ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 12, No. 3, 2017, *pp*.210-215

Gene expression data classification using exponential locality sensitive discriminant analysis

Chunming Xu [1]
[1] School of Mathematics and Statistical, Yancheng Teachers University,
Yancheng 224002,PR China
E-mail: ycxcm@126.com.

(Received April 05, 2017, accepted June 20, 2017)

Abstract. Locality sensitive discriminant analysis is a typical and very effective graph-based dimensionality reduction method which has been successfully applied in pattern recognition problems. LSDA aims to find a projection which maximizes the margin between data points from different classes at each local area. As a result, it can discover the local geometrical structure of the data samples. However, just as linear discriminant analysis, it has the small sample size (SSS) problem. To overcome this limitation, we propose a novel exponential locality sensitive discriminant analysis algorithm in this paper. The proposed algorithm can make nearby objects with the same labels in the input space also nearby in the new representation; while nearby objects with different labels in the input space should be far apart. In addition, it can also deal with the SSS problem. The experiments on gene expression data sets verify the effectiveness of the proposed algorithm.

Keywords: gene expression data classification; dimensionality reduction; locality sensitive discriminant analysis; exponential locality sensitive discriminant analysis

1. Introduction

In recent years, with the rapid development of microarray gene-expression technology, it is now possible to simultaneously monitor the expression of all genes in the genome with a single experiment. One important application of gene expression data is the classification of cancer or other diseases, which draws a great number of researchers' attention [1-3]. Typically, the gene expression data sets are characterized by thousands of variables on only a few observations. It has been observed that although there are a lot of genes for each observation, the number of tissue samples ranges from tens to hundreds. In other words, there is much redundant information resided in the high-dimensional gene-expression data. To remove redundant information, dimensionality reduction technique is an effective way.

During the past decades, many dimensionality reduction algorithms have been developed. Classical dimensionality reduction methods can be categorized into two class: unsupervised methods and supervised methods. Representatives of unsupervised methods are principal component analysis (PCA) [4], independent component analysis (ICA) [5] and locality preserving projections (LPP) [6]. Unfortunately, unsupervised dimensionality reduction methods don't utilize any class information so they are not suitable for classification problems. For supervised methods, linear discriminant analysis (LDA) [7] is one of the most popular dimensionality reduction techniques. LDA seeks the optimal transformation that maximizing the between-class scatter while at the same time minimizing the within-class scatter. LDA has been widely used in many practical applications such as image retrieval and face recognition due to the fact that it can extract the most discriminatory features. Many extended LDA algorithms have been developed, for example, regularized discriminant analysis (RDA) [8], kernel linear discriminant analysis (KLDA) [9], two dimensional linear discriminant analysis (2DLDA) [10], locality sensitive discriminant analysis (LSDA) [11] and so on.

However, in many case, the number of samples is smaller than the dimensionality of the samples which will leads to the SSS problem in linear discriminant analysis based methods. In this paper, the SSS problem of LSDA algorithm is considered and a novel exponential LSDA (ELSDA) algorithm is proposed to overcome the shortcoming. The proposed ELSDA method not only inherits the advantages of the LSDA algorithm, but also avoids the SSS problem.

The rest of this paper is organized as follows. Section 2 gives a brief review of LSDA. Our ELSDA approach is proposed in section 3. Experimental results and some conclusions are provided in sections 4 and 5, respectively.

2. Locality sensitive discriminant analysis

LDA is a supervised learning algorithm, which can take advantage of the classification information of samples, and has received wide attention in the field of pattern recognition. However, LDA does not consider the distribution of samples, so it is not very good to deal with the data with nonlinear geometric distribution. To resolve this drawback, a local sensitive discriminant analysis (LSDA) method is proposed in literature [11]. LSDA can make full use of both the label information and the local manifold structure information of labeled samples to guide the dimensionality reduction process.

