Neurocomputing 173 (2016) 163–171

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# A novel dimensionality reduction method with discriminative generalized eigen-decomposition

Xiumei Wang, Weifang Liu, Jie Li, Xinbo Gao *

*VIPS Lab, School of Electronic Engineering, Xidian University, Xi'an 710071, China*

ARTICLE INFO

*Article history:*
Received 1 July 2014
Received in revised form
1 December 2014
Accepted 14 December 2014
Available online 2 September 2015

*Keywords:*
Dimensionality reduction
Signal to noise ratio
Generalized eigen-decomposition
Supervised learning

ABSTRACT

Dimensionality reduction plays a critical role in machine learning and computer vision for past decades. In this paper, we propose a discriminative dimensionality reduction method based on generalized eigen-decomposition. Firstly, we define the discriminative framework between pairwise classes inspired by the signal to noise ratio. Then the metric is given for intra-class compactness and inter-class separation. Finally, the framework for one against one class can be easily extended to one against all classes. Compared with traditional supervised dimensionality reduction methods, the proposed method can catch discriminative directions for pairwise classes rather than for all classes. Furthermore, it also can deal with non-Gaussian distributed data. The experimental results show that the proposed model can achieve high precisions in classification tasks.



## 1. Introduction

In machine learning and pattern recognition, dimensionality reduction is a standard tool in dealing with high-dimensional data sets, which can efficiently avoid the problem of "curse of dimensionality". Dimensionality reduction is generally used as the pre-processing method for high-dimensional data tasks, such as biometric recognition, image retrieval and disaster prediction. The purpose of dimensionality reduction techniques is to discover a new low-dimensional space, where the analysis of intrinsic structure for various applications would be more efficient.

Most of the conventional dimensionality reduction methods are established on Gaussian distribution assumption, such as principal component analysis (PCA) [1,2] and linear discriminant analysis (LDA) [1–3], which may lose the local structure when encountering non-Gaussian distributed data. Non-linear manifold learning methods [4–7], such as Laplacian eigenmaps (LE) [8,9], locally linear embedding (LLE) [10] and ISOMAP [11], can deal with non-Gaussian distributed data by preserving the local geometry of the nearest neighbors. However, the local neighborhood relationships are defined only on the training samples, and it is unclear how to evaluate the maps of new testing samples. Locality preserving projections (LPP) [12] is the linear extension of LE and can locate a new sample in the low-dimensional space easily with a linear transformation matrix. The same as other manifold learning methods, it ignores the discriminative information of samples with different labels. Discriminative locality alignment (DLA) [13] tries to minimize the distances between nearest neighbors from the same class and simultaneously maximize the distances between nearest neighbors from different classes. DLA can capture the intra-class non-Gaussian structure and inter-class discriminative information. The directions obtained by DLA complex discriminative property of all classes, but for two given classes, there may not exist the most discriminative one among all directions. Therefore, it is difficult to intuitively explain the directions in DLA. On the other hand, this kind of non-Gaussian distribution dimensionality reduction techniques tends to find the linear sub-space of the high-dimensional data, and it may lose effectiveness when facing with the nonlinear structure data.

In order to overcome the aforementioned drawbacks, we propose a generalized dimensionality reduction method named discriminative generalized eigen-decomposition (DGE). DGE is inspired by generalized eigenvectors for multiclass (GEM) [14] which deal with the data in a framework of signal to noise ratio by solving the generalized eigen-decomposition problem. GEM is still a feature extraction method based on Gaussian distribution assumption. To discover the local structure hidden in non-Gaussian distributed data, DGE defines local signal and noise around each training sample, and combines them in a "one-against-one" way to construct signal to noise ratio functions. DGE can extract discriminative features between pairwise classes in an intuitive and implementable way. The contributions of DGE can be summarized as follows,

* Corresponding author. Tel.: +86 2988201838.
*E-mail address:* xbgao.xidian@gmail.com (X. Gao).

