Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

LDA Extension via Oblique Projection

IOP Conf. Series: Journal of Physics: Conf. Series **1060** (2018) 012049 doi:10.1088/1742-6596/1060/1/012049

LDA Extension via Oblique Projection

B J Yang[1], W G Shang[2] and Y J Wu[2, *]

[1] School of Mathematic Sciences, Guizhou Normal University, Guiyang, P.R. China
[2] School of Information Engineering, Guangdong Medical University, Dongguan, P.R. China

[*]E-mail: wuyingjiang@163.com

Abstract. The classical linear discriminant analysis (LDA) was previously modified by orthogonal projection into null space LDA (N_LDA) and direct LDA (D_LDA) for solving small sample size (SSS) problem. In this paper, the author proposes an extension of LDA by oblique projection, wherein N_LDA and D_LDA are included as special cases, to reduce discriminative information loss resulted from single N_LDA or D_LDA. The effectiveness of the proposed algorithm is tested by image forensics and face recognition.

1. Introduction

Linear discriminant analysis (LDA) is one of the most popular means in pattern recognition [1]. The objective of LDA is to find a projection W that maximizes the ratio of the between-class scatter matrix, S_b, against the within-class scatter matrix, S_w, i.e., via the following Fisher criterion:

$$J(W_{opt}) = \arg \max_{W} \frac{\left|WS_bW^{\mathrm{T}}\right|}{\left|WS_wW^{\mathrm{T}}\right|}. \tag{1}$$

One of the major drawbacks of LDA is the so-called small sample size (SSS) problem encountered in many practical applications involving high dimensional data [1]. Many techniques have been proposed to solve this problem, and Ref. [2] may be consulted in detail. Among them, the most notable approaches are null space LDA (N_LDA) and direct LDA (D_LDA) [2, 3], which are based on the following modified Fisher criterion [4]:

$$WS_wW^{\mathrm{T}} = 0 \quad and \quad WS_bW^{\mathrm{T}} \neq 0. \tag{2}$$

Since N_LDA and D_LDA are based on orthogonal projection, and generally the two conditions in equation (2) cannot be simultaneously satisfied, the processes are usually fulfilled in two steps: in N_LDA the null space of S_w is found first and then that of S_b is discarded, while in D_LDA the reverse order is taken. In this manner, much discriminated information may be lost.

In this paper, an extension of LDA is proposed by oblique projection, and it is called OP_LDA, for reduction of possible loss of discriminative information in N_LDA or D_LDA. Experimental results on image forensics and face recognition show satisfactory performance of the proposed method.

The rest of this paper is organized as follows. An overview of N_LDA and D_LDA is briefly introduced in Section 2. Section 3 describes OP_LDA and discusses the relationship among N_LDA, D_LDA and OP_LDA. In Section 4, we evaluate the performance of OP_LDA and compare its

IOP Conf. Series: Journal of Physics: Conf. Series **1060** (2018) 012049 doi:10.1088/1742-6596/1060/1/012049

performance with LDA, N_LDA, D_LDA, and 2-dimensional LDA (2D_LDA) [5]. Finally, conclusions are given in Section 5.

2. N_LDA and D_LDA

2.1. N_LDA

In N_LNA, the discriminative information can be found in the null space of Sw by two steps [2]:

Step 1: Find the null space of S_w, Null(S_w).

Step 2: Let $S_b'=P_{\text{Null}\,(Sw)}(S_b)$, and then

$$W_1 = P_{\text{Col}(S_b')}, \tag{3}$$

where $P_{\text{Null}(Sw)}(S_b)$ is a matrix obtained by orthogonal projecting the column vectors of S_b onto Null(S_w), Col(S_b') is the column space of S_b', $P_{\text{Col}(Sb')}$ is an orthogonal projection operator onto Col(S_b').

Figure 1 depicts a simple example of this algorithm, where v is an arbitrary vector in the input space, and v' is a new vector obtained from v by N_LDA.

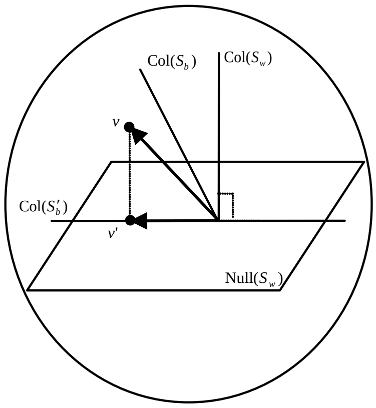

Figure 1. A simple example of N_LDA.

