Pattern Recognition 43 (2010) 1269–1281

Contents lists available at ScienceDirect

Pattern Recognition

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

A regularization framework for robust dimensionality reduction with applications to image reconstruction and feature extraction

Zhizheng Liang [a,*], Youfu Li [b]

[a] School of Computer Science and Technology, China University of Mining and Technology, China
[b] Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong

A R T I C L E I N F O

Article history:
Received 26 March 2009
Received in revised form
12 October 2009
Accepted 16 October 2009

Keywords:
Regularization framework
Nonlinear eigenvalue problem
SCF iteration
Robust
Feature extraction
Image reconstruction

A B S T R A C T

Dimensionality reduction has many applications in pattern recognition, machine learning and computer vision. In this paper, we develop a general regularization framework for dimensionality reduction by allowing the use of different functions in the cost function. This is especially important as we can achieve robustness in the presence of outliers. It is shown that optimizing the regularized cost function is equivalent to solving a nonlinear eigenvalue problem under certain conditions, which can be handled by the self-consistent field (SCF) iteration. Moreover, this regularization framework is applicable in unsupervised or supervised learning by defining the regularization term which provides some types of prior knowledge of projected samples or projected vectors. It is also noted that some linear projection methods can be obtained from this framework by choosing different functions and imposing different constraints. Finally, we show some applications of our framework by various data sets including handwritten characters, face images, UCI data, and gene expression data.

1. Introduction

In data analysis problems where there are a large number of input variables, it is often beneficial to reduce the dimension of data in order to improve the efficiency and accuracy of data analysis. Consequently, dimensionality reduction becomes one of key techniques in data analysis. Dimensionality reduction aims at reducing the dimensionality of data such that the extracted features are as representable as possible. During the past several decades, a variety of algorithms and techniques [1–8] for dimensionality reduction have been developed. Among them, principal component analysis (PCA) and linear discriminant analysis (LDA) are regarded as the most powerful tools of dimensionality reduction. In general, PCA is to find an orthogonal set of vectors by maximizing the variance of the projected data, whereas LDA is to seek discriminant vectors by maximizing the ratio of the between-class distance to the within-class distance. It is shown that LDA is a more effective method for extracting features in the classification problem as compared to PCA in general cases. However, LDA often suffers from the small sample size (3S) problem when the dimension of data is much larger than the number of data points.

In recent years, many approaches [10–16] have been proposed to deal with high dimensional data and the 3S problem. For example, the Fisherface method [2] first applies PCA to reduce the dimension of samples to obtain a full-rank within-class scatter matrix. Then standard LDA is used to extract features. In [15], Chen et al. proposed the null space-based LDA, where the between-class scatter is maximized in the null space of the within-class scatter matrix. In [12], Howland and Park proposed the LDA/GSVD algorithm which circumvents the singularity problem by using the generalized singular value decomposition. Direct LDA [17] first removes the null space of the between-class scatter matrix and then seeks the projection to minimize the within-class scatter. In order to reduce the computational cost of LDA, Ye and Li [5] proposed a two-stage LDA extension (LDA/QR). Their method first applies the QR decomposition on a small matrix, and then followed by LDA. Further, Zhang and Sim [10] analyzed LDA via the Fukunaga–Koontz transform, which provides a unified framework for understanding some variants of LDA. In [14], Li et al. proposed an efficient and stable method to calculate discriminant vectors based on the maximum margin criterion (MMC). The difference between Fisher's criterion and MMC is that the former maximizes the Fisher quotient while the latter maximizes the average distance. In [18], the authors proposed a unified framework for generalized LDA via a transfer function. It is shown that uncorrelated LDA is a special case of PCA plus LDA and regularized LDA.

Although PCA and LDA have been successfully used in solving some problems in pattern recognition and machine learning, they are prone to the presence of outliers due to the fact they do not involve robust functions in the cost function. In order to deal with this problem, some researchers proposed robust algorithms [19–24] for dimensionality reduction in recent years. In [24], the

* Corresponding author.
E-mail address: cuhk_liang@yahoo.cn (Z. Liang).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.10.012

authors formulated matrix factorization as an L1 norm minimization problem, which can be efficiently solved by alternate convex programming. However, the solution does not have rotational invariance. Considering this point, the authors [21] proposed rotational invariant L1 norm principal component analysis which combines some merits of PCA and L1 PCA [24]. Their method can suppress the effect of outliers by defining a modified covariance matrix which softens contributions from outliers. In [19], the authors proposed a method of principal component analysis based on a new L1 norm optimization technique. The L1 norm optimization algorithm is robust to outliers and is easy to implement.

