L1-Norm Distance Discriminant Analysis with Multiple Adaptive Graphs and Sample Reconstruction

Guowan SHAO [a, 1] , Chunjiang PENG [a], Wenchu OU [a] and Kai DUAN [b]

[a] *College of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China*
[b] *Engineering Research Center for Advanced Mine Equipment of Ministry of Education, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China*

Abstract. Linear discriminant analysis (LDA) is sensitive to noise and its performance may decline greatly. Recursive discriminative subspace learning method with an L1-norm distance constraint (RDSL) formulates LDA with the maximum margin criterion and becomes robust to noise by applying L1-norm and slack variables. However, the method only considers inter-class separation and intra-class compactness and ignores the intra-class manifold structure and the global structure of data. In this paper, we present L1-norm distance discriminant analysis with multiple adaptive graphs and sample reconstruction (L1-DDA) to deal with the problem. We use multiple adaptive graphs to preserve intra-class manifold structure and simultaneously apply the sample reconstruction technique to preserve the global structure of data. Moreover, we use an alternating iterative technique to obtain projection vectors. Experimental results on three real databases demonstrate that our method obtains better classification performance than RDSL.

Keywords. Discriminant analysis, dimensionality reduction, adaptive graphs, sample reconstruction

1. Introduction

High dimensional data are available in pattern recognition and computer vision applications. However, it is difficult to directly use the data for some reasons, such as the curse of dimensionality and high computational complexity [1]. A natural solution to the difficulty is to perform dimensionality reduction or subspace learning for data representation. Many methods are presented to realize the fundamental task under the assumption that the most useful information lies in a latent subspace [2]. Among them, linear discriminant analysis (LDA) [3] and principal component analysis (PCA) [4] are two classical and representative methods for supervised and unsupervised dimensionality reduction, respectively.

---

[1] Corresponding Author, Guowan SHAO, College of Mechanical and Electrical Engineering, Hunan University of Science and Technology, Xiangtan 411201, Hunan, PR China; E-mail: gwshao_ezhou@163.com, 821211624@qq.com.

According to the definition of LDA, it measures within-class and between-class distances in Euclidean space, namely LDA is based on L2-norm. Therefore, LDA is sensitive to noise and its performance may decline greatly. Some methods use L1-norm to measure the distances for enhancing the robustness [5-9]. Rotational invariant LDA (RILDA) [5] uses RI L1-norm or L2,1-norm as the metric and constructs a more generalized dimensionality reduction framework. However, the proposed algorithm is not convergent and thus the performance may be further improved. L1-norm LDA [6] extracts local optimal projection vectors by replacing L2-norm with L1-norm based on the formulation of LDA. It avoids the singular problem existing in LDA and obtains better results than RILDA. L1-LDA [7] applies the Bayes optimality theory for excellent performance. However, the iterative algorithm enlarges the Bayes error upper bound and cannot guarantee the optimality. Besides, its robustness cannot be enhanced for the overemphasis on the minimum distance between boundary points and the center of all points. In this situation, non-greedy L1-LDA [8] combines two L1-norm terms in a difference form and gives a non-greedy iterative algorithm, which obtains better performance. L1-norm distance-based effective LDA [9] is similar to non-greedy L1-LDA. However, the two methods are sensitive to initialization for they adopt the Rayleigh quotient as the learning rate.

Recently, a novel recursive discriminative subspace learning method with an L1-norm distance constraint (RDSL) has been proposed to obtain a robust and discriminative subspace [10]. The method formulates LDA with the maximum margin criterion and avoids the sensitiveness to initialization. In addition to L1-norm, RDSL introduces slack variables and thus it is more robust to noise. However, the method only considers inter-class separation and intra-class compactness and ignores the intra-class manifold structure and the global structure of data. It is beneficial to utilize structure information for better performance.