2.2. D_LDA

In D_LNA, the discriminative information can be found in the column space of S_b by two steps [3]:

Step 1: Find the column space of S_b.

Step 2: Let $S_w'=P_{\text{Col}(Sb)}(S_w)$, and then

$$W_2 = P_{\text{Null}|_{\text{Col}(S_b)}(S_w')}, \tag{4}$$

where $P_{\text{Col}(Sb)}(S_w)$ is a matrix obtained by orthogonal projecting the column vectors of S_w onto Col(S_b), $\text{Null}|_{\text{Col}(Sb)}(S_w')$ is the null space of S_w' in Col(S_b), $P_{\text{Null}|_{\text{Col}(Sb)}(Sw')}$ is an orthogonal projection operator onto $\text{Null}|_{\text{Col}(Sb)}(S_w')$.

Since rank(S_b), the rank of matrix S_b, is usually smaller than rank(S_w), and in this case Col(S_w')=Col(S_b) and $\text{Null}|_{\text{Col}(Sb)}(S_w')=\{0\}$, the solution of D_LDA is given by

$$W_2 = P_{\text{Col}(S_b)}. \tag{5}$$

Figure 2 depicts a simple example of this algorithm, where v is an arbitrary vector in the input space, and v' is a new vector obtained from v by D_LDA.

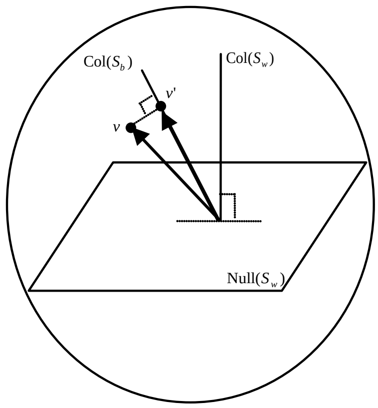

Figure 2. A simple example of D_LDA.

3. Proposed method

3.1. OP_LDA

We present a new LDA algorithm based on oblique projection, named OP_LDA. The objective of OP_LDA is to find a projection operator W that satisfies equation (6), which also is special circumstance of equation (2).

$$WS_w = 0 \quad and \quad WS_b \neq 0. \tag{6}$$

The key idea of our algorithm is that W is an oblique projection operator, whereas the traditional methods take it as orthogonal projection. A particular solution of equation (6) is given by

$$W_3 = P_{\mathrm{Col}(S_b)|\mathrm{Col}(S_w)}, \tag{7}$$

where $P_{\mathrm{Col}(Sb)|\mathrm{Col}(Sw)}$ is an oblique projection operator onto Col(S_b) along Col(S_w), and W_3 can be obtained from Col(S_w) and Col(S_b), Ref. [6] may be consulted in detail.

Figure 3 depicts a simple example of this algorithm, where v is an arbitrary vector in the input space, and v' is a new vector obtained from v by OP_LDA.

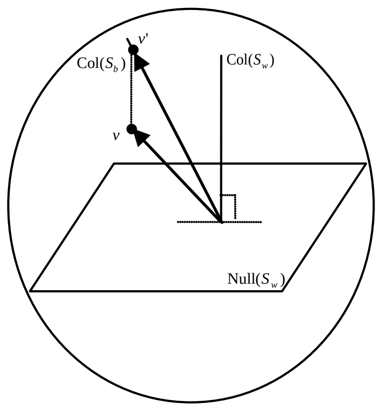

Figure 3. A simple example of OP_LDA.

3.2. Discussion

As shown in Figure 1, Figure 2 and Figure 3, there are three means to obtain v': a. discarding the component of Col(S_w), b. keeping that of Col(S_b), and c. discarding that of Col(S_w) and keeping that of Col(S_b) simultaneously. So it is clear that OP_LDA including N_LDA and D_LDA as special cases when Col(S_w) and Col(S_b) are orthogonal.

The traditional methods emphasize S_w or S_b respectively, and our method emphasizes S_w and S_b simultaneously. For discarding Col(S_w) in N_LDA, S_b is converted into S_b' by projecting onto Null(S_w). Since there are differences between S_b and S_b', discriminated information may be lost by adopting S_b'. The similar drawback may appear in D_LDA. Therefore discriminative information loss in N_LDA or D_LDA may be reduced by adopting S_w and S_b directly in OP_LDA.