Given l data points $\boldsymbol{x}_1,\dots,\boldsymbol{x}_l$ that are distribduted on a underlying submanifold. Let $l(\boldsymbol{x}_i)$ be the class label of $\boldsymbol{x}_i$ and its k nearest neighbors be $N(\boldsymbol{x}_i)=\{\boldsymbol{x}_i^1,\boldsymbol{x}_i^2,\dots,\boldsymbol{x}_i^k\}$ .By the label information, the set $N(\boldsymbol{x}_i)$ can be further divided into two non overlapping subsets, $N_b(\boldsymbol{x}_i)$ and $N_w(\boldsymbol{x}_i)$. $N_w(\boldsymbol{x}_i)$ contains the neighbors having the same label with $\boldsymbol{x}_i$, while $N_b(\boldsymbol{x}_i)$ contains the neighbors sharing different labels. Specifically,

$$N_w(\boldsymbol{x}_i)=\{\boldsymbol{x}_i^j \,|\, l(\boldsymbol{x}_i^j)=l(\boldsymbol{x}_i),1\le j\le k\} \tag{1}$$

$$\mathrm{N_b}(\boldsymbol{x}_i)=\{\boldsymbol{x}_i^j \,|\, \mathrm{l}(\boldsymbol{x}_i^j)\ne \mathrm{l}(\boldsymbol{x}_i),1\le \mathrm{j}\le \mathrm{k}\} \tag{2}$$

Define the weight matrices W_b and W_w respectively as follows:

$$W_{b,ij}=\begin{cases}1, \boldsymbol{x}_i\in N_b(\boldsymbol{x}_j) \quad or \quad \boldsymbol{x}_j\in N_b(\boldsymbol{x}_i)\\ 0, otherwise\end{cases} \tag{3}$$

$$W_{w,ij}=\begin{cases}1, \boldsymbol{x}_i\in N_w(\boldsymbol{x}_j) \quad or \quad \boldsymbol{x}_j\in N_w(\boldsymbol{x}_i)\\ 0, otherwise\end{cases} \tag{4}$$

The goal of LSDA is to maximize $\sum_{ij}(\boldsymbol{y}_i-\boldsymbol{y}_j)^2 W_{b,ij}$ while at the same time minimize

$$\sum_{ij}(\boldsymbol{y}_i-\boldsymbol{y}_j)^2 W_{w,ij}\ .$$

It is easy to derive that

$$\sum_{ij}(\boldsymbol{y}_i-\boldsymbol{y}_j)^2 W_{b,ij}=\frac{1}{2}(\boldsymbol{p}^T\boldsymbol{x}_i-\boldsymbol{p}^T\boldsymbol{x}_j)^2W_{b,ij}=\frac{1}{2}\boldsymbol{p}^T\boldsymbol{X}(\boldsymbol{D_b}-\boldsymbol{W_b})\boldsymbol{X}^T\boldsymbol{p},$$

where $y_i=\boldsymbol{p}^T\boldsymbol{x}_i$; $\boldsymbol{D_b}$ is a diagonal matrix which is satisfied that $D_{b,ii}=\sum_j W_{b_{ij}}$. On the other hand,

$$\sum_{ij}(\boldsymbol{y}_i-\boldsymbol{y}_j)^2 W_{w,ij}=\frac{1}{2}(\boldsymbol{p}^T\boldsymbol{x}_i-\boldsymbol{p}^T\boldsymbol{x}_j)^2W_{w,ij}=\frac{1}{2}\boldsymbol{p}^T\boldsymbol{X}(\boldsymbol{D_w}-\boldsymbol{W_w})\boldsymbol{X}^T\boldsymbol{p},$$

where $\boldsymbol{D_w}$ is a diagonal matrix which is satisfied that $D_{w,ii}=\sum_j W_{w,ij}$.

Thus the objective function of LSDA can be written as:

$$J(p)=\arg\max_p \alpha\boldsymbol{p}^T\boldsymbol{X}(\boldsymbol{D_b}-\boldsymbol{W_b})\boldsymbol{X}^T\boldsymbol{p}-(1-\alpha)\boldsymbol{p}^T\boldsymbol{X}(\boldsymbol{D_w}-\boldsymbol{W_w})\boldsymbol{X}^T\boldsymbol{p} \tag{5}$$

where α is a positive parameter and $0\le\alpha\le 1$.

