ICIC Express Letters
Volume **10**, Number **9**, September **2016**

ICIC International ©2016 ISSN 1881-803X
pp. 2097–2104

A PROBABILISTIC METHOD OF LINEAR DISCRIMINANT ANALYSIS WITH RANDOMIZED INPUT

Congcong Xiong[1], Xin Tao[1], Yarui Chen[1,*], Ying Deng[1], Yancui Shi[1] and Xiaoman Zhang[2]

[1]College of Computer Science and Information Engineering
Tianjin University of Science and Technology
No. 1038, Dagu Nanlu, Hexi District, Tianjin 300222, P. R. China
*Corresponding author: yrchen@tust.edu.cn

[2]State Grid Jibei Electric Economic Research Institute
Beijing 100038, P. R. China

Received January 2016; accepted April 2016

Abstract. *Linear discriminant analysis (LDA) is one of the effective methods for feature extraction and dimensionality reduction in most linear classification problems. However, the recognition performance of traditional LDA is poor on some samples with incomplete feature information or small sized data. Motivated by the success of extreme learning machine and probabilistic generative model, the paper presents a probabilistic method of linear discriminant analysis with randomized input (PLDA-R). As feature information of sample is incomplete, LDA with utilization of randomized input could improve the classification accuracy since it generates a randomized space, where samples are projected onto this space and feature information could be enlarged. Moreover, probabilistic generative model is introduced as a classifier, which gives a probabilistic discrimination result, and will enhance the classification effect on small sized training samples. These conclusions are confirmed by extensive experiments on various datasets.*

Keywords: LDA, Randomized input, Probabilistic generative model, KNN

1. **Introduction.** Traditional LDA, also known as Fisher criterion, was proposed by Fisher in 1936 [1,2], which is also one of effective algorithms for dimensionality reduction and feature extraction in many pattern recognition applications. The key idea of Fisher criterion is to find an optimal projection (or transformation matrix) $\mathbf{W}$ by maximizing the radio of the between-class scatter matrix $\mathbf{S}_b$ to the within-class scatter matrix $\mathbf{S}_w$, which can be completely expressed as $J(\mathbf{W}) = \left|\mathbf{W}^{\mathrm{T}}\mathbf{S}_b\mathbf{W}\right| / \left|\mathbf{W}^{\mathrm{T}}\mathbf{S}_w\mathbf{W}\right|$, where $J(\mathbf{W})$ is defined as loss function [3]. The transformation matrix $\mathbf{W} \in \mathbb{R}^{d\times h}$ $(h < d)$ is determined by eigenvectors corresponding to the $k-1$ largest eigenvalues of $\mathbf{S}_w^{-1}\mathbf{S}_b$ by eigenvalue decomposition (EVD), where k is the number of samples class [4]. Particularly, $\mathbf{S}_w$ and $\mathbf{S}_b$ need to be non-singular [5]. Actually, the loss of feature information may occur in the procedure of collecting or processing samples, and the recognition performance of traditional LDA often suffers from these samples with incomplete feature information. Therefore, this paper proposes a new method of LDA with randomized input, which is motived by extreme learning machine [6]. This method sets up a q-nodes hidden layer, and randomly determines the input weights matrix and bias matrix linking the hidden layer and the output layer to generate a randomization space where samples are projected on it, and feature information of original sample is enlarged. Then we combine randomization procedure with traditional LDA to extract feature information.

After the feature extraction, training samples will be classified using appropriate classifiers, such as k nearest neighbor (KNN) and Naive Bayes [7]. For the case of small number of samples, Naive Bayes classifier is superior to KNN classifier since the latter is easy to over-fit. However, as the number of training samples increases, KNN classifier will be better due to a lower asymptotic error so that the former is not sufficient to provide an accurate classification model [8,9]. Motived by above analysis, we present a classifier algorithm based on probabilistic generative model. This algorithm gives a new perspective to classify samples by the methods of probability, which gives a probabilistic discrimination result, and will enhance the classification effect on small size training samples.

The later chapters of this paper are organized as follows. Section 2 reviews the procedure of LDA algorithm. In Section 3, we make a detailed description of PLDA-R. Section 4 presents numerical experiments to show the efficiency of the PLDA-R on the standard data sets of UCI. Finally, Section 5 summarizes our work with some considerations on future directions.