3.3. Analysis

The following condition needs to be satisfied in using $P_{\mathrm{Col}(Sb)|\mathrm{Col}(Sw)}$: the direct sum of Col($S_b$) and Col($S_w$) must be equal to the input space [6]. Since the direct sum is the sum of two disjoint spaces, it includes two meanings: the sum of Col(S_b) and Col(S_w) must be equal to the input space, and Col(S_b) and Col(S_w) must be disjoint, i.e., Col(S_w)∩Col(S_b)={0}.

Since the sum of Col(S_b) and Col(S_w) is a subspace of the input space, $P_{\mathrm{Col}(Sb)|\mathrm{Col}(Sw)}$ is invalid for vectors which do not belong to the sum space. Therefore we project vectors in this case onto Col(S_t) first, where S_t denotes the total scatter matrix. Since Null(S_w)∩Null(S_b)=Null(S_t) has been proven in Ref. [7], and S_w, S_b and S_t are symmetric matrixes, it is clear that Col(S_t) is the sum of Col(S_b) and Col(S_w).

If Col(S_w)∩Col(S_b)≠{0}, $P_{\mathrm{Col}(Sb)|\mathrm{Col}(Sw)}$ does not exist. This is a more complicated problem. In this paper we consider a simple scheme as follow:

$$W_3 = \begin{cases} P_{\mathrm{Col}(S_b)|\mathrm{Col}(S_w)} P_{\mathrm{Col}(S_t)}, \mathrm{Col}(S_w) \cap \mathrm{Col}(S_w) = \{0\} \\ P_{\mathrm{Col}(S_b)|P_{\mathrm{Col}(S_w)}\mathrm{Null}(S_b)} P_{\mathrm{Col}(S_t)}, \textit{otherwise.} \end{cases} \tag{8}$$

where $P_{\mathrm{Col}(St)}$ is a orthogonal projection operator onto Col(S_t), $P_{\mathrm{Col}(Sw)}$Null(S_b) is a space obtained by orthogonal projecting the vectors of Null(S_b) onto Col(S_w), and $P_{\mathrm{Col}(Sb)|P_{\mathrm{Col}(Sw)}\mathrm{Null}(Sb)}$ is an oblique projection operator onto Col(S_b) along $P_{\mathrm{Col}(Sw)}$Null(S_b).

Figure 4 depicts a simple example of our method, where v is an arbitrary vector in the input space, and v' is a new vector obtained from v by equation (8). It is clear that we want v' to lie in Col(S_b) and discard the partial components of Col(S_w).

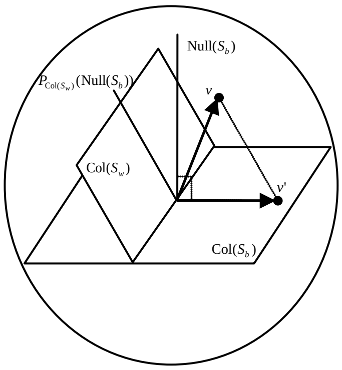

Figure 4. A simple example of proposed method for Col(S_w)∩Col(S_b)≠{0}.

4. Experimental results

4.1. Experimental setup

We illustrate the efficacy of the proposed method on image forensics and face recognition. As application-based development, some parameters and schemes have been empirically determined in our implementation.

For finding the null space and the column space of an arbitrary symmetric matrix, M, we perform first the singular value decomposition of M as $M=U\sum V^{\mathrm{T}}$, and then let Col(M)=span{$u_1,\ldots,u_r$} and

Null(M)=span{$u_{r+1},\ldots,u_n$}, where $U=[u_1,\ldots,u_r,u_{r+1},\ldots,u_n]$, $u_1,\ldots,u_r$ correspond to the nonzero eigenvalues, and $u_{r+1},\ldots,u_n$ correspond to the zero eigenvalues [8].

In order to conform whether Col(S_b) and Col(S_w) are disjoint, we consider a simple scheme via the following dimensional formula [9]: if rank(S_t)=rank(S_b)+rank(S_w), then Col(S_w)∩Col(S_b)={0}, otherwise Col(S_w)∩Col(S_b) ≠{0}.

To evaluate the performance of the proposed method, the training samples are randomly selected about half of the image dataset to train the classifier, and the remaining images are used in testing.

4.2. Image forensics

OP_LDA is tested by blurring detection in image forensics. The dataset and feature extraction are same as Ref. [10], wherein details can be found. The dataset consists of 183 authentic images and 183 forged images, and three types of feature vectors respectively consisted of the singular values of gray image matrix, correlation coefficients for double blurring operation, and image quality metrics are extracted and then fused for forgery detection.