In this paper, we present L1-norm distance discriminant analysis with multiple adaptive graphs and sample reconstruction (L1-DDA) to solve the problem existing in RDSL. We apply Shannon entropy to adaptively construct a graph for each class as a regularization term and thus intra-class structure information may be incorporated into projection vectors. L1-norm is used in the construction of adaptive graphs in order to improve the robustness. Meanwhile, we add a sample reconstruction term to approximately preserve the global structure of data, and sample reconstruction is in the form of cutting squared L2-norm. Finally, we use an alternating iterative technique to optimize the objective function. Experimental results on three publicly available real databases demonstrate that our method obtains better classification performance than RDSL.

2. Related Work

In this section, we briefly review RDSL for our method is based on it and then simply explain Shannon entropy which is applied to multiple adaptive graphs. Given a data set $X = [x_1, x_2, \cdots, x_n]$, $x_j \in R^m$, $j = 1, 2, \cdots, n$, and the n points in the m dimensional space come from c classes. Let x_{ij} denote the j th point from class i and n_i the number of points from class i. In addition, let $\bar{x}_i$ denote the mean of class i and $\bar{x}$ the mean of all the classes.

2.1. Recursive Discriminative Subspace Learning

LDA uses L2-norm to construct an objective function and thus its performance may deteriorate when noisy data are available. In this case, some methods [8-9] adopt L1-norm to cope with the problem. However, the methods are sensitive to initialization. RDSL applies the maximum margin criterion to formulate LDA and avoids the sensitiveness.

RDSL is formulated as [10]

$$\min_{w_q,\xi_i} \sum_{i=1}^{c}\sum_{j=1}^{n_i}\left\|w_q^T(x_{ij}-\bar{x}_i)\right\|_1 + \lambda\sum_{i=1}^{c}\xi_i$$

$$\text{s.t. } \left\|w_q^T(\bar{x}_i-\bar{x})\right\|_1 + \xi_i \geq 1$$

$$\xi_i \geq 0, \quad i=1,2,\cdots,c \tag{1}$$

where w_q is the q th projection vector, $1 \leq q \leq d$, d is the number of projection vectors, and accordingly the projection matrix $W=[w_1,w_2,\cdots,w_d]$. ξ_i is a slack variable, which is introduced to avoid the overfitting and λ is a regularization parameter which can control the generalization or the robustness of RDSL.

Eq. (1) can be transformed into a linear programming problem or a convex quadratic programming problem and then the solution can be obtained by an iterative algorithm [10].

2.2. Shannon Entropy

Information theoretic learning techniques have been extensively applied in the pattern recognition field and many methods for dimensionality reduction have involved entropies. Among different types of entropies, Shannon entropy is most widely used since, in general, there is a close-form solution with the entropy.

Shannon entropy is defined as follows [11]:

$$H(P)=-\sum_{i=1}^{N}p_i\log p_i$$

$$\text{s.t. } \sum_{i=1}^{N}p_i=1, p_i \geq 0$$

where $P=(p_1,p_2,\cdots,p_N)$ is a probability distribution. The definition measures the uncertainty of the distribution. More specifically, the value of $H(P)$ will increase as the distribution becomes more uniform.

According to the graph theory, the similarity values between samples from the same class tend to the same value. In other words, the values are roughly uniformly distributed. Therefore, the similarity matrix may be optimized by maximizing the Shannon entropy of the distribution. In section 3, we construct an adaptive graph for each class based on the analysis.

3. L1-Norm Distance Discriminant Analysis

As formulated in Eq. (1), RDSL considers inter-class separation and intra-class compactness. However, RDSL ignores intra-class structure. It is beneficial to exploit within-class structure information [12] and hence there is room to enhance the performance.

3.1. L1-Norm Distance Discriminant Analysis with Multiple Adaptive Graphs

Since the local geometry structure of data can be modeled by a nearest neighbor graph or a similarity matrix, we use multiple graphs to construct within-class structure. It is difficult to select the number of nearest neighbors of a graph. In order to avoid the difficulty, we intend to construct an adaptive graph for each class. According to the graph theory, the similarity between two samples is large when the samples belong to the same class with high probability. Therefore, it is reasonable to infer that any two samples from the same class have a roughly equal similarity value in most cases. Based on the inference, we use Shannon entropy to adaptively construct similarity matrices.