2. Linear Discriminant Analysis.

For dataset $\{\mathbf{x}_i, t_i\}_{i=1}^{n}$, $\mathbf{x}_i \in \mathbb{R}^d$, $t_i \in \{1, 2, \ldots, k\}$, n is the total number of samples, and k is the number of classes. The dataset can be partitioned into k subsets $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k]$, where $\mathbf{X}_j$ belonging to class j and consisting of n_j number of samples can be described by $\mathbf{X}_j \in \mathbb{R}^{d \times n_j}$, and $\sum_{j=1}^{k} n_j = n$. The optimal projection (or transformation matrix) $\mathbf{W} \in \mathbb{R}^{d \times h}$ $(h < d)$ will be obtained by LDA that projects sample from d-dimensional space $\mathbf{x}_i \in \mathbb{R}^d$ to h-dimensional space $\mathbf{x}_i^L \in \mathbb{R}^h$ as the following

$$\mathbf{x}_i^L = \mathbf{W}^{\mathrm{T}} \mathbf{x}_i. \tag{1}$$

In order to obtain transformation matrix $\mathbf{W}$, we introduce the Fisher criterion here:

$$\mathbf{W}^{LDA} = \max \left(\frac{|\mathbf{S}_b^L|}{|\mathbf{S}_w^L|} \right) = \arg \max_{\mathbf{W}} \frac{|\mathbf{W}^{\mathrm{T}} \mathbf{S}_b \mathbf{W}|}{|\mathbf{W}^{\mathrm{T}} \mathbf{S}_w \mathbf{W}|}, \tag{2}$$

where $\mathbf{S}_w$, $\mathbf{S}_b$ are respectively called *within-class* scatter matrix and *between-class* scatter matrix. These scatters are further defined as follows:

$$\mathbf{S}_w = \sum_{j=1}^{k} \sum_{\mathbf{x}_i \in \mathbf{X}_j} (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^{\mathrm{T}}, \tag{3}$$

$$\mathbf{S}_b = \sum_{j=1}^{k} n_j (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^{\mathrm{T}}, \tag{4}$$

where $\mathbf{m}_j$ is the *centroid* of the j-th class samples and $\mathbf{m}$ is *globe centroid* of all samples [10].

The optimal projection $\mathbf{W}$ is obtained to reach the maximum of $\mathbf{W}^{LDA}$ by method of Lagrange multiplier, and the final result is given by $\mathbf{W} \propto \mathbf{S}_w^{-1} \mathbf{S}_b$. Therefore, the $\mathbf{W}$ is further computed by the EVD of $\mathbf{S}_w^{-1} \mathbf{S}_b$, and retains the top $k - 1$ eigenvectors corresponding to the nonzero eigenvalues since $\mathbf{S}_b$ has rank at most equal to $k - 1$ [2].

3. Probabilistic LDA with Randomized Input.

In this section, we describe the PLDA-R algorithm with a two-stage procedure where all samples will be preprocessed by randomized input and extracted feature information using RLDA firstly, and then projected samples are classified by the method of probabilistic generative model. The schematic diagram of PLDA-R is shown in Figure 1.

3.1. LDA with randomized input.

For dataset $\{\mathbf{x}_i, t_i\}_{i=1}^{n}$, $\mathbf{x}_i \in \mathbb{R}^d$, $t_i \in \{1, 2, \ldots, k\}$, n is the total number of samples, and k is the number of classes. We define the randomization function $g(\mathrm{x})$ as follows:

$$g(\mathbf{u}_r, \mathbf{b}_r, \mathbf{x}_i) = \mathbf{u}_r \mathbf{x}_i + \mathbf{b}_r, \tag{5}$$

where $\mathbf{u}_r \in \mathbb{R}^{q\times d}$ is input weight matrix that connects to the q-th hidden node, $\mathbf{b}_r \in \mathbb{R}^{q\times n}$ is the bias matrix of the q-th hidden node and $r \in \{1, 2, \ldots, q\}$. Then the output function $\mathbf{\Phi}(\mathbf{x}_i)$ is defined as:

$$\mathbf{\Phi}(\mathbf{x}_i) = [g(\mathbf{u}_1, b_1, \mathbf{x}_i), g(\mathbf{u}_2, b_2, \mathbf{x}_i), \ldots, g(\mathbf{u}_q, b_q, \mathbf{x}_i)], \tag{6}$$

which aims to enlarge feature space of sample $\mathbf{x}_i$ from d-dimension to q-dimension. In addition, $\mathbf{u}_r$ and $\mathbf{b}_r$ are random matrix which are determined by random number in the range of $[0, 1]$. A nonlinear processing would be carried out for sample $\mathbf{x}_i$ using sigmoid function $F(\mathbf{\Phi}(\mathbf{x}_i))$:

$$\hat{\mathbf{x}}_i = F(\mathbf{\Phi}(\mathbf{x}_i)) = \frac{1}{1 + e^{-\varphi\times\mathbf{\Phi}(\mathbf{x}_i)}}, \tag{7}$$

where $\hat{\mathbf{x}}_i \in \mathbb{R}^q$ denotes output sample after randomization, and regularization parameter φ in activation function and hidden node number q are both artificially given. Then the output samples $\hat{\mathbf{x}}_i$ will be further projected from q-dimensional space $\hat{\mathbf{x}}_i \in \mathbb{R}^q$ to h-dimensional space $\hat{\mathbf{x}}_i^L \in \mathbb{R}^h$ $(h < q)$ using traditional LDA as follows:

$$\hat{\mathbf{x}}_i^L = \mathbf{W}^{\mathrm{T}}\hat{\mathbf{x}}_i, \tag{8}$$

where $\hat{\mathbf{x}}_i^L$ denotes output sample in optimal projection space. These samples would be classified using probabilistic generative model later.

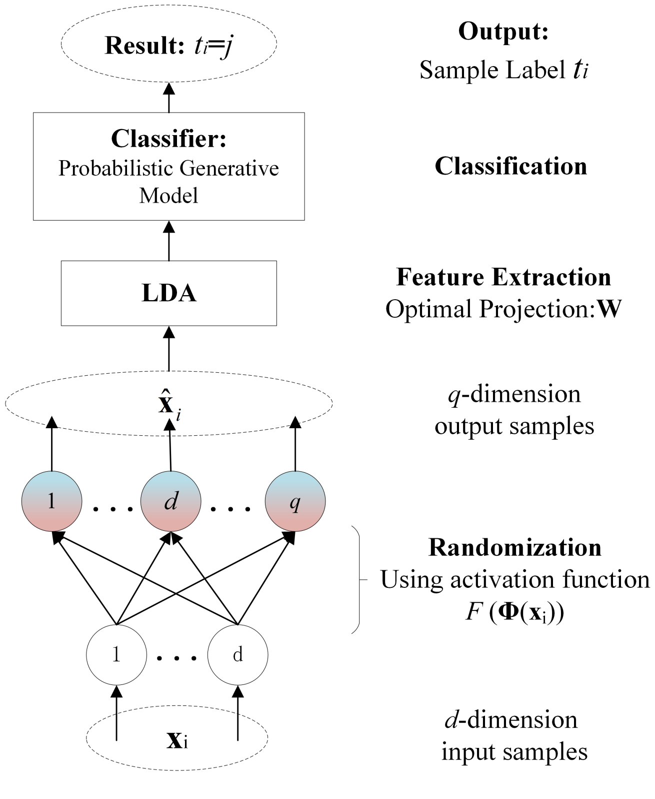


Figure 1. Schematic diagram of PLDA-R algorithm

3.2. **Classifier: Probabilistic generative model.** In this paper, we present a classifier based on probabilistic generative model. For the dataset $\left\{\hat{\mathbf{x}}_i^L, t_i\right\}_{i=1}^n$, $\hat{\mathbf{x}}_i^L \in \mathbb{R}^h$, the paper adopts a generative approach in which we model the class-conditional probabilistic densities $P\left(\hat{\mathbf{x}}_i^L|C_j\right)$ and the class priors $P(C_j)$, and then use them to compute posterior probabilities $P\left(C_j|\hat{\mathbf{x}}_i^L\right)$. The classification result is determined by the maximum of class posterior probabilities $P\left(C_j|\hat{\mathbf{x}}_i^L\right)$. In order to simplify calculation, the paper assumes that the class-conditional probabilistic densities are Gaussian distribution and that all classes share the same covariance matrix $\mathbf{S}$, and then $P\left(C_j|\hat{\mathbf{x}}_i^L\right)$ is given by:

$$P\left(C_j|\hat{\mathbf{x}}_i^L\right) = \frac{P\left(\hat{\mathbf{x}}_i^L|C_j\right) P\left(C_j\right)}{\sum\limits_{l=1}^{k} P\left(\hat{\mathbf{x}}_i^L|C_l\right) P\left(C_l\right)}, \tag{9}$$

where $P\left(\hat{\mathbf{x}}_i^L|C_j\right)$ with Gaussian distribution denotes

$$P\left(\hat{\mathbf{x}}_i^L|C_j\right) = \frac{1}{(2\pi)^{1/2}} \frac{1}{|\mathbf{S}|^{1/2}} \exp\left\{-\frac{1}{2}\left(\hat{\mathbf{x}}_i^L - \mathbf{m}_j\right)^{\mathrm{T}} \mathbf{S}^{-1} \left(\hat{\mathbf{x}}_i^L - \mathbf{m}_j\right)\right\}, \tag{10}$$

and $P\left(C_j\right)$ is

$$P\left(C_j\right) = \frac{n_j}{n}. \tag{11}$$

By maximizing parameters with centroid of each class $\mathbf{m}_j$ and covariance matrix $\mathbf{S}$ respectively using maximum likelihood estimation, we will obtain that

$$\mathbf{m}_j = \frac{1}{n_j} \sum_{\hat{\mathbf{x}}_i \in \mathbf{X}_j} \hat{\mathbf{x}}_i^L, \tag{12}$$

$$\mathbf{S} = \frac{n_1}{n}\mathbf{S}_1 + \frac{n_2}{n}\mathbf{S}_2 + \cdots + \frac{n_k}{n}\mathbf{S}_k, \tag{13}$$

where

$$\mathbf{S}_j = \frac{1}{n_j} \sum_{\hat{\mathbf{x}}_i \in \mathbf{X}_j} \left(\hat{\mathbf{x}}_i^L - \mathbf{m}_j\right) \left(\hat{\mathbf{x}}_i^L - \mathbf{m}_j\right)^{\mathrm{T}} \tag{14}$$

is equivalent to the variance of the j-th class samples. The posterior probabilities $P\left(C_j|\hat{\mathbf{x}}_i^L\right)$ can be computed by Formula (9), and we select the highest posterior probability corresponding to sample class as class label t_i for each test sample. In summary, PLDA-R algorithm could be expressed as extracting feature using LDA with randomized samples and classifying samples with probabilistic generative model, and the pseudo code description of it is illustrated as **Algorithm 1**.