In comparison, Table 1 lists the detection rates of the following four different detection schemes: feature fusion obtained respectively by LDA, N_LDA, D_LDA, and OP_LDA, respectively plus Euclidean distance (ED) classifier. In this table, "--" means that LDA is ineffective, due to the SSS problem.

Table 1. Detection rate vs. different schemes in image forensics.

Scheme	Detection rate (%)
LDA+ED	--
N_LDA+ED	86.81
D_LDA+ED	60.55
OP_LDA+ED	86.81

Since image forensics is in fact a binary classification problem, and in this case Null(S_w) is lost in D_LDA, D_LDA performs poor. While Col(S_b) is kept in N_LDA , so N_LDA obtained the same result as OP_LDA in this experiment.

4.3. Face recognition

OP_LDA is tested by face recognition using three face datasets: YALE, ORL, and PIE, details can be found in Ref. [5]. The YALE dataset contains 165 face images of 15 persons. The image size is 100×100. We subsample the images down to a size of 25×25. The ORL dataset contains 400 face images of 40 persons. The image size is 92×112. We subsample the images down to a size of 23×28. The PIE dataset is a subset of the CMU_PIE face image dataset. It contains 6615 face images of 63 persons. The image size is 486×640. We subsample the images down to a size of 98×128. Table 2 lists the recognition rates of the following five different schemes: feature extraction obtained respectively by LDA, N_LDA, D_LDA, OP_LDA, and 2D_LDA, respectively plus ED classifier.

Table 2. Recognition rate vs. different schemes in face recognition.

Database / Scheme	YALE	ORL	PIE
LDA+ED	--	--	--
N_LDA+ED	92.93	93.6	100
D_LDA+ED	87.87	89.1	96.39
OP_LDA+ED	93.07	93.75	98.59
2D_LDA+ED	92.53	93.4	99.99

The results demonstrate that our proposed method can achieve comparable performance to 2D_LDA in [5]. Compared to the prior arts in [2, 3], OP_LDA performs better than N_LDA and D_LDA in using YALE and ORL face datasets. Due to the SSS problem, LDA is ineffective in these experiments.

5. Experimental results

An extension of LDA by oblique projection for solving SSS problem is proposed in this paper. Instead of using orthogonal projection, oblique projection is adopted to reduce possible discriminative information loss in N_LDA or D_LDA, two popular modified versions of the traditional LDA. Experimental results on image forensics and face recognition showed satisfactory performance of the proposed method. OP_LDA may be converted into a nonlinear version by kernel method, which is one of our future tasks.

Acknowledgments

This work was supported in part by the Doctor Start-up Fund of Guizhou Normal University (2016)

References

[1] Fukunaga K 1990 *Introduction to Statistical Pattern Recognition 2th ed.* (New York: Academic Press) chapter 10
[2] Chen L F, Liao H Y M, Ko M T, Lin J C and Yu G J 2000 A new LDA-based face recognition system which can solve the small sample size problem *Pattern Recognition* **33(10)**, 1713–26
[3] Yu H and Yang J 2001 A direct LDA algorithm for high-dimensional data: with application to face recognition *Pattern Recognition* **34(10)** 2067–70
[4] Liu K, Cheng Y Q and Yang J Y 1993 Algebraic feature extraction for image recognition based on an optimal discriminant criterion *Pattern Recognition* **26(6)** 903–911
[5] Ye J P, Janardan R and Li Q 2004 Two-dimensional linear discriminant analysis *Advances in Neural Information Processing Systems* **17** 1569–76
[6] Liu B Y 2012 Incremental learning of oblique projection-based kernel discriminator: Theory and algorithm *Sciencepaper Online*
[7] Yang J, Zhang D and Yang J Y 2003 A generalised K-L expansion method which can deal with small sample size and high-dimensional problems *Pattern Analysis and Applications* **6(1)** 47–54
[8] Zhang X D 2004 *Matrix Analysis and Applications* (Beijing: Tsinghua University Press) pp 609–612
[9] Li D C 1995 On Theorems of Dimensions and Bases of Subspace’s Intersections *Journal of Southwest Nationalities College, Natural Science Edition* **21(2)** 224–226
[10] Yang B J and Liu B Y 2013 Image forgery detection via multifeature classification *Journal of Pattern Recognition and Image Processing* **4(4)** 413–421