Our model is formulated as follows:

$$\min_{w_q,\xi_i,s_{kj}^i} \sum_{i=1}^{c}\sum_{j=1}^{n_i}\left\|w_q^T(x_{ij}-\bar{x}_i)\right\|_1 + \sum_{i=1}^{c}\alpha_i\left(\sum_{k=1}^{n_i}\sum_{j=1}^{n_i}s_{kj}^i\left\|w_q^T(x_{ik}-x_{ij})\right\|_1 + \eta_i s_{kj}^i\log s_{kj}^i\right) + \lambda\sum_{i=1}^{c}\xi_i$$

$$\text{s.t. } \left\|w_q^T(\bar{x}_i-\bar{x})\right\|_1 + \xi_i \geq 1$$

$$\xi_i \geq 0, \quad \sum_{j=1}^{n_i}s_{kj}^i = 1, \quad s_{kj}^i \geq 0, \quad i=1,2,\cdots,c, \quad k,j=1,2,\cdots,n_i \qquad (2)$$

where α_i and η_i are regularization parameters, $\alpha_i \geq 0$, $\eta_i \geq 0$, and s_{kj}^i is the similarity between the k th and j th samples from class i.

We use L1-norm to measure distances between samples from the same class, which may enhance the robustness of the graphs. We then use multiple adaptive graphs as a regularization term to improve the performance of RDSL. However, intra-class structure is only local geometry structure and our model does not involve global structure. Therefore, the performance may be further improved.

3.2. Improvement Strategy of the Model

The data reconstruction trick is usually applied to preserve the global structure of data [13]. However, the trick may be less than effective when the data are corrupted by noise, especially strong noise. In other words, it is necessary to eliminate the adverse effect of noise on data reconstruction. Inspired by the cutting L1-norm loss function [14], we use cutting squared L2-norm to construct a sample reconstruction term in order to realize the purpose.

The improved model is formulated as follows:

$$\min_{w_q, u_q, \xi_i, s_{kj}^i} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left\| w_q^T (x_{ij} - \bar{x}_i) \right\|_1 + \sum_{i=1}^{c} \alpha_i \left(\sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \left\| w_q^T (x_{ik} - x_{ij}) \right\|_1 + \eta_i s_{kj}^i \log s_{kj}^i \right)$$

$$+ \beta \sum_{i=1}^{c} \sum_{j=1}^{n_i} \min \left(h, \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 \right) + \lambda \sum_{i=1}^{c} \xi_i$$

s.t. $\left\| w_q^T (\bar{x}_i - \bar{x}) \right\|_1 + \xi_i \geq 1$

$$\xi_i \geq 0, \quad \sum_{j=1}^{n_i} s_{kj}^i = 1, \quad s_{kj}^i \geq 0, \quad i = 1, 2, \cdots, c, \quad k, j = 1, 2, \cdots, n_i \qquad (3)$$

where u_q is the q th reconstruction vector, $1 \leq q \leq d$, β is a regularization parameter, $\beta \geq 0$, and h is a given value, $h > 0$.

We do not use cutting L1-norm to measure reconstruction errors for it is easier to optimize the objective function. Although L2-norm is sensitive to noise, it is simple and convenient for optimization. Moreover, cutting squared L2-norm can constrain noisy samples and thus may become robust to noise.

4. Optimization Procedure

Eq. (3) involves the variables: w_q, u_q, s_{kj}^i and ξ_i, and we adopt an alternating iterative method [15] to obtain the solutions.

4.1. Fix u_q and s_{kj}^i and Update w_q and ξ_i

When u_q and s_{kj}^i are fixed, Eq. (3) reduces to the following form:

$$\min_{w_q, \xi_i} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left\| w_q^T (x_{ij} - \bar{x}_i) \right\|_1 + \sum_{i=1}^{c} \alpha_i \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \left\| w_q^T (x_{ik} - x_{ij}) \right\|_1$$

$$+ \beta \sum_{i=1}^{c} \sum_{j=1}^{n_i} \min \left(h, \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 \right) + \lambda \sum_{i=1}^{c} \xi_i$$

$$\text{s.t. } \left\| w_q^T (\bar{x}_i - \bar{x}) \right\|_1 + \xi_i \geq 1$$

$$\xi_i \geq 0, \quad i = 1, 2, \cdots, c \qquad (4)$$

In order to optimize Eq. (4), we first reformulate the first item of the objective function. Clearly,

$$\left\| w_q^T (x_{ij} - \bar{x}_i) \right\|_1 = \frac{w_q^T (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T w_q}{\left\| w_q^T (x_{ij} - \bar{x}_i) \right\|_1}$$