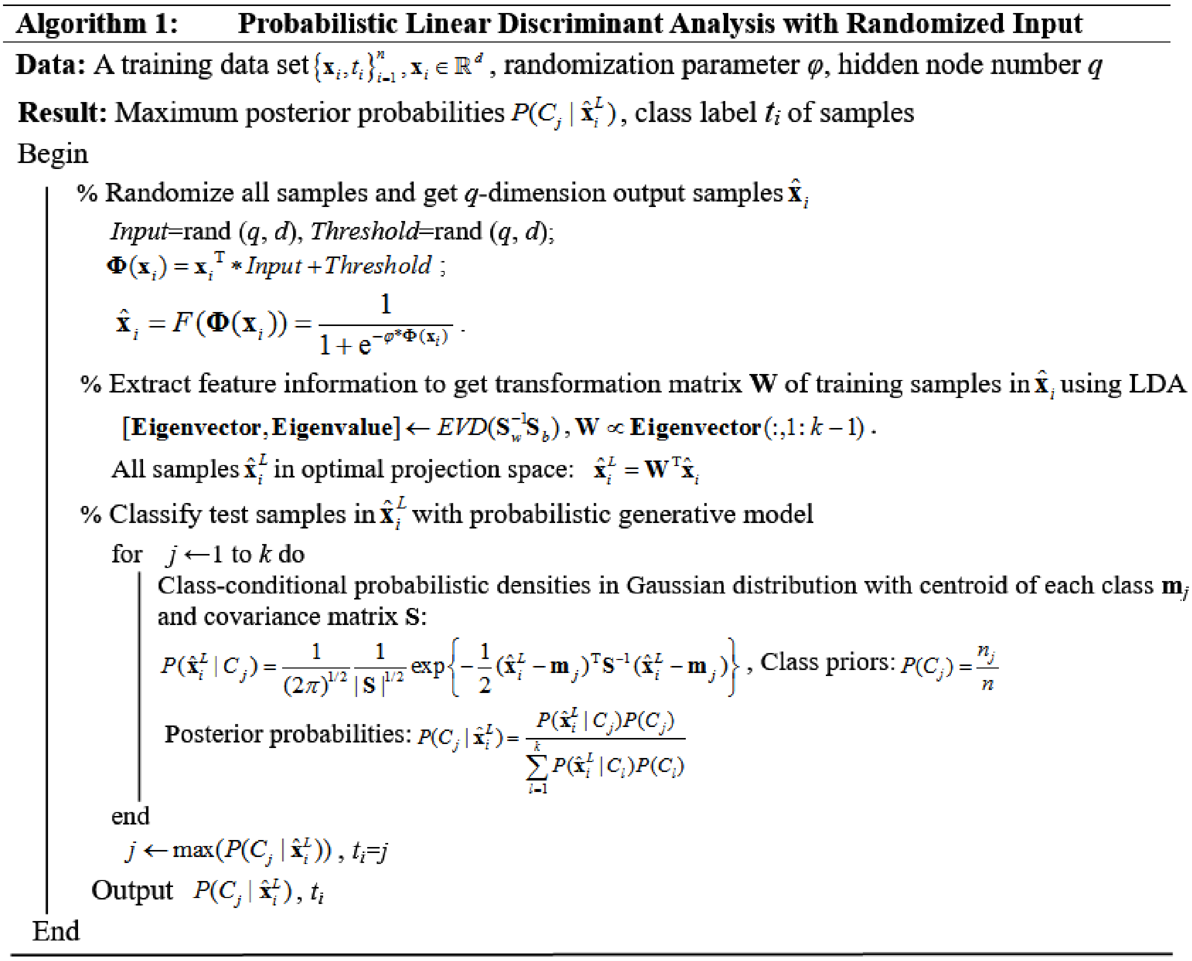
4. **Experiments.**

In this section, we present some experiments of PLDA-R algorithm with data sets that contain Iris, Wine and Banknote from UCI. In order to comprehensively examine the classification performance of proposed algorithm, we divided it into traditional LDA with randomized input (RLDA) and PLDA-R where the former does not use probabilistic generative model and latter does. The experiments are conducted to compare the proposed algorithm with other classifiers such as KNN and Naive Bayes. The more information of data sets is shown in Table 1.

Table 1. Experimental data

Datasets	Class	Dimension	Number of available samples
Iris	3	4	150([50,50,50])
Banknote	2	4	1372([762,610])
Wine	3	13	178([47,57,38])

