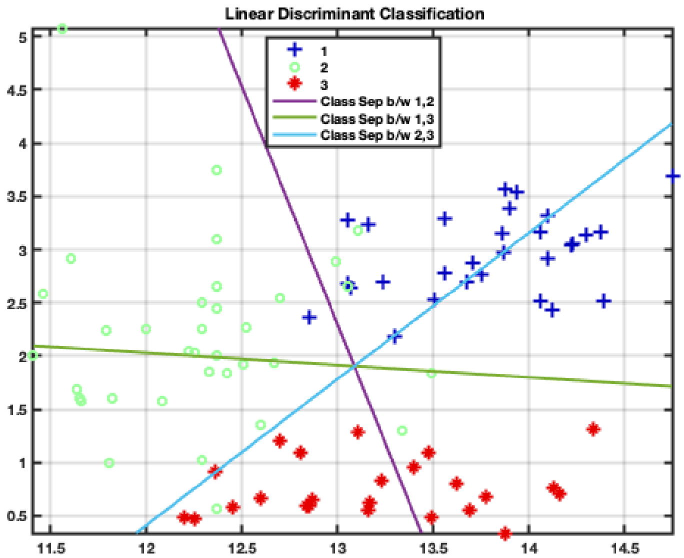

Figure 1: Least Discriminant Classsification for the wine data set, where only the first two features the wine data set were plotted in the image

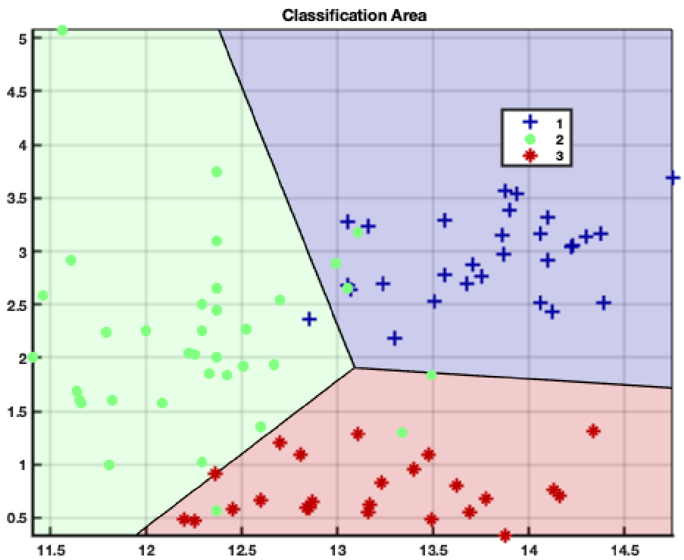

Figure 2: Least Discriminant Classsification Area for the wine data set, showing the color coded regions corresponding to they own colored points.

4.1.2 Test Results

Table 1: Accuracy of Least Squares Discriminant Function over the data sets

Dataset	Train Accuracy	Train STD	Test Accuracy	Test STD
Wine	1	0	0.98095238	0.032991443

Table 2: Accuracy of Fisher Discriminant Function over the data sets

Dataset	Train Accuracy	Train STD	Test Accuracy	Test STD
Wine	0.0419435246	0.086126044	0.87580733442	0.072208064