Let $\rho_{ij} = 1 / \left\| w_q^T (x_{ij} - \bar{x}_i) \right\|_1$, then $\sum_{i=1}^{c} \sum_{j=1}^{n_i} \left\| w_q^T (x_{ij} - \bar{x}_i) \right\|_1 = w_q^T B_{q1} w_q$, where

$$B_{q1} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \rho_{ij} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T .$$

Similarly, we reformulate the second item of the objective function and

$$\sum_{i=1}^{c} \alpha_i \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \left\| w_q^T (x_{ik} - x_{ij}) \right\|_1 = w_q^T B_{q2} w_q$$

where $B_{q2} = \sum_{i=1}^{c} \alpha_i \sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \mu_{kj}^i (x_{ik} - x_{ij})(x_{ik} - x_{ij})^T$, $\mu_{kj}^i = 1 / \left\| w_q^T (x_{ik} - x_{ij}) \right\|_1$.

Moreover, we reformulate the third item of the objective function, which can be written as

$$\min \left(h, \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 \right) = (1 - \theta_{ij}) h + \theta_{ij} \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2$$

where $\theta_{ij} = 1$ if $\left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 < h$, otherwise $\theta_{ij} = 0$. Clearly,

$$\left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 = x_{ij}^T x_{ij} - 2 x_{ij}^T u_q x_{ij}^T w_q + w_q^T x_{ij} u_q^T u_q x_{ij}^T w_q$$

Therefore, we have

$$\sum_{i=1}^{c} \sum_{j=1}^{n_i} \min \left(h, \left\| x_{ij} - u_q w_q^T x_{ij} \right\|_2^2 \right) = b_{q1} + a_{q1}^T w_q + w_q^T G_{q1} w_q$$

where

$$b_{q1} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left((1 - \theta_{ij}) h + \theta_{ij} x_{ij}^T x_{ij} \right)$$

$$a_{q1} = -2 \sum_{i=1}^{c} \sum_{j=1}^{n_i} \theta_{ij} (x_{ij}^T u_q) x_{ij}$$

$$G_{q1} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \theta_{ij} x_{ij} u_q^T u_q x_{ij}^T$$

Finally, we reformulate the constrain of the objective function. Clearly,

$$\left\| w_q^T (\bar{x}_i - \bar{x}) \right\|_1 = \text{sign} \left(w_q^T (\bar{x}_i - \bar{x}) \right) w_q^T (\bar{x}_i - \bar{x}) = \text{sign} \left(w_q^T (\bar{x}_i - \bar{x}) \right) (\bar{x}_i - \bar{x})^T w_q$$

where $\text{sign}(\cdot)$ is the sign function. Let e_c denote the c dimensional column vector which includes c ones, and $\gamma = [\xi_1, \xi_2, \cdots, \xi_c]^T$. Let $A_q = [\upsilon_1, \upsilon_2, \cdots, \upsilon_c]^T$, where $\upsilon_i = \text{sign} \left(w_q^T (\bar{x}_i - \bar{x}) \right) (\bar{x}_i - \bar{x})$, Then $\left\| w_q^T (\bar{x}_i - \bar{x}) \right\|_1 + \xi_i \geq 1$, $i = 1, 2, \cdots, c$, can be reformulated as follows: $A_q w_q + \gamma \geq e_c$.

Therefore, Eq. (4) can be reformulated as

$$\min_{w_q, \gamma} w_q^T \left(B_{q1} + B_{q2} + \beta G_{q1} \right) w_q + \beta a_{q1}^T w_q + \lambda e_c^T \gamma$$

$$\text{s.t. } A_q w_q + \gamma \geq e_c , \gamma \geq 0 \qquad (5)$$

Eq. (5) is a quadratic programming problem, which can be easily solved.