http://dx.doi.org/10.1016/j.neucom.2014.12.122
0925-2312/

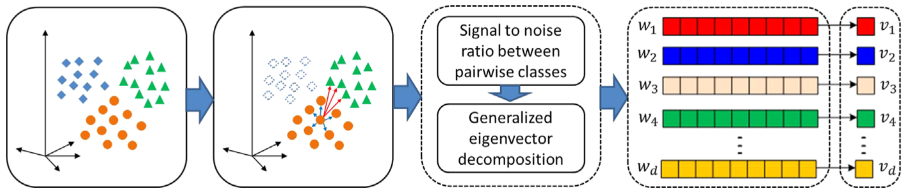


**Fig. 1.** Flowchart of DGE.

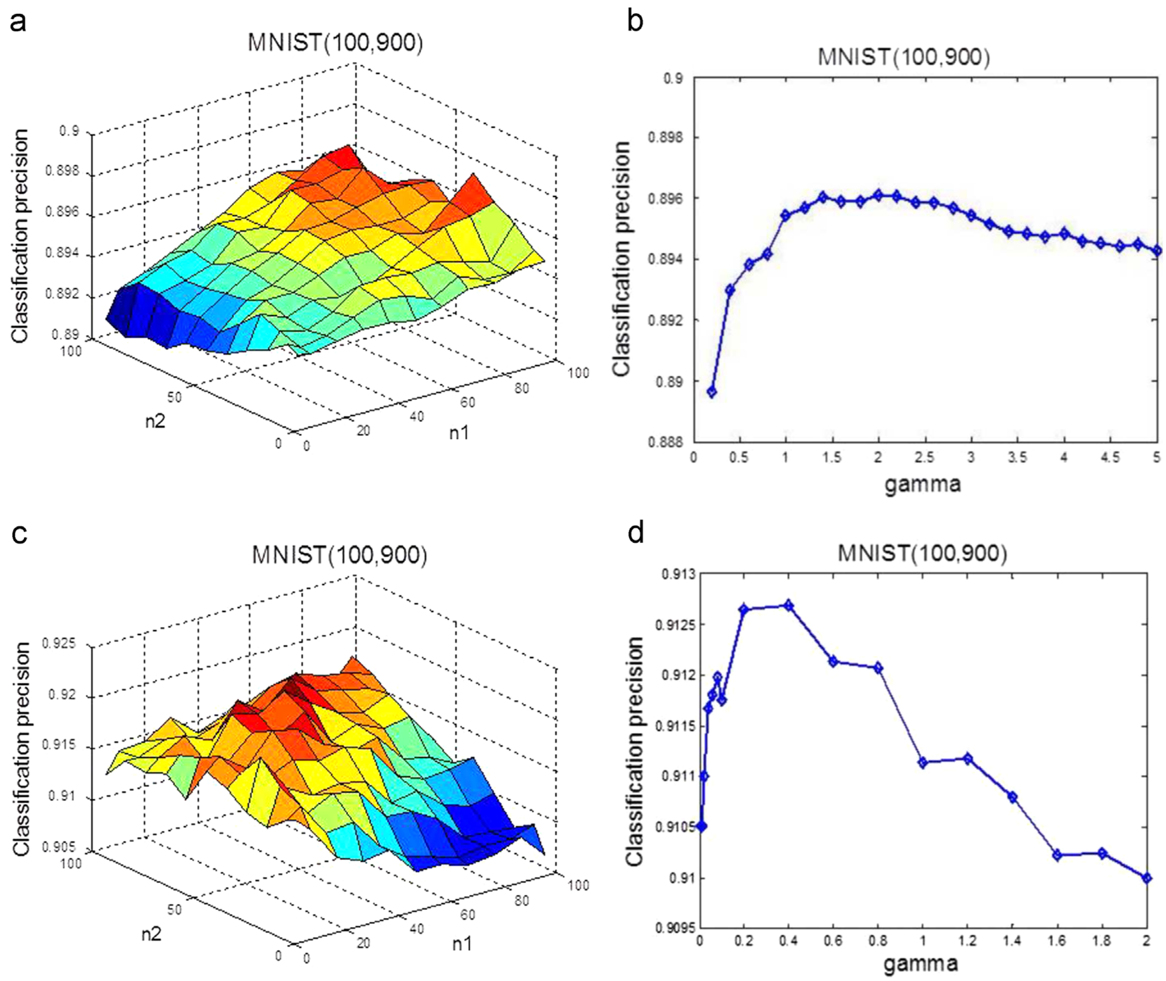


**Fig. 2.** Parameter settings for classification with HDGE on MISNT dataset. (a) and (b) are the classification precisions of the first layer of HDGE under different parameter settings. (c) and (d) are the classification precisions of the second layer of HDGE under different parameter settings.

1. DGE can generally deal with complex distribution data, including Gaussian distribution and non-Gaussian distribution.
2. DGE can extract the more discriminative directions for pairwise classes rather than complex the discriminative information of all classes.
3. DGE can be easily extended to solve nonlinear dimensionality reduction problems by establishing hierarchical structure with intermediate nonlinear transformation.
4. DGE extracts the directions for each pair of classes, so it can deal with the case that the number of data dimensions is less than the number of classes.

The rest of this paper is organized as follows. We review the related work in Section 2. In Section 3, the proposed method is presented in detail. In Section 4, experimental results on datasets with different feature types are analyzed. In the last section, we draw the conclusion.

## 2. Related work

In this section, we will first review the fundamental theory of GEM. Formally, given a set of labeled data $\{x_i, y_i\}_{i=1}^{n}$ sampled iid a distribution on $\mathbb{R}^m \times [k]$. GEM treats the pairwise classes as signal and noise respectively. For pairwise classes $(i, j) \in T = \{(i, j) | i, j \in [k], i \neq j\}$, signal and noise on direction $w$ are defined by the conditional second moments given class labels $E[(w^T x)^2 | y = i]$ and $\mathbb{E}[(w^T x)^2 | y = j]$. Then the motivation of GME is to find the discriminative direction $w$ through maximizing the signal to noise ratio

$$R_{ij}(w) = \frac{E[(w^T x)^2 | y = i]}{E[(w^T x)^2 | y = j]} = \frac{w^T E[x^T x | y = i] w}{w^T E[x^T x | y = j] w} = \frac{w^T C_i w}{w^T C_j w}. \tag{1}$$

The local maximizers of Eq. (1) are equal to solve the generalized eigenvectors $C_i w = \lambda C_j w$. Since the objective function is homogeneous in $w$, it is assumed that $w^T C_j w = 1$. Intuitively, GEM extracts the