Note that the algorithms for robust PCA are to minimize the error between original data and reconstructed data in terms of different objective functions. However, they may produce undesirable classification performances due to the fact they are devised from the viewpoint of data reconstruction. Furthermore, they do not make full use of prior knowledge of data points such as the geometrical structure of data points. To this end, we develop a regularization framework of discriminant analysis by using prior knowledge of data points. In this framework, one can flexibly choose robust functions to suppress the presence of outliers. Moreover, a regularization parameter is used to control the tradeoff between the data reconstruction error and prior knowledge of data points. It is found that the optimization problem can be formulated as a nonlinear eigenvalue problem under proper conditions. Further, we propose a projected nonlinear eigenvalue problem. In addition, we also conduct extensive experiments to evaluate the proposed framework on various data sets including handwritten numerals, UCI data sets, face images and gene expression data. Overall, the main contributions of this paper include

(1) We develop a regularization framework of discriminant analysis for dimensionality reduction. In this framework, one can choose robust functions to suppress the presence of outliers. Moreover, we are also capable of using this framework to implement the data reconstruction problem.
(2) We give the detailed analysis on the relationship among some linear projected methods. In particular, we show that regularized MMC is a special case of our framework, which helps explain why regularized MMC is a robust feature extraction method, and also point out the range of the regularization parameter in regularized MMC.
(3) We conduct extensive experiments on various data sets to evaluate the effectiveness of our framework and compare it with some linear projected methods.

The rest of this paper is organized as follows. Section 2 overviews linear projection methods including PCA, LDA, MMC, and regularized MMC. In Section 3, we give a regularization framework of discriminant analysis for dimensionality reduction and show how to solve the optimization problem. In Section 4, links to some existing linear projected methods are given. Section 5 gives the detailed experimental results. Section 6 contains some concluding remarks and further directions.

2. PCA, LDA, MMC and regularized MMC

Assume that $x_1,\ldots,x_m$ are a set of n-dimensional samples of size m, $x_i \in \Re^n (i=1,\ldots,m)$. Each sample belongs to exactly one of c object classes $\{l_1,\ldots,l_c\}$ and the number of samples in the ith class is m_i. The between-class scatter matrix, the within-class scatter matrix, and the total scatter matrix are defined as:

$$S_b = \sum_{i=1}^{c} m_i(\mu_i-\mu)(\mu_i-\mu)^T,$$

$$S_w = \sum_{i=1}^{c}\sum_{x\in l_i}(x-\mu_i)(x-\mu_i)^T, \quad S_t = \sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T,$$

where $_i$ is the centroid of the ith class and μ is the global centroid of the sample set.

2.1. PCA

Principal component analysis, also called Karhumen–Loeve transform in some sense, extracts the desired number of principal components for data by minimizing the mean squared error criterion. The optimal linear transformation $U \in \Re^{n\times k}$ for PCA is the one that maximizes the total scatter in a reduced dimensional space. The matrix U can be obtained by performing the eigendecomposition on S_t and the columns of U are eigenvectors of S_t corresponding to the first k largest eigenvalues. It is easy to verify that the ith eigenvalue is the variance of data that is projected onto the ith eigenvector. A good property of PCA is that it decorates the data.

2.2. Classical LDA

Classical LDA seeks the direction on which data points of different classes are far from each other while requiring data points of the same class to be close to each other. To be specific, LDA is to find the optimal projection by optimizing the objective function in the following:

$$\max \mathit{trace}((U^TS_wU)^{-1}(U^TS_bU)). \quad (1)$$

The optimal transformation U can be obtained by solving the generalized eigenproblem: $S_bu=\lambda S_wu$. In general, there are at most $c-1$ eigenvectors corresponding to nonzero eigenvalues since the rank of the matrix S_b is not bigger than $c-1$. When S_w is singular, one can overcome it by applying some methods such as LDA/QR [6], PCA plus LDA [2], LDA/GSVD [9], and LDA/FKT [10].

2.3. MMC and regularized MMC

MMC aims at maximizing the average margin between different classes. To be specific, MMC is to optimize the objective function: $\mathit{trace}(U^T(S_b-S_w)U)$ under the proper constraint. The optimal transformation U can be obtained by performing the eigen-decomposition on the matrix (S_b-S_w). The matrix U is composed of the first k eigenvectors of S_b-S_w corresponding to the first k largest eigenvalues. The regularized MMC is to maximize $\mathit{trace}(U^T(S_b-\gamma S_w)U)$ with a nonnegative regularization parameter γ. As pointed out in [25], the MMC or regularized MMC can also be performed within the range space of S_t since the null space of S_t does not contain any discriminant information. As a result, the computational complexity of MMC or regularized MMC can be further reduced.

3. The regularization framework of discriminant analysis

3.1. The regularization framework

In this section, we assume that the data is centralized without loss of generality. In fact, this is easily obtained by a translation of data. It is shown [26] that the standard PCA is equivalent to