4.2. Fix w_q, ξ_i and s_{kj}^i and Update u_q

When w_q, ξ_i and s_{kj}^i are fixed, Eq. (3) reduces to the following form:

$$\min_{u_q} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \min\left(h, \left\|x_{ij} - u_q w_q^T x_{ij}\right\|_2^2\right) \tag{6}$$

Eq. (6) can be reformulated as follows:

$$\min_{u_q} g_{q2} u_q^T u_q + a_{q2}^T u_q + b_{q2} \tag{7}$$

where

$$b_{q2} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left((1-\theta_{ij})h + \theta_{ij} x_{ij}^T x_{ij}\right)$$

$$a_{q2} = -2\sum_{i=1}^{c} \sum_{j=1}^{n_i} \theta_{ij}\left(x_{ij}^T w_q\right)x_{ij}$$

$$g_{q2} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \theta_{ij}\left(x_{ij}^T w_q\right)^2$$

Eq. (7) is also a quadratic programming problem and the solution can be easily obtained.

4.3. Fix w_q, ξ_i and u_q and Update s_{kj}^i

When w_q, ξ_i and u_q are fixed, Eq. (3) reduces to the following form:

$$\min_{s_{kj}^i} \sum_{i=1}^{c} \alpha_i \left(\sum_{k=1}^{n_i} \sum_{j=1}^{n_i} s_{kj}^i \left\|w_q^T(x_{ik} - x_{ij})\right\|_1 + \eta_i s_{kj}^i \log s_{kj}^i\right)$$

$$\text{s.t.} \sum_{j=1}^{n_i} s_{kj}^i = 1,\ s_{kj}^i \geq 0,\ i=1,2,\cdots,c,\ k,j=1,2,\cdots,n_i \tag{8}$$

Eq. (8) can be easily solved and

$$s_{kj}^i = \frac{\exp\left(-\left\|w_q^T(x_{ik} - x_{ij})\right\|_1 / \eta_i\right)}{\sum_{j=1}^{n_i} \exp\left(-\left\|w_q^T(x_{ik} - x_{ij})\right\|_1 / \eta_i\right)} \tag{9}$$

It is worth noting that w_q can be initiated by other methods, such as LDA, and obtained by an iterative algorithm. After w_q is obtained, training samples are updated by the factor $\left(I_m - WW^T\right)$, where I_m is the $m \times m$ dimensional identity matrix [10]. For Eq. (5), since an iterative technique is used, the next value of w_q can be obtained with the current one. Namely, B_{q1}, B_{q2}, G_{q1} and a_{q1} are also fixed. Similarly, for Eq. (7), the next value of u_q can be obtained with its current one. The alternating iterative algorithm is summarized in Table 1.

Table 1. Algorithm 1. The alternating iterative algorithm

Input: X --- labeled data set, $X \in R^{m \times n}$;

α_i , η_i , β , λ --- regularization parameters;

d --- number of projection vectors ;

Output: W --- projection matrix, $W \in R^{m \times d}$.

Procedure:

Initiate the projection matrix $W = \phi$, where ϕ is the empty set;

For $q = 1$ **to** d do

Initiate w_q ;

Compute u_q according to Eq. (7);

Compute s_{kj}^i according to Eq. (9);

While not converge **do**

Update w_q according to Eq. (5);

Update u_q according to Eq. (7);

Update s_{kj}^i according to Eq. (9);

End while

Let $W = [W, w_q]$, and normalize and orthogonalize it;

Update the data set X , i.e., $X = \left(I_m - WW^T \right) X$;