Table 3: Confusion Matrix LS Train data sets

30	0	0
0	36	0
0	0	24

Table 4: Confusion Matrix LS Test data sets

29	0	0
0	34	1
0	0	24

Table 5: Confusion Matrix Fisher Train data sets

29	1	0
7	29	0
0	4	20

Table 6: Confusion Matrix Fisher Test data sets

27	2	0
6	29	0
1	4	19

4.2 Data set:WALLPAPER

4.2.1 Least Squares Classifier

Visual boundaries

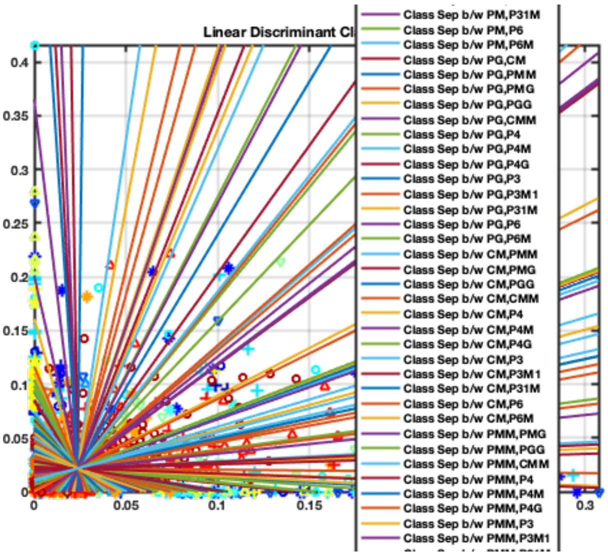

Figure 3: Least Discriminant Classsification for the wallpaper data set, where only the first two features the wallpaper data set were plotted in the image

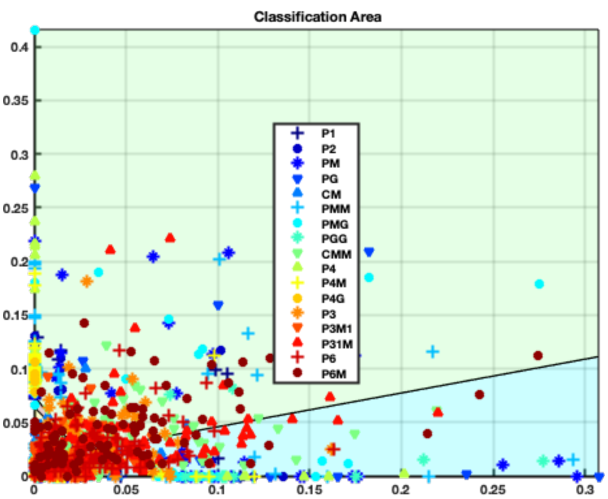

Figure 4: Least Discriminant Classsification Area for the wallpaper data set, showing the color coded regions corresponding to they own colored points.

4.2.2 Test Results

Table 7: Accuracy of Least Squares Discriminant Function over the data sets

Dataset	Train Accuracy	Train STD	Test Accuracy	Test STD
Wallpaper	0.967647059	0.032887956	0.617058824	0.269577556

Table 8: Accuracy of Fisher Discriminant Function over the data sets

Dataset	Train Accuracy	Train STD	Test Accuracy	Test STD
Wallpaper	0.975882352941176	0.0316344004	0.704117647058824	0.028888732

Table 9: Confusion Matrix LS Train data set

96	1	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0
2	90	1	5	0	0	0	0	0	0	1	0	0	0	1	0	0
1	0	97	0	0	0	2	0	0	0	0	0	0	0	0	0	0
2	2	0	91	0	0	1	3	0	1	0	0	0	0	0	0	0
0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
0	0	3	0	0	0	97	0	0	0	0	0	0	0	0	0	0
0	0	1	2	0	0	0	91	0	0	0	6	0	0	0	0	0
0	0	0	0	3	0	0	0	97	0	0	0	0	0	0	0	0
1	1	0	0	0	1	0	0	0	95	2	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	99	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0	0	0	97	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	99	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	3
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	99

