ENEE633 Project 1 Report

Yang Jiao

October 19, 2015

Abstract

This project is composed of four parts, each with a different classification strategy. In each part, a classifier or a combination of pattern recognition techniques are employed over three data sets. This report details how experiments are conducted and presents the outcomes. Brief discussion follows to highlight the trends in the results. Finally, a conclusion section compares strategies for each data set.

1 Data Preparation

The experiments are conducted over three data sets of face images. The classification methods are run directly over those cropped images without further preprocessing. Each data sets is divided into a training set and a testing set. A summary of data sets are presented in Table 1. Some experiments may vary the size of training sets and not detailed in this table.

Table 1: Data Sets Overview

Data Set	Image Size	no. Classes	no. Train Samp.	no. Test Samp.
data.mat	24 * 21	200	2 per class	1 per class
pose.mat	48 * 40	68	9 per class	4 per class
illumination.mat	48 * 40	68	14 per class	7 per class

Data Set	Description
data.mat	3 faces for each of 200 classes, 1 neutral and 1 expression faces for training, 1 illumination face for testing
pose.mat	13 faces with different poses for each of 68 classes
illumination.mat	21 faces with different illuminations for each of 68 classes

2 ML Estimation with Gaussian Assumption Followed by Bayes Rule

2.1 Train the Bayes Classifier

The Bayes classifier is trained based on ML estimation with Gaussian assumption. It is assumed different classes have different covariance matrices. Experiments under same covariance matrix assumption are discussed in other sections.

2.2 Test the Bayes Classifier

To handle the problem of singular covariance matrices, an identity matrix is added to the estimated matrix. Other methods are discussed in later sections.

2.3 Experiments and Results

Table 2: ML Estimation Followed by Bayes Classifier

Data Set	Error Rate (%)	no. Train Samp.	Dim. of Feature
data.mat	0.3800	2 per class	504
pose.mat	0.2500	9 per class	1920
illumination.mat	0	14 per class	1920

2.4 Discussion

Bayes classifier works best on illumination.dat, the data set that has most training samples; then ok on pose.mat, the data set with moderate training points; and worst on data.mat, that has very limited training samples.

3 Nearest Neighbor Rule

3.1 Implementation

A 1-nearest neighbor rule is implemented. If there are multiple 1-nearest neighbor, the majority of the labels is chosen.

3.2 Experiments and Results

3.3 Discuss

- In Table 3, the performance of Bayes and NN classifiers are compared. Bayes over-performs NN over three data sets.

Table 3: Nearest Neighbor v.s. Bayes Rule

Data Set	Error Rate	no. Train Samp.	Bayes Results	no. Train Samp.
data.mat	0.4050	400	0.3800	2 per class
pose.mat	0.3015	612	0.2500	9 per class
illumination.mat	0.0462	952	0	14 per class

Table 4: NN with Varying Training Sets

Data Set	Error Rate	no. Train Samp.	Dim. of Feature
illumination.mat	0.5420	7 * 68	1920
illumination.mat	0.4244	8 * 68	1920
illumination.mat	0.4244	9 * 68	1920
illumination.mat	0.4223	10 * 68	1920
illumination.mat	0.4202	11 * 68	1920
illumination.mat	0.1723	12 * 68	1920
illumination.mat	0.0693	13 * 68	1920
illumination.mat	0.0462	14 * 68	1920
pose.mat	0.6471	2 * 68	1920
pose.mat	0.5588	3 * 68	1920
pose.mat	0.5588	4 * 68	1920
pose.mat	0.4485	5 * 68	1920
pose.mat	0.3309	6 * 68	1920
pose.mat	0.3309	7 * 68	1920
pose.mat	0.3309	8 * 68	1920
pose.mat	0.3015	9 * 68	1920

- In Table 4, for both illumination.mat and pose.mat, the trend is clear as we varying the size of the training set. The more training samples, the better performance measured in error rate we can obtain.

4 PCA Followed by Bayes Classifier and NN Rule

4.1 Implementation

The amount of variance kept is measured by cumulative sum of descending eigenvalues over sum of all eigenvalues. By default, the lowest dimension that keeps more than 95% of variance is kept by PCA, if output dimension is not specified.

Table 5: PCA Followed by Bayes

Data Set	99%	95%	90%	80%
data.mat	0.3750	0.3900	0.4150	0.5000
pose.mat	0.2426	0.2279	0.1985	0.2316
illumination.mat	**0**	**0.0021**	**0.0651**	**0.7962**

Table 6: PCA Followed by NN

Data Set	99%	95%	90%	80%
data.mat	0.4150	0.4200	0.4200	0.5050
pose.mat	0.3088	0.2941	0.2904	0.3676

4.2 Experiments and Results

4.3 Discuss

- In Table 5, the trend is vague for data.mat and pose.mat. On the one hand, PCA relieves the problem of small training set. On the other hand, PCA discards components that could be helpful. Thus, the trend is vague for these two data sets.
- However, illumination.dat has more training samples. Thus the trend is clear in Table 5.
- Similarly, in Table 6, the trend is vague for data.mat and pose.mat.
- And in Table 8, for illumination.dat, the more components discarded, the worse the performance measured in error rate.
- In Table 7, it is studied how PCA can help relieve the problem of singular covariance matrix. The identity matrix applied in part 1 is removed. Also, the assumption is changed to all classes have the same covariance matrix.
- It is seen that although more samples are used to estimate the covariance matrix, the matrix is still singular for both data sets.
- PCA helps to relieve the problem of lack of training set. The trend is in 'U' shape, i.e. as the dimension goes down, the performance goes better, reach an optimal point, then goes worse.