End for

5. Experiments

In this section, we evaluate the classification accuracy of our method, and LDA, coined complete large margin linear discriminant analysis (CLMLDA) [16], RILDA, and RDSL are also run respectively on three real datasets, namely the USPS database [17], the COIL-20 database[2] and the Yale database[3]. For the datasets, the grayscale of images is normalized to $[0,1]$. For simplicity, PCA is first used to reduce the dimensionality to $n - c$ after the images are represented with vectors. For CLMLDA, the optimal values of two parameter are chosen from the set $[10^{-6}, 10^{-3}, 0, 10^3, 10^6]$. For RDSL, the optimal value of λ is chosen from the same set. For our method, η_i and β are empirically set to 10^3 and 0.1, respectively, and other regularization parameters are chosen from the set. The cutting parameter h is empirically set and as a result, five percent of training samples with the largest reconstruction errors are eliminated. w_q is initiated by the solution of LDA and the convergence condition is in accordance with [10]. The nearest neighbor classifier is used to measure classification accuracy and directly used on the preprocessed data by PCA as the baseline method. We repeat all the experiments ten times and obtain average lowest classification error rates.

---

[2] https://www1.cs.columbia.edu/CAVE/software/softlib/coil-20.php

[3] http://vision.ucsd.edu/content/yale-face-database

5.1. Experiments on the USPS Database

The USPS database consists of 9,298 grayscale images of digits from envelopes and their size is 16×16 pixels, namely each image can be described as a 256-dimensional vector. For each digit, we randomly select 100 samples and construct a subset for experiments. Furthermore, the images in the subset are artificially contaminated by the
Gaussian noise at a random position with the proportion of forty percent. In the experiments, we randomly select 8, 10, 12 and 14 images as training samples, respectively, and the rest images as testing samples.

Average lowest classification error rates with different dimensions are given in Figure 1, which shows that our method slightly outperforms RDSL.
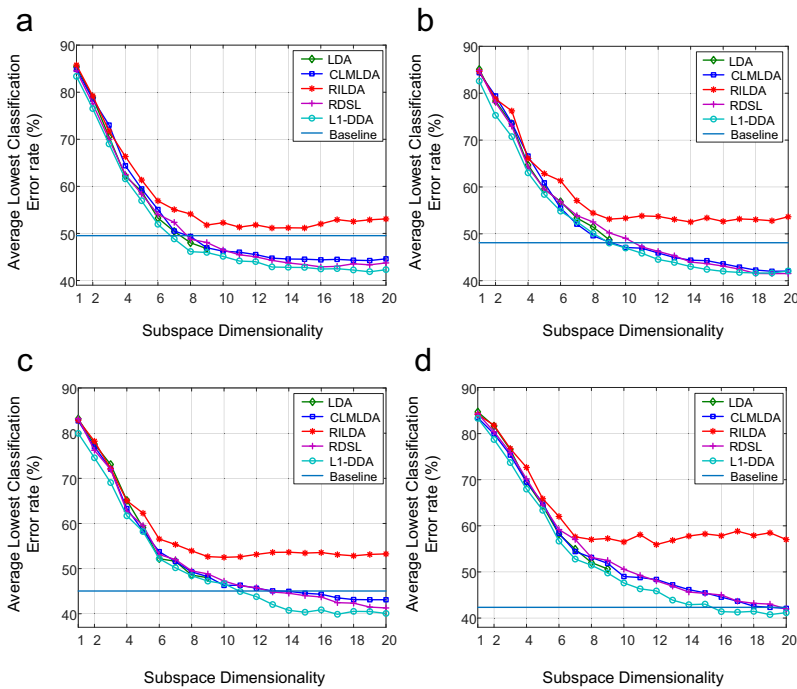

Figure 1. Average lowest classification error rates on the USPS database. (a) 8 samples, (b) 10 samples, (c) 12 samples, and (d) 14 samples.

5.2. Experiments on the Coil-20 Database

There are 1,440 grayscale images from twenty objects in the COIL-20 database and each object has 72 images which are taken at an interval of five degrees. In the experiments, the images are downsampled and their size becomes 32×32 pixels. We then add the Gaussian noise to each image at a random position with the proportion of forty percent. For each object, 6, 8, 10 and 12 images are randomly selected as training samples, respectively, and the rest images are used for testing.