Denote that $\boldsymbol{L_b}=\boldsymbol{D_b}-\boldsymbol{W_b}$, $\boldsymbol{L_w}=\boldsymbol{D_w}-\boldsymbol{W_w}$, then $J(p)$ can be further written as

$$J(p)=\arg\max_p \boldsymbol{p}^T\boldsymbol{X}(\alpha\boldsymbol{L_b}-(1-\alpha)\boldsymbol{L_w})\boldsymbol{X}^T\boldsymbol{p} \tag{6}$$

Constrain that $\boldsymbol{p}^T\boldsymbol{X}\boldsymbol{D_w}\boldsymbol{X}^T\boldsymbol{p}=1$, then the objective function (4) can be recast as the following optimization problem:

$$J(p)=\underset{\substack{\mathbf{p} \\ \mathbf{p}^T\mathbf{X}\mathbf{D}_w\mathbf{X}^T\mathbf{p}=1}}{\arg\max}\ \mathbf{p}^T\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T\mathbf{p} \tag{7}$$

Construct the Lagrangian function as follows:

$$\mathbf{p}^T\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T\mathbf{p}-\lambda\mathbf{p}^T\mathbf{X}\mathbf{D}_w\mathbf{X}^T\mathbf{p}=0 \tag{8}$$

The projective directions are the leading eigenvectors of the following equation:

$$(\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T)\mathbf{p}=\lambda\mathbf{X}\mathbf{D}_w\mathbf{X}^T\mathbf{p} \tag{9}$$

Then for a given sample x, its features can be obtained using the projective directions.

(7)

3. Exponential locality sensitive discriminant analysis (ELSDA)

As a supervised manifold learning algorithm, LSDA can use both the class information and the geomery structure information of samples to aid the feature extraction process. However, LSDA doesn't consider the small sample problem. In this section, the ELSDA algorithm is proposed to solve this problem.

3.1. Matrix exponential

In this subsection, we will introduce the definition of matrix exponential and its properties [12]. For an square matrix A with the size $n\times n$, its exponential matrix A can be expressed as

$$\exp(A)=I+A+\frac{A^2}{2}+\cdots+\frac{A^K}{K}+\cdots \tag{10}$$

where I is the identity matrix.

The matrix $\exp(A)$ has many merits, which can be summarized as follows:

(1) $\exp(A)$ is a finite and full rank matrix.

(2) If there exists matrix B satisfied that $AB=BA$, then $\exp(A+B)=\exp(A)\exp(B)$.

(3) The exponential matrix $\exp(A)$ is invertible. In fact, $\exp(A-A)=\exp(A)\exp(-A)=I$ .Therefore, the inverse matrix of $\exp(A)$ is $\exp(-A)$.

(4) For any nonsingular matrix B, we have $\exp(B^{-1}AB)=B^{-1}\exp(A)B$.

(5) If $v_1,v_2,\cdots,v_d$ are eigenvectors of A corresponding to the eigenvalues $\lambda_1,\lambda_2,\cdots,\lambda_d$, then $v_1,v_2,\cdots,v_d$ are eigenvectors of $\exp(A)$ corresponding to the eigenvalues $e^{\lambda_1},e^{\lambda_2},\cdots,e^{\lambda_d}$.

Among the characteristics of the matrix $\exp(A)$, the most important is that it is a invertible matrix. In addition, it has the same eigenvectors as matrix A. As a result, we can substitute $\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T$ and $\mathbf{X}\mathbf{D}_w\mathbf{X}^T$ with $\exp(\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T)$ and $\exp(\mathbf{X}\mathbf{D}_w\mathbf{X}^T)$ in Eq.(7) respectively to get the objective function of ELSDA.