5 LDA Followed by Bayes Classifier and NN Rule

5.1 Implementation

PCA is employed before LDA to handle the problem of singular covariance matrices. By default, PCA reduces the space to N - c, where N is the number

Table 7: Varying no. Principle Components Followed by Bayes

Data Set	Error Rate	Total Train Samples	Feature Dimension	Percent
data.mat	0.4200	2 * 200	100	0.91691
data.mat	0.3650	2 * 200	150	0.96245
data.mat	**0.3350**	**2 * 200**	**160**	**0.96813**
data.mat	0.3550	2 * 200	170	0.97303
data.mat	0.4100	2 * 200	180	0.97726
data.mat	0.4650	2 * 200	200	0.98398
data.mat	Singular	2 * 200	$24 * 21 = 504$	1
pose.mat	0.5735	9 * 68	10	0.69195
pose.mat	0.3346	9 * 68	20	0.77175
pose.mat	**0.3309**	**9 * 68**	**40**	**0.83945**
pose.mat	0.3419	9 * 68	50	0.85936
pose.mat	0.9853	9 * 68	100	0.9135
pose.mat	0.9853	9 * 68	200	0.95828
pose.mat	0.9853	9 * 68	500	0.99652
pose.mat	Singular	9 * 68	1920	1

Table 8: Varying no. Principal Components Followed by NN

Data Set	99%	98%	97%	96%	95%	94%	93%	92%
illumination.mat	0.0861	0.1303	0.1408	0.1555	0.1807	0.1933	0.2101	0.2416
Data Set	91%	90%	89%	87%	85%	83%	81%	80%
illumination.mat	0.2521	0.2920	0.3235	0.4454	0.5546	0.6849	0.8046	0.8046

of training samples, and c is the number of classes.

5.2 Experiments and Results

5.3 Discussion

- In Table 9, it is shown that lack of training samples results in high error rate, even with PCA, LDA, plus identity matrix is added to the covariance matrix.

- Further look into Table 9, it is shown that there is still some flexibility in the application of PCA in pose.mat. The dimension can be further reduced while stay above c - 1. This is not possible for data.mat, as N - c is close to c - 1.

- In Table 10, the dimension is further reduced from N - c to c - 1. It is presented in the table that this is a helpful strategy to improve the LDA performance. As the dimension is reduced to the minimum available, the performance is the best.

Table 9: PCA LDA Bayes and PCA LDA NN

Data Set	Method	Error Rate	N - c	c - 1
data.mat	PCA LDA Bayes	0.4400	200	199
pose.mat	**PCA LDA Bayes**	**0.4007**	**544**	**67**
illumination.mat	PCA LDA Bayes	0	816	67
data.mat	PCA LDA NN	0.47	200	199
pose.mat	PCA LDA NN	0.3676	544	67
illumination.mat	PCA LDA NN	0	816	67

Table 10: Varying no. Principal Components, LDA, Same Covariance Matrix for Bayes

Data Set	Error Rate	Feature Dimension	Percent
pose.mat	**0.3199**	**67**	**0.88333**
pose.mat	0.3235	80	0.89703
pose.mat	0.3235	90	0.90584
pose.mat	0.3272	100	0.9135
pose.mat	0.3787	200	0.95828
pose.mat	0.4081	500	0.99652
pose.mat	Singular	1920	1

- In Table 11, several methods applied to the data.mat data sets are summarized. Bayes and NN suffers from the problem of small training set. The usage of PCA before LDA is limited, as N - c is close to c - 1. PCA to even lower dimension plus same covariance matrix assumption outperforms other results.

6 Conclusion

In this project, 4 classification strategies are applied to 3 data sets. Based on the experiments results presented in this report, it can be concluded that:

- data.mat requires more training samples to improve classification performance. Otherwise, Bayes classifier with 160 PCA plus same covariance matrix assumption works best with 0.3350 % error rate.

- pose.mat has moderate training samples. 67, 88.33 % PCA + LDA + Bayes classifier + different covariance assumption + identity matrix works best on it (not presented in the previous table, 0.1875 compared to 0.3199 in Table 10 under same covariance assumption.) This combination reduces the dimension to the minimum possible (so that LDA can be further applied) and addresses the problem of singular covariance matrix. It also takes into account of suitable model assumption by not assuming all classes have same covariance matrix.

Table 11: Methods for data.mat Summary

Data Set	Error Rate	methods
data.mat	Singular	different sigma
data.mat	Singular	same sigma
data.mat	0.3800	different sigma + identity
data.mat	0.3750	same sigma + identity
data.mat	0.3900	identity
data.mat	**0.3350**	**160 PCA + same sigma**
data.mat	0.4400	different sigma add identity + 200 PCA + LDA
data.mat	0.4650	same sigma + 200 PCA + LDA
data.mat	0.3900	different sigma add identity + 133, 95% PCA
data.mat	0.4000	same sigma + 133, 95% PCA
data.mat	0.4050	NN
data.mat	0.4200	PCA + NN
data.mat	0.4700	200 PCA + LDA + NN

- illumination.dat has the most training samples. Most strategies renders acceptable performance. It helps to study several trends in this report. PCA and LDA help to reduce computation complexity and not necessarily compromise error rate.