The setting up of randomized input technology is described as follows. Regularization parameter φ and number of hidden node q in activation function are required to be given artificially. The optimal value of φ should be controlled in the range of $[0, 1]$. The recognition accuracy is obtained by the average of 50 times experimental results for the algorithm involving randomization procedure. The values of these variables are shown in Table 2 and Table 3.



```
Algorithm 1:    Probabilistic Linear Discriminant Analysis with Randomized Input
Data: A training data set $\{\mathbf{x}_i, t_i\}_{i=1}^{n}, \mathbf{x}_i \in \mathbb{R}^d$, randomization parameter $\varphi$, hidden node number $q$
Result: Maximum posterior probabilities $P(C_j \mid \hat{\mathbf{x}}_i^L)$, class label $t_i$ of samples
Begin
    % Randomize all samples and get $q$-dimension output samples $\hat{\mathbf{x}}_i$
      $Input$=rand $(q, d)$, $Threshold$=rand $(q, d)$;
      $\mathbf{\Phi}(\mathbf{x}_i) = \mathbf{x}_i^{\mathrm{T}} * Input + Threshold$ ;
      $\hat{\mathbf{x}}_i = F(\mathbf{\Phi}(\mathbf{x}_i)) = \dfrac{1}{1 + \mathrm{e}^{-\varphi * \mathbf{\Phi}(\mathbf{x}_i)}}$ .
    % Extract feature information to get transformation matrix $\mathbf{W}$ of training samples in $\hat{\mathbf{x}}_i$ using LDA
      $[\mathbf{Eigenvector}, \mathbf{Eigenvalue}] \leftarrow EVD(\mathbf{S}_w^{-1}\mathbf{S}_b)$, $\mathbf{W} \propto \mathbf{Eigenvector}(:, 1 : k-1)$ .
      All samples $\hat{\mathbf{x}}_i^L$ in optimal projection space:  $\hat{\mathbf{x}}_i^L = \mathbf{W}^{\mathrm{T}}\hat{\mathbf{x}}_i$
    % Classify test samples in $\hat{\mathbf{x}}_i^L$ with probabilistic generative model
      for  $j \leftarrow 1$ to $k$ do
          Class-conditional probabilistic densities in Gaussian distribution with centroid of each class $\mathbf{m}_j$
          and covariance matrix $\mathbf{S}$:
          $P(\hat{\mathbf{x}}_i^L \mid C_j) = \dfrac{1}{(2\pi)^{1/2}} \dfrac{1}{|\mathbf{S}|^{1/2}} \exp\left\{ -\dfrac{1}{2} (\hat{\mathbf{x}}_i^L - \mathbf{m}_j)^{\mathrm{T}} \mathbf{S}^{-1} (\hat{\mathbf{x}}_i^L - \mathbf{m}_j) \right\}$, Class priors: $P(C_j) = \dfrac{n_j}{n}$
          Posterior probabilities: $P(C_j \mid \hat{\mathbf{x}}_i^L) = \dfrac{P(\hat{\mathbf{x}}_i^L \mid C_j) P(C_j)}{\sum_{i=1}^{k} P(\hat{\mathbf{x}}_i^L \mid C_i) P(C_i)}$
      end
      $j \leftarrow \max(P(C_j \mid \hat{\mathbf{x}}_i^L))$ , $t_i$=$j$
    Output  $P(C_j \mid \hat{\mathbf{x}}_i^L)$, $t_i$
End
```

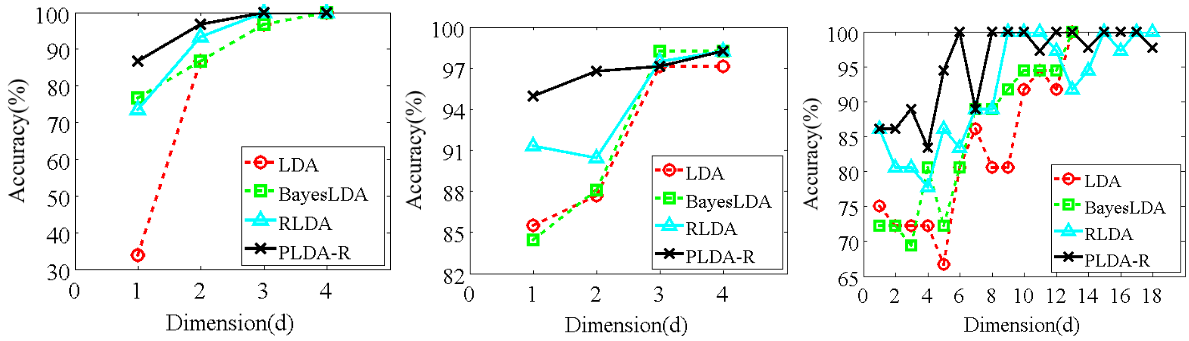
Table 2. Recognition accuracy (in percentage) with different sample dimensions