3.2. ELSDA

Based on Eq.(7) and Matrix exponential theory, the criterion of ELSDA can be specially established as follows:

$$J(p)=\underset{\substack{\mathbf{p} \\ \mathbf{p}^T\mathbf{exp}(\mathbf{X}\mathbf{D}_w\mathbf{X}^T\mathbf{p})=1}}{\arg\max}\ \mathbf{p}^T\mathbf{exp}(\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T)\mathbf{p} \tag{11}$$

The solution of Eq. (11) can be easily achieved by the following generalized eigenvalue problem:

$$\exp(\mathbf{X}(\alpha\mathbf{L}_b+(1-\alpha)\mathbf{W}_w)\mathbf{X}^T)\mathbf{p}=\lambda\exp(\mathbf{X}\mathbf{D}_w\mathbf{X}^T)\mathbf{p} \tag{12}$$

Specifically, the coefficient matrix $\mathbf{P}$ can be constructed by the eigenvectors of Eq.(9) associated with the first d largest eigenvalues $\mathbf{p}_1,\mathbf{p}_2,\cdots,\mathbf{p}_d$, i.e. the coefficient matrix $\mathbf{P}$ can be constructed a

$$\mathbf{P}=(\mathbf{p}_1,\mathbf{p}_2,\cdots,\mathbf{p}_d).$$

Therefore the new data representation of $\mathbf{x}$ can be expressed as:

$$\mathbf{y}=\mathbf{P}^T\mathbf{x} \tag{13}$$

One thing that should be pointed out is that we should normalize the matrix before calculating its matrix exponential due to that its matrix exponential may involve large numbers.

3.3. The steps of the ELSDA algorithm

In summary of the preceding description, the main steps of the proposed method are described below:

Step 1. Normalize the data samples $\boldsymbol{x_1},\ldots,\boldsymbol{x_l}$.

Step 2. Construct the weight matrices W_b and W_w according to Eq. (3) and Eq. (4) respectively.

Step 3. Calculate the matrices L_b and L_w.

Step 4 . Calculate the matrices $\exp(\boldsymbol{X}(\alpha\boldsymbol{L_b}+(1-\alpha)\boldsymbol{W_w})\boldsymbol{X}^T)$ and $\exp(\boldsymbol{XD_wX}^T)$.

Step 5. Compute the eigenvectors $\boldsymbol{p_1},\boldsymbol{p_2},\cdots,\boldsymbol{p_d}$ of $(\exp(XD_w\boldsymbol{X}^T))^{-1}\exp(\boldsymbol{X}(\alpha\boldsymbol{L_b}+(1-\alpha)\boldsymbol{W_w})\boldsymbol{X}^T)$ associated with the first d largest eigenvalues, then the optimal projection matrix is formed by $\boldsymbol{P}=(\boldsymbol{p_1},\boldsymbol{p_2},\cdots,\boldsymbol{p_d})$.

Step 6. Feature extraction: $\boldsymbol{y}=\boldsymbol{P}^T\boldsymbol{x}$.

4. Experimental results

BBDED problem is solved using SOS algorithm for a system of 10 generators, 6 customers in 12 dispatch periods. The parameters set as: population size-50, maximum number of iteration-100.

4.1. Convergence analysis

In this section, we will evaluate the performances of the proposed ELSDA algorithm for gene expression data classification based on two public microarray data sets. The details of the data sets used in our experiments are summarized as follows:

The Colon dataset contains 62 samples collected from colon-cancer patients. Among them, 40 tumor biopsies are from tumors and 22 normal. There are 7129 genes in the dataset[13].

The Leukemia dataset contains two types of acute leukemia: 47 acute lymphoblastic leukemia and 25 acute myeloid leukemia. There are 2000 genes in the dataset [14].

For comparison, we also present the results of two competing dimensionality reduction algorithms, i.e., linear discriminant analysis (LDA) and locality sensitive discriminant analysis (LSDA). In this paper, the values of the regularization parameter α in LSDA an ELSDA is fixed to 0.1, the number of nearest neighbors is empirically set to 8. For all the dimensionality reduction algorithms, we reduce the dimension to 20.

Many classifiers have been successfully used to classify gene expression data, such as K-neighbor [15], Bayesian [16], and Support Vector Machines [17] and so on. In this paper, the K-neighbor classifier with K=1 is adopted because of its simplicity.