Table 10: Confusion Matrix LS Test data set

13	19	4	23	1	5	2	4	3	13	2	2	1	4	4	0	0
12	19	2	16	1	10	7	9	0	10	3	2	3	1	1	2	2
2	2	62	0	0	3	26	0	1	1	0	1	1	0	1	0	0
8	11	2	37	0	4	6	13	0	5	3	4	1	3	3	0	0
0	0	0	0	90	0	0	0	2	0	0	0	4	2	0	2	0
9	6	7	3	3	36	9	3	3	8	11	0	0	0	0	0	2
0	2	14	2	0	2	71	1	0	0	3	2	0	0	0	1	2
1	4	0	6	1	0	4	46	0	0	0	37	0	0	0	1	0
0	3	2	0	16	1	0	0	64	0	0	0	1	0	4	5	4
5	15	0	10	0	3	1	3	1	28	20	5	5	1	0	1	2
0	2	0	2	0	11	0	0	1	10	70	3	0	0	0	0	1
0	0	0	0	0	0	0	7	0	0	0	93	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	72	11	15	1	0
0	0	0	0	0	0	0	0	0	0	0	0	2	98	0	0	0
0	1	0	0	0	0	0	0	0	1	0	0	12	1	82	3	0
1	1	1	0	0	0	0	0	0	0	0	0	2	0	2	86	7
0	0	4	0	0	0	0	0	4	1	0	0	0	0	0	9	82

Table 11: Confusion Matrix Fisher Train data sets

96	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	88	0	5	0	0	0	0	0	1	0	0	0	0	0	0	0
1	0	97	1	0	0	1	0	0	0	0	0	0	0	0	0	0
1	1	0	96	0	0	0	1	0	1	0	0	0	0	0	0	0
0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	99	0	0	0	0	0	0	0	0	0	0	0
0	0	3	0	0	0	97	0	0	0	0	0	0	0	0	0	0
0	0	1	3	0	0	0	93	0	0	0	3	0	0	0	0	0
0	0	0	0	2	0	0	0	98	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	97	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	99	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Table 12: Confusion Matrix Fisher Test data sets

30	27	2	22	0	4	0	1	0	12	2	0	0	0	0	0	0
22	28	3	13	0	10	3	3	0	15	3	0	0	0	0	0	0
2	1	70	2	0	3	22	0	0	0	0	0	0	0	0	0	0
16	15	2	43	0	1	5	8	0	9	1	0	0	0	0	0	0
0	0	0	0	93	0	0	0	0	0	0	0	6	0	0	1	0
12	10	4	0	0	58	5	0	0	5	6	0	0	0	0	0	0
1	4	19	1	0	0	71	2	0	0	2	0	0	0	0	0	0
2	4	0	7	0	0	3	61	0	0	0	23	0	0	0	0	0
0	0	1	0	11	0	0	0	76	0	0	0	0	0	6	6	0
15	22	0	5	0	0	1	7	0	33	15	2	0	0	0	0	0
1	1	0	2	0	6	0	0	0	15	75	0	0	0	0	0	0
0	0	0	0	0	0	0	9	0	0	0	91	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	94	2	3	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	6	0	91	3	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	93	4
0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	4	90

4.3 Data set:TAIJI

4.3.1 Least Squares Classifier

Visual boundaries

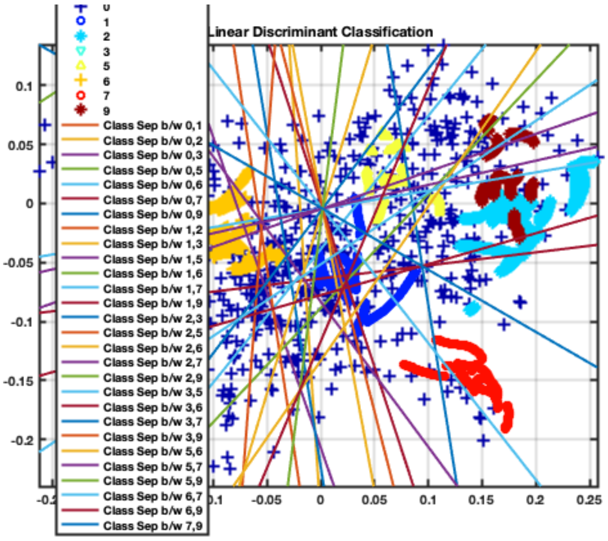

Figure 5: Least Discriminant Classsification for the taiji data set, where only the first two features the taiji data set were plotted in the image