Average lowest classification error rates with different dimensions are shown in Figure 2. From the figure we conclude that the performance of L1-DDA is much better than that of RDSL.
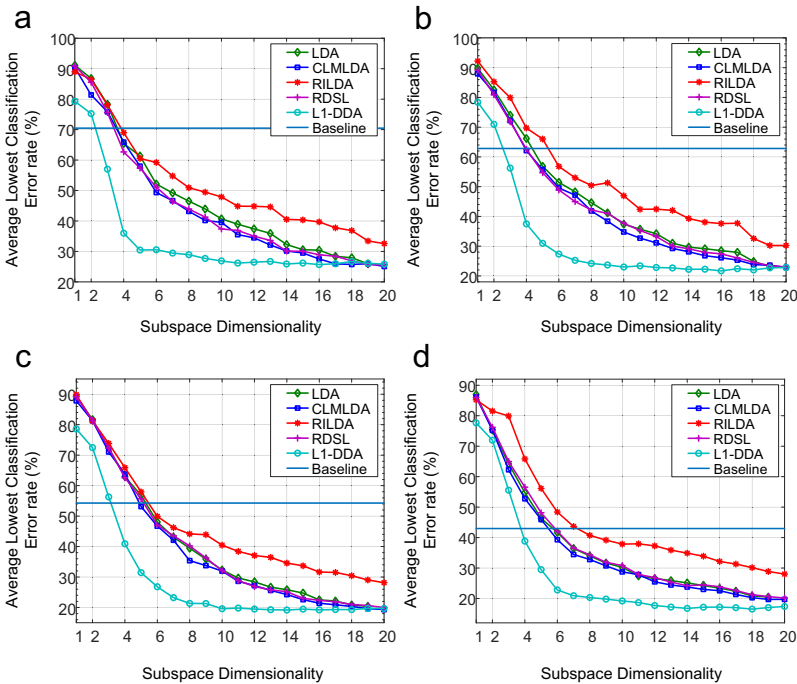

Figure 2. Average lowest classification error rates on the COIL-20 database. (a) 6 samples, (b) 8 samples, (c) 10 samples, and (d) 12 samples.

5.3. Experiments on the Yale Database

The Yale database consists of 165 grayscale face images and the images are from fifteen subjects and different in three aspects: facial expression, lighting condition and whether or not wearing glasses. In the experiments, we use the processed images whose size is 32×32 pixels[4]. Similar to the experiments on the USPS and COIL-20 databases, we use the Gaussian noise to contaminate each image at a random position with the proportion of forty percent. Moreover, we randomly select 5, 6, 7 and 8 images from each subject as training samples, respectively, and the rest ones as testing samples.

Average lowest classification error rates with different dimensions are given in Figure 3. From the experimental results, we can conclude that the proposed method is significantly effective for classification.
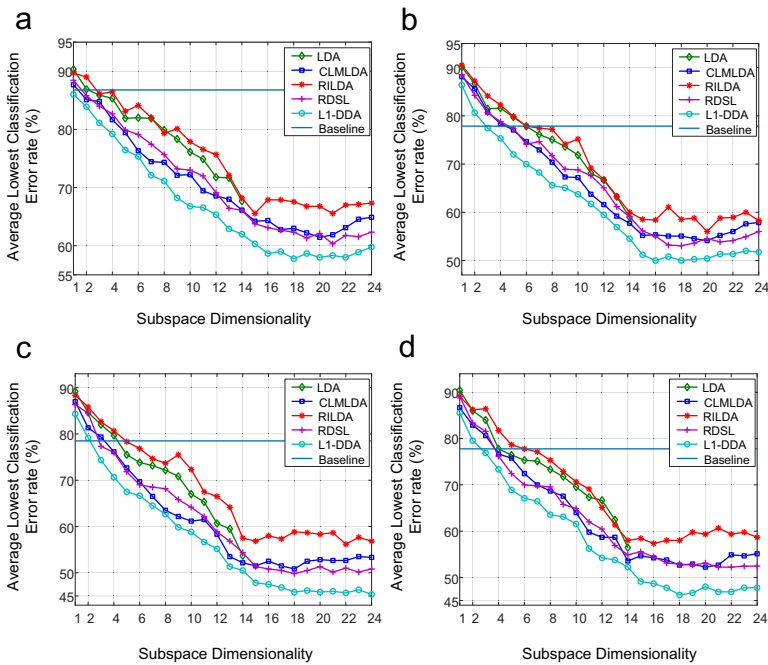
---

[4] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

Figure 3. Average lowest classification error rates on the Yale database. (a) 5 samples, (b) 6 samples, (c) 7 samples, and (d) 8 samples.