In order to obtain reliable experimental results, the tests are carried within the framework of 3-fold cross validation, a statistical method of evaluating and comparing learning algorithms. Firstly, the data is partitioned into three data sets of approximately equal size respectively. Then the training data set, which contains two parts of the subsets, is used to learn a classification model while the remaining subset is used to validate the effectiveness of the proposed model. The procedure is repeated three times and the performance is evaluated by the averaged recognition results over the three subsets.

In general, the recognition rate varies with the dimension of the feature subspace. Figure 1-2 shows the plots of recognition rates versus dimensionality reduction for the LDA, LSDA and ELSDA algorithms on the Colon dataset and the Leukemia dataset respectively. As can be seen, our proposed ELSDA algorithm outperforms LDA and LSDA in recognition accuracy with different number of dimensions. In addition, ELSDA is more stable than the other two methods.

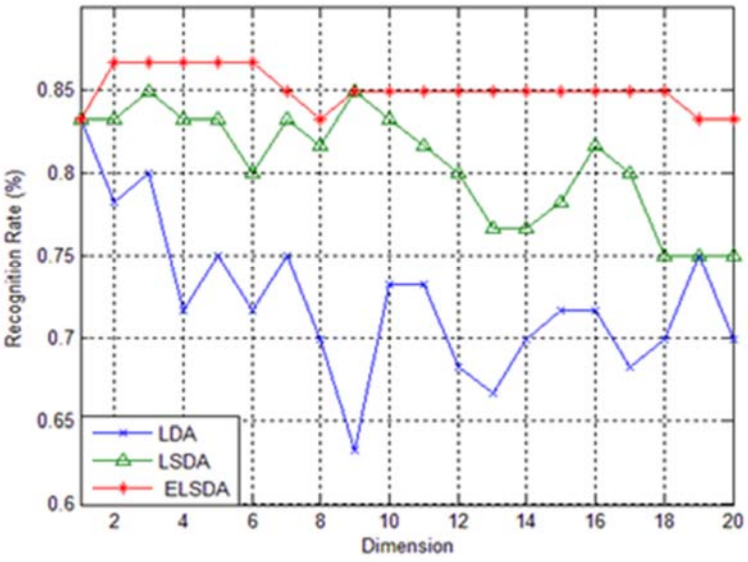

Fig. 1 Recognition results of different algorithms on the Colon dataset

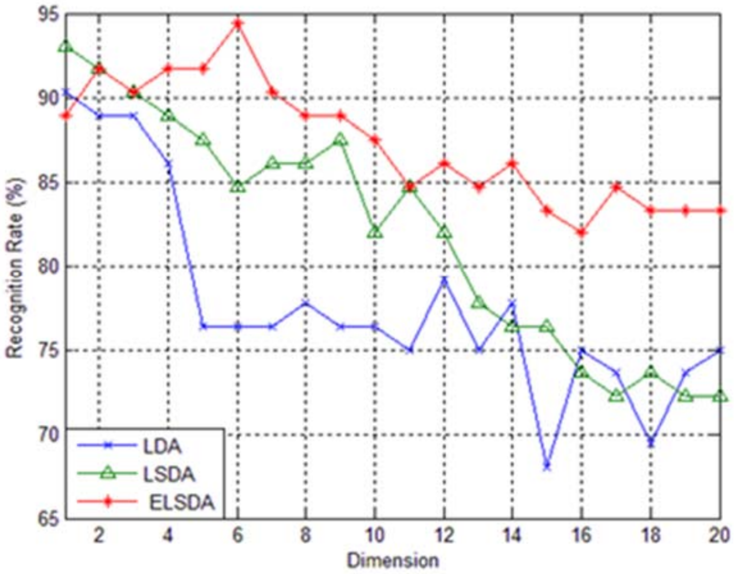

Fig. 2 Recognition results of different algorithms on the Leukemia dataset

Table.1 The top recognition rates of LDA, LSDA and ELSDA

Method	LDA	LSDA	ELSDA
The Colon dataset	83.33%	85.00%	86.67%
The Leukemia dataset	90.28%	93.06%	94.44%

The experiment has been conducted to demonstrate the properties of the proposed algorithm. From the experiment results we can have three observations:

(1).LDA is a traditional linear dimensionality reduction method and gets the poorest results among the three methods.