Datasets	LDA+KNN		LDA+Bayes		RLDA+KNN		PLDA-R	
	Dim	Accu	Dim	Accu	Dim	Accu	Dim	Accu
Iris ($\varphi = 0.1$, $q = 8$)	1	33.70	1	76.67	1	73.33	1	**86.67**
	2	86.67	2	86.67	2	93.33	2	**96.67**
	3	96.67	3	96.67	3	**100**	3	**100**
	4	**100**	4	**100**	4	**100**	4	**100**
Banknote ($\varphi = 0.1$, $q = 8$)	1	85.51	1	84.42	1	91.30	1	**94.93**
	2	87.68	2	88.04	2	90.42	2	**96.74**
	3	97.10	3	**98.19**	3	97.46	3	97.10
	4	97.10	4	**98.19**	4	**98.19**	4	**98.19**
Wine ($\varphi = 0.01$, $q = 18$)	2	72.22	2	72.22	2	80.56	2	**86.11**
	4	72.22	4	80.56	4	77.78	4	**83.33**
	8	80.56	8	88.89	8	88.89	8	**97.22**
	13	**100**	13	**100**	18	**100**	18	97.22

4.1. **PLDA-R with different sample dimensions.** In order to verify the advantage of proposed algorithm on samples with incomplete feature information, the paper selects 80% of all samples as training set and the remaining for test, and evaluates them with training samples in initial dimensions, and major results are described in Table 2 (the highest classification accuracy are depicted in bold font). It can be observed from Table 2 that recognition accuracy with randomization procedure has significantly improved in different dimensions on three data sets. Particularly, the recognition accuracy of methods

TABLE 3. Recognition accuracy with different proportion training samples

Datasets	Proportion	LDA+KNN	LDA+Bayes	RLDA+KNN	PLDA-R
Iris ($\varphi = 0.1$, $q = 8$)	0.2	95.00%	89.17%	95.50%	**96.67%**
	0.4	95.56%	**97.78%**	96.67%	96.67%
	0.6	96.67%	68.33%	**98.83%**	**98.83%**
	0.8	**100%**	93.33%	**100%**	**100%**
Banknote ($\varphi = 0.5$, $q = 8$)	0.2	97.19%	98.19%	97.72%	**99.36%**
	0.4	97.71%	97.95%	**99.28%**	**99.28%**
	0.6	97.86%	98.75%	**99.64%**	**99.64%**
	0.8	97.24%	98.28%	**99.31%**	**99.31%**
Wine ($\varphi = 0.001$, $q = 18$)	0.2	97.18%	**97.89%**	95.77%	96.48%
	0.4	97.20%	97.20%	95.33%	**98.13%**
	0.6	**98.59%**	80.28%	98.13%	**98.59%**
	0.8	**100%**	83.33%	**100%**	97.22%

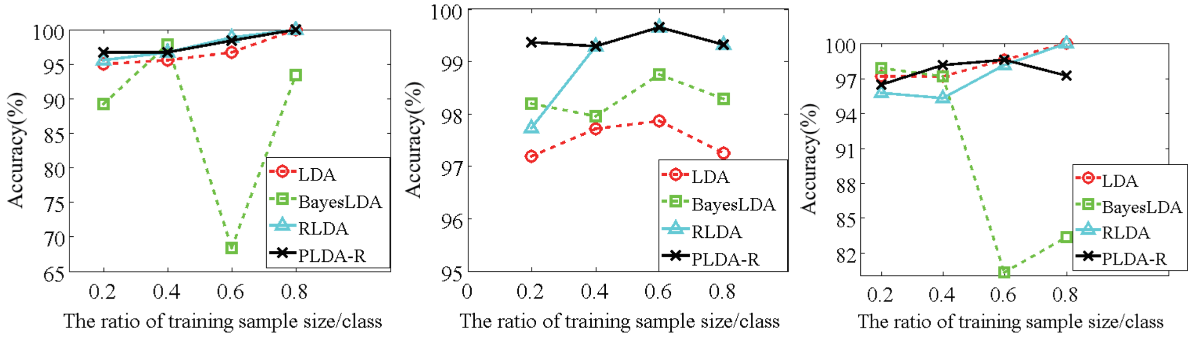


(a) Data set 1: Iris (b) Data set 2: Banknote (c) Data set 3: Wine

FIGURE 2. The experimental results with different sample dimensions

using randomization is much higher than traditional LDA at the minimum of dimension. Besides, PLDA-R algorithm has more favorable results than traditional LDA, which owns 10 times the highest accuracy, and RLDA and LDA with Naive Bayes are followed by. Therefore, we have a conclusion that PLDA-R algorithm has a better performance as feature information of sample is incomplete, and recognition results of each dimension are shown in Figure 2 where traditional LDA with several classifiers is represented by dotted line and PLDA-R with variant forms is described by solid line.