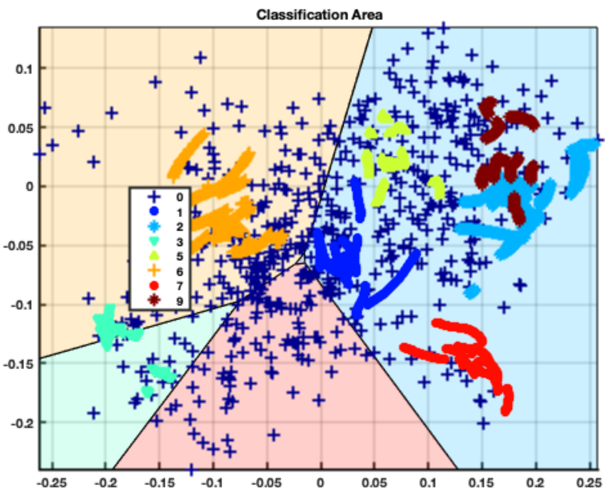

Figure 6: Least Discriminant Classsification Area for the taiji data set, showing the color coded regions corresponding to they own colored points.

4.3.2 Test accuracy

Table 13: Accuracy of Least Squares Discriminant Function over the data sets

Dataset	Train Accuracy	Train STD	Test Accuracy	Test STD
Taiji	0.961304471	0.109447484	0.928689884	0.201695466

Table 14: Accuracy of Fisher Discriminant Function over the data sets

Dataset	Train Accuracy	Train STD	Test Accuracy	Test STD
Taiji	0.981324278	0.052822917	0.940922926	0.089409522

Table 15: Confusion Matrix LS Train data set

1220	63	109	59	66	92	54	104
0	1066	0	0	0	0	0	0
0	0	2132	0	0	0	0	0
0	0	0	1066	0	0	0	0
0	0	0	0	1066	0	0	0
0	0	0	0	0	2132	0	0
0	0	0	0	0	0	1066	0
0	0	0	0	0	0	0	1066

Table 16: Confusion Matrix LS Test data set

259	32	50	27	106	43	21	65
0	369	0	0	0	0	0	0
0	0	738	0	0	0	0	0
0	0	0	369	0	0	0	0
0	0	0	0	369	0	0	0
0	0	0	0	0	738	0	0
0	0	0	0	0	0	369	0
0	0	0	0	0	0	0	369

Table 17: Confusion Matrix Fisher Train data sets

1503	45	71	25	31	44	23	25
0	1066	0	0	0	0	0	0
0	0	2132	0	0	0	0	0
0	0	0	1066	0	0	0	0
0	0	0	0	1066	0	0	0
0	0	0	0	0	2132	0	0
0	0	0	0	0	0	1066	0
0	0	0	0	0	0	0	1066

Table 18: Confusion Matrix Fisher Test data sets

461	13	24	11	41	26	12	15
59	310	0	0	0	0	0	0
15	0	723	0	0	0	0	0
18	0	0	351	0	0	0	0
0	0	0	0	369	0	0	0
6	0	0	0	0	732	0	0
0	0	0	0	0	0	369	0
0	0	0	0	0	0	0	369

5 Conclusion

From the results we can observe that Fisher Discriminant function performs better than Least Squares Discriminant function in case of a large number of outliers. We also observe that Fisher has better accuracy than Least Squares.

6 Extra Credit

Table 19: Confusion Matrix Fisher Train data sets

30	0
3	33

Table 20: Confusion Matrix Fisher Test data sets

29	0
4	32

References

[1] Bishop, Christopher M. Pattern Recognition and Machine Learning. *Pattern Recognition and Machine Learning.*

[2] $https://www.csd.uwo.ca/~olga/Courses/CS434a_541a/Lecture8.pdf$ 4