(2).Comparatively, LSDA and ELSDA are based on the graph theory and contain valuable geometry structure information which is helpful for nonlinear classification purpose, so they have stronger recognition power and their recognition results are better than LDA.

(3).ELSDA outperforms the other two techniques, etc. LDA and LSDA. This is mainly because it not only exploits the geometry of the data manifold but also does not suffer from the SSS problem, so that it can reserve more useful information and achieve the best results among the three algorithms.

5. Conclusion

This paper presents a new dimensionality reduction method called ELSDA. The proposed ELSDA algorithm exploits not only the label information but also the local manifold structure of given labeled samples. In addition, it can preferably solve the SSS problem. As a result, ELSDA is expected to perform better when there are no sufficient training samples. The ELSDA method is tested on gene expression data sets and the experimental results demonstrate the superior performance of our proposed algorithm.

6. Acknowledgement

This work is supported by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No.13KJD520010).

7. References

[1] Jia Lv, Qinke Peng , Xiao Chen, A multi-objective heuristic algorithm for gene expression microarray data classification, Expert Systems With Applications, 59 (2016), pp.32-39.

[2] M Pan, J Zhang. Correlation-based linear discriminant classification for gene expression data, Genetics & Molecular Research,16 (2017), pp.63–76.

[3] P Mohapatra, S Chakravarty, PK Dash. Microarray medical data classification using kernel ridge regression and modified cat swarm optimization based gene selection system, Swarm & Evolutionary Computation, 28 (2016), pp. 144–160.

[4] M.A. Turk, A.P. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neuroscience, 3 (1991), pp. 71-96.

[5] Bartlett, M.S., Movellan, J.R. and Sejnowski, T.J., Face Recognition by Independent Component Analysis, IEEE Trans. Neural Networks, 13 (2002), pp. 1450–1464.

[6] X He，P Niyogi. Locality Preserving Projections (LPP). Advances in Neural Information Processing Systems, 16(2002): pp.186-197.

[7] Etemad K., Chellapa R., Discriminant Analysis for Recognition of Human Face Images, Journal of the Optical Society of America A , 14(1997): pp. 1724–1733.

[8] Friedman J. Regularized discriminant analysis, Journal of the American Statistical Association, 84 (1989): pp.165-175.

[9] Yang M H. Kernel Eigenfaces vs. kernel Fisherfaces: face recognition using kernel method. In: Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 215-220, 2002.

[10] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix. Pattern Recognition Letters, 26(2005): pp.527-532.

[11] D. Cai, X.F. He, K. Zhou, Locality sensitive discriminant analysis, Proc. Int. Joint Conf. on Artificial Intelligence, Hyderabad, India, pp. 708-713, 2007.

[12] T. Zhang, B. Fang, Y.Y. Tang, Z. Shang, B. Xu, Generalized discriminant analysis: a matrix exponential approach, IEEE Transactions on Systems, Man, and Cybernetics, Part B. 40 (2010) pp.186–197.

[13] T.R. Golub, D.K. Slonim, P. Tamayo, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science, 286(1999): pp.531-537.

[14] U. Alon, N. Bkraai, D.A. Notterman, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proceedings of the national academy of sciences, 96(1999): pp.6745-6750.

[15] S. Deegalla and H. Bostrom, Classification of microarrays with kNN: comparison of dimensionality reduction methods, Lecture notes in computer science, 4881(2007): pp.800-809.

[16] P. Helman, R. Veroff and S.R. Atlas, A bayesian network classification methodology for gene expression data, Journal of computational biology, 11(2004): pp.581-615.

[17] T.S. Furey, N. Cristianini, N. Duffy, Support vector machines classification and validation of cancer tissue samples using microarray expression data, Bioinformatics, 16(2000): pp.906-914.