4.2. **PLDA-R with different sample proportions.** For the proposal of examining the advantages of probabilistic generative model on small sized training sample, the paper selects different sample proportions as training set and the remaining for test with each class sample, and the experimental results are shown in Table 3. We can find that the PLDA-R still performs higher accuracy than traditional LDA with other classifiers on three data sets. In particular, the classification accuracy with randomization procedure has averagely increased by 2% at low training sample proportion, and the accuracy further increases using probabilistic generative model as classifier on Banknote. For Naive Bayes classifier, it lacks universality since the recognition accuracy is superior to LDA on the Banknote and poor on the other data sets. As a result, PLDA-R owns favorable results in a majority of data sets with small size training sample and Figure 3 describes this. However, there is no denying that the computational cost of PLDA-R is slightly higher than LDA due to the introduction of randomized input and probabilistic generative model.



(a) Data set 1: Iris (b) Data set 2: Banknote (c) Data set 3: Wine

Figure 3. The experimental results with different proportion of training sample

5. Conclusion.

The paper has proposed a probabilistic method of linear discriminant analysis with randomized input. By integrating LDA with randomization procedure, we are able to create a randomized space where samples project onto it so that the feature information of samples could be enlarged. The experimental results indicate that the proposed algorithm is superior to traditional LDA on classification performance as feature information is incomplete. In addition, probabilistic generative model has been used to classify samples as a classifier, which gives by giving a probabilistic discrimination result, and favorable classification accuracy is obtained on small sized training samples.

The PLDA-R combines randomized input and probabilistic generative model with LDA, and the experiment also demonstrates it provides better recognition accuracy than LDA.

There are several interesting directions for our future research. (1) The improved PLDA-R algorithm could be applied to biometric recognition such as face recognition, handwriting recognition and gait recognition. (2) Since the computational cost of PLDA-R is slightly higher than LDA, the paper will further simplify procedure of proposed algorithm and reduce its computational complexity in condition of ensuring the recognition accuracy.

Acknowledgment. This paper was funded by the youth academic team of the 12th Five-Year Comprehensive Investment Planning and Discipline Construction Projects. This work was supported by the National Natural Science Foundation of China (Grant Nos. 61272509, 61402331, 61402332) and supported by the Foundation of Educational Commission of Tianjin City, China (Grant No. 20140803).

REFERENCES

[1] S. Wang, *The Linear Discriminant Analysis based on Clustering Regularization*, Tianjin University, 2013.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st Edition, Springer, 2007.

[3] C. J. Zhou, L. Wang, L. Lv, X. D. Zheng, X. P. Wei and Q. Zhang, Face recognition based on independent component analysis image reconstruction and linear discriminant analysis, *ICIC Express Letters*, vol.8, no.9, pp.2457-2462, 2014.

[4] J. P. Ye, Least squares linear discriminant analysis, *International Joint Conference on Neural Networks*, vol.3, no.10, pp.1087-1093, 2007.

[5] A. Sharma and K. K. Paliwal, A new perspective to null linear discriminant analysis method and its fast implementation using random matrix multiplication with scatter matrices, *Pattern Recognition*, vol.45, no.6, pp.2205-2212, 2012.

[6] G. B. Huang, What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John Von Neumann'S Puzzle, *Cognitive Computation*, vol.7, no.3, pp.263-278, 2015.

[7] F. L. Zhou and Y. W. Wang, Quick online spam classification method based on active and incremental learning, *Journal of Intelligent and Fuzzy Systems*, vol.30, no.1, pp.17-27, 2016.

[8] A. Sharma and K. K. Paliwal, A deterministic approach to regularized linear discriminant analysis, *Neurocomputing*, pp.207-214, 2015.

[9] H. Yu and J. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, *Pattern Recognition*, vol.34, pp.2067-2070, 2001.

[10] T. V. Bandos, L. Bruzzone and G. Camps-Valls, Classification of hyperspectral images with regularized linear discriminant analysis, *IEEE Trans. Geoscience & Remote Sensing*, vol.47, no.3, pp.862-873, 2009.