Information Sciences 303 (2015) 1–14

Contents lists available at ScienceDirect

Information Sciences

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

Regression analysis of locality preserving projections via sparse penalty

Zhonglong Zheng [a,b,*], Xiaoqiao Huang [a], Zhongyu Chen [a], Xiaowei He [a], Huawen Liu [a], Jie Yang [b]

[a] Department of Computer Science, Zhejiang Normal University, China
[b] Department of Computer Science, University of California, Merced, USA

ARTICLE INFO

Article history:
Received 5 September 2013
Received in revised form 31 October 2014
Accepted 4 January 2015
Available online 20 January 2015

Keywords:
Subspace learning
Locality preserving projections
Sparse penalty
Regression analysis

ABSTRACT

Recent studies have shown that linear subspace algorithms, such as Principal Component Analysis, Linear Discriminant Analysis and Locality Preserving Projections, have attracted tremendous attention in many fields of information processing. However, the projection results obtained by these algorithms are linear combination of the original features, which is difficult to be interpreted psychologically and physiologically. Motivated by Compressive Sensing theory, we formulate the generalized eigenvalue problem under CS framework, which then allows us to apply a sparsity penalty and minimization procedure to locality preserving projections. The proposed algorithm is called sparse locality preserving projections, which performs locality preserving projections in the lasso regression framework that dimensionality reduction, feature selection and classification are merged into one analysis. The method is also extended to its regularized form to improve its generalization. The proposed algorithm is a combination of locality preserving with sparse penalty. Additionally, the algorithm can be performed in either supervised or unsupervised tasks. Experimental results on toy and real data sets show that our methods are effective and demonstrate much higher performance.

1. Introduction

During the past decades, problems in which the number of features is much larger than the number of observations have attracted much interests. Linear dimensionality reduction methods are popular due to the advantages of reasonable motivation in principle and the simplicity in form.

Principal component analysis (PCA) is an unsupervised linear method of variables technique used in data compression, classification, and visualization [24]. The essence of PCA is to extract principal components, linear combinations of input variables that together best account for the variance in a data set. Linear discriminant analysis (LDA) is a favored tool for supervised linear classification in many areas because of its simplicity and robustness [1]. The goal of LDA is to provide low dimensional projections of data onto the most discriminative directions. Locality preserving projection (LPP) is a recently proposed method [12], which can be regarded as the linearization of Laplacian EigenMap [2]. When applied to face recognition tasks, LPP is also called LaplacianFaces. The idea behind LPP is that it considers the manifold structure of the data

* Corresponding author at: 5210 Lake Rd, 208 Room, SE Building, University of California, Merced 95340, USA.
E-mail addresses: zhonglong@zjnu.cn, zzheng@ucmerced.edu (Z. Zheng).

http://dx.doi.org/10.1016/j.ins.2015.01.004
0020-0255/© 2015 Published by Elsevier Inc.

set, and preserves the locality of data in the embedding space. LPP has shown the superiority in terms of image indexing and face recognition.

One of the major disadvantages of these three methods is that the derived projections are linear combinations of all the original features. Hence the learned results are difficult to interpret. Since the features are likely to be correlated, it is possible that a subset of these features can be chosen such that the others may not contain substantial additional information and may be deemed redundant in the presence of this subset of features. In other words, the data in the reduced subspace is represented as a linear combination of a subset of the features which are the most informative. Recent achievements in psychology and physiology have shown that the representation of objects in human brain may be sparsity-based [22]. Sparse representations have attracted a great deal of attention in signal processing and information theory [4,21,26,27,29,25]. Recent progress has been made on the surprising effectiveness of the ℓ_1 norm for recovering sparse representations. Zou et al. proposed sparse principal component analysis by using ℓ_1 penalized regression on conventional PCA [34]. Wu et al. proposed sparse linear discriminant analysis in a similar framework [18,28]. Another sparse variation of discriminant analysis, called SDA, was presented based on formulating classification as regression [5]. Han et al. extended single-task SDA to the multi-task scenario with a method called multi-task sparse discriminant analysis MtSDA [11] Qiao et al. proposed sparse preserving projection method, which aims to preserve the sparse reconstructive relationship of the data by minimizing a ℓ_1 regularization-related objective function [19]. by maximizing the ratio of the ℓ_1-norm-based locality preserving between-class dispersion and the ℓ_1-norm-based locality preserving within-class dispersion, DLPP-ℓ_1 was presented and proved to be more robust to outliers [31].

In this paper, we extend traditional sparse framework to handling the generalized eigenvalue problems, which then allows us to apply a sparsity penalty and minimization procedure to locality preserving projections. The proposed algorithm is called sparse locality preserving projections (SpLPP), which is based on lasso regression framework for learning sparse projections by incorporating ℓ_1 penalty with conventional locality preserving projections. The affinity graph constructed in LPP encodes both discriminant and geometrical structure in the data [12]. Once the Laplacian matrix is computed, we recast the generalized eigenvalue problem of LPP in the lasso regression framework to obtain sparse basis functions. The proposed SpLPP is a combination of locality preserving with sparsity. Additionally, our algorithm can be performed in either supervised or unsupervised mode.

The rest of this paper is organized as follows: A short review of LPP is described in Section 2. In Section 3, we provide SpLPP algorithm in detail, together with the regularized SpLPP. A variety of experimental results are presented in Section 4. Finally, we provide some concluding remarks and suggestions for future work in Section 5.

2. Review of locality preserving projection

Given the original data set $\{x_1, \ldots, x_n\} \in \mathbb{R}^m$, let $X = [x_1, \ldots, x_n]$, then X is an $m \times n$ size matrix. Let S be a similarity matrix defined on all pairwise data points. LPP can be achieved by optimizing the following minimization problem:

$$\begin{aligned} w_{opt} = &\ \arg\min_w \sum_{ij} (y_i - y_j)^2 S_{ij} \\ = &\ \arg\min_w \sum_{ij} (w^t x_i - w^t x_j)^2 S_{ij} \\ = &\ \arg\min_w \ w^T X L X^T w \\ &\ \text{s.t.} \ \ w^T X D X^T w = 1 \end{aligned} \tag{1}$$

where $L = D - S$ is the *graph Laplacian matrix*, and $D_{ii} = \sum_j S_{ij}$ is the local density measure around x_i. The symmetry similarity matrix S_{ij} in LPP is defined as:

$$S_{ij} = \begin{cases} exp(-\|x_i - x_j\|^2/t), & \|x_i - x_j\|^2 < \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\epsilon > 0$ defines the radius of the local neighborhood. Here S_{ij} is actually heat kernel weight, the proof of such choice and how to select parameter t can be referred to [2].

The objective function in LPP incurs a heavy penalty if neighboring points x_i and x_j are mapped far apart. Therefore, the minimization of the objective function is an attempt to ensure that if x_i and x_j are "close", then y_i and y_j are close as well. The optimization will ultimately lead to the following generalized eigenvalue problem:

$$XLX^T w = \lambda XDX^T w \tag{3}$$

Let $w_0, w_1 \ldots, w_{k-1}$ be the solutions of Eq. (3), ordered according to their eigenvalues, $0 \leqslant \lambda_0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_{k-1}$. Then $W = [w_0, w_1 \ldots, w_{k-1}]$ is the final transformation projection matrix of LPP.