6. Conclusions

In this paper, we propose L1-norm distance discriminant analysis with multiple adaptive graphs and sample reconstruction (L1-DDA) based on RDSL. Our method uses multiple adaptive graphs to preserve intra-class manifold structure and applies L1-norm and Shannon entropy to robustly and adaptively construct the graphs. Moreover, it uses the sample reconstruction technique to preserve the global structure of data and cutting squared L2-norm is applied to constrain noisy samples for better robustness. Finally, we use an alternating iterative technique to obtain projection vectors. Experimental results on three real datasets show that our method obtains better classification performance than RDSL.

Acknowledgement

References

[1] Duda RO, Hart PE, Stork DG, editors. Pattern classification. 2nd ed. NewYork: Wiley Press; c2005.

[2] Hastie T, Tibshirani R, Friedman J, editors. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Berlin: Springer Press; c2001.

[3] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell. 1997 Jul;19(7):711-20.

[4] Jolliffe IT, editors. Principal component analysis. 2nd ed. New York: Springer-Verlag Press; c2002.

[5] Lai Z, Xu Y, Yang J, Shen L, Zhang D. Rotational invariant dimensionality reduction algorithms. IEEE Trans Cybern. 2017 Nov;47(11):3733-46.

[6] Zhong F, Zhang J. Linear discriminant analysis based on L1-norm maximization. IEEE Trans Image Process. 2013 Aug;22(8):3018-27.

[7] Zheng W, Lin Z, Wang H. L1-norm kernel discriminant analysis via Bayes error bound optimization for robust feature extraction. IEEE Trans Neural Netw Learn Syst. 2014 Apr;25(4):793-805.

[8] Liu Y, Gao Q, Miao S, Gao X, Nie F, Li Y. A non-greedy algorithm for L1-norm LDA. IEEE Trans Image Process. 2017 Feb;26(2):684-95.

[9] Ye Q, Yang J, Liu F, Zhao C, Ye N, Yin T. L1-norm distance linear discriminant analysis based on an effective iterative algorithm. IEEE Trans Circuits Syst Video Technol. 2018 Jan;28(1):114-29.

[10] Zhang D, Sun Y, Ye Q, Tang J. Recursive discriminative subspace learning with L1-norm distance constraint. IEEE Trans Cybern. 2020 May;50(5):2138-51.

[11] Fukunnaga K, editors. Introduction to statistical pattern recognition. New York: Academic Press; c1991.

[12] Hu L, Zhang W. Orthogonal neighborhood preserving discriminant analysis with patch embedding for face recognition. Pattern Recogn. 2020 Oct;106(10):1579-94.

[13] Wen J, Han N, Fang X, Fei L, Yan K, Zhan S. Low-rank preserving projection via graph regularized reconstruction. IEEE Trans Cybern. 2019 Apr;49(4):1279-91.

[14] Cevikalp H. Best fitting hyperplanes for classification. IEEE Trans Pattern Anal Mach Intell. 2017 Jun;39(6):1076-88.

[15] Li X, Zhang H, Zhang R, Liu Y, Nie F. Generalized uncorrelated regression with adaptive graph for unsupervised feature selection. IEEE Trans Neural Netw Learn Syst. 2019 May;30(5):1587-95.

[16] Chen X, Yang J, Zhang D, Liang J. Complete large margin linear discriminant analysis using mathematical programming approach. Pattern Recogn. 2013 Jun;46(6):1579-94.

[17] Simard P, LeCun Y, Denker JS. Efficient pattern recognition using a new transformation distance. In: Hanson S, Cowan J, Giles C, editors. Advances in Neural Information Processing Systems; 1992 Nov 30- Dec 3; San Francisco: Morgan-Kaufmann Press; c1993. p. 50-8.