Neurocomputing 200 (2016) 35–46

Contents lists available at ScienceDirect

Neurocomputing

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

Sparsity induced locality preserving projection approaches for dimensionality reduction

Qi Zhang [a], Kuiying Deng [b], Tianguang Chu [a,*]

[a] State Key Laboratory for Turbulence and Complex Systems, College of Engineering, Peking University, Beijing 100871, China
[b] Department of Engineering Mechanics, Hebei University, Baoding 071002, China

ARTICLE INFO

Article history:
Received 18 June 2015
Received in revised form
8 November 2015
Accepted 8 March 2016
Available online 29 March 2016

Keywords:
Dimensionality reduction
Sparse subspace learning
Structural sparsity
Data classification
Face recognition

ABSTRACT

We consider the problem of sparse subspace learning for data classification and face recognition. New approaches called l_α-regularization-based sparse locality preserving projection (α-SLPP) and structural sparse locality preserving projection (SSLPP) are proposed by incorporating theories of sparse representation and structural sparse regularization into spectral embedding. The proposed methods can efficiently exploit the local geometric information of the data. Also, by inducing sparsity, they facilitate the interpretation of the projection results and the detection of more discriminating features for classification and recognition. In addition, α-SLPP induces sparsity by using non-convex l_α-norm regularizer, which is much closer to l_0-norm. SSLPP derives a more organized sparse pattern through structural sparse regularization, and thus overcomes the problem that merely decreasing the cardinality may not be enough in certain situations. We formulate the sparse subspace learning problem as feasible optimization problems and present efficient methods to solve them. Experiments in data classification, face recognition, and pixel-corrupted face recognition are carried out to verify the feasibility and effectiveness of the proposed approaches.

1. Introduction

Dimensionality reduction is an important issue in machine learning. It focuses on exploring subspaces that best preserve the information of the original data. Actually, most signal data, such as images, texts, and sounds are high-dimensional, which are usually redundant and may degrade the performance of pattern classifiers. For such cases, dimensionality reduction can be a useful approach to map the high-dimensional data into low-dimensional subspaces, yet retain most of the intrinsic information content in the original data. This facilitates data manipulation and visualization. So far, dimensionality reduction has found wide applications in many fields, such as artificial intelligence, data mining, and biomedical science [1–6].

Among the efficient dimensionality reduction methods, principal component analysis (PCA) [7] is one of the most popular. It projects the original data into a subspace spanned by the eigenvectors corresponding to the leading eigenvalues of the data covariance matrix. In recent years, numbers of researches have shown that the data resource (e.g., face images) may reside on a nonlinear submanifold embedded in the original high-dimensional space [8]. In this case, PCA and other global methods may suffer deviation in subspace learning because they compute the global Euclidean distances between data samples and cannot explore nonlinear features.

To explore the nonlinear features inhered in data, many approaches have been developed, including kernel-based methods, such as kernel principal component analysis (KPCA) [9], kernel Fisher discriminant analysis (KFDA) [10], and manifold-learning-based nonlinear methods, such as locally linear embedding (LLE) [11], isometric feature mapping (ISOMAP) [12], Laplacian eigenmaps (LE) [13], maximum variance unfolding (MVU) [14,15], and local tangent space alignment (LTSA) [16,17]. However, due to the implicity of the nonlinear map, these manifold learning approaches usually cannot readily yield the test samples' images in the embedding subspaces, in terms of the low-dimensional embedding results of the training data. This is referred to as "out-of-sample" problem [18]. Locality preserving projections (LPP) [19], as an efficient linear approach to solve this problem, preserves the local geometric structure of the data and approximates the eigenfunction of the Laplace Beltrami operator [13]. Data samples can then be readily mapped into the embedding subspace, with the nonlinear characters of the data structure well detected.

* Corresponding author.
E-mail addresses: zhang_qi@pku.edu.cn (Q. Zhang), rossdeng@pku.edu.cn (K. Deng), chutg@pku.edu.cn (T. Chu).

http://dx.doi.org/10.1016/j.neucom.2016.03.019
0925-2312/

Observe that a common feature of the linear approaches mentioned above is that the embedding results (i.e., low-dimensional projective data) linearly combine all the original data variables. So it is difficult to interpret the results that which variables or features play more important roles in feature extraction, especially when the number of the data variables is very large [20]. Actually, the interpretation of the projection results is of much importance especially when the variables have physical meanings in many applications (e.g., gene representation, face recognition). To facilitate interpretation, there are several ways such as the rotation technique used in [21] and the simple component method presented in [22], but the most popular way is to incorporate sparsity into dimensionality reduction, i.e., sparse subspace learning approaches. These methods explore sparse subspaces, whose basis vectors (i.e., also named as "loading vectors" [20]) are with few nonzero elements. Thus, the sparse subspace learning approaches can explicitly lead to sparse loadings and the zero-loading variables contribute nothing to the low-dimensional representations. This provides a way to interpret the low-dimensional results by finding features that linearly combine a small set of variables.

Actually, the sparse subspace learning problem has seen a surge of interests in recent years. Several sparse subspace learning approaches have been developed by taking advantage of the sparse representation theory [23–26]. For example, sparse PCA algorithm (SPCA) [27,28] regularizes the principle components by l_1-norm and is efficiently solved by least angle regression (LARS) [29]. Later, the hard cardinality constraint of SPCA was relaxed in [30]. Then, a spectral bounds framework for sparse subspace learning was proposed in [31]. And a unified sparse subspace learning (USSL) method for getting sparse projections in a regression framework was developed by [32]. Moreover, a number of researches incorporating sparsity into locality preserving analysis have been conducted, such as sparse locality preserving embedding (SLPE) method introduced in [33], the research of discriminant locality preserving projections based on $l1$-norm maximization (DLPP-L1) [34], the regression analysis of sparse locality preserving projection (spLPP) conducted in [35], and sparse local discriminant projections (SLDP) approach proposed in [36].

However, these sparse subspace learning algorithms all use l_1-regularization to induce sparsity due to the fact that l_1-norm is the convex surrogate of l_0-norm. Actually, recent studies have shown the advantages of using concave penalties in addressing sparsity problems, e.g., l_α $(0<\alpha<1)$ quasi-norms, which are much closer to l_0-norm and can induce a more aggressive penalization, especially when dealing with unregularized problems [37–39]. To our knowledge, there have been few l_α-regularization-based sparse subspace learning algorithms that incorporate the exploration of local geometric information inhered in data with non-convex sparse subspace learning. The purpose of this paper is to present some results in this direction. In particular, we will propose an l_α-regularization-based sparse locality preserving projection (α-SLPP) method, which is based on the locality preserving projection approach [19] with both "locality" and "sparsity" taken into account, and can be efficiently solved in a regression framework. To be specific, we apply the α-SLPP method by implementing successively the projection learning and l_α-regularization-based sparse subspace learning steps. We also present feasible approaches to solve the related optimization problem. The resulting sparse subspaces can efficiently exploit the nonlinear characters of the data structure and is also of much help in interpreting the projection results.

Another focus of our study is "sparse pattern". In many cases, it may not be enough to merely decrease the cardinality (namely, the number of nonzero elements) of the considered vectors, because the interrelation and structure information of the variables are also very important. For example, in face recognition, variables localized on specific positions in face images are naturally related to each other, and thus sets of pixels form small convex regions in face images. In genomics, factors explaining the gene expression patterns are expected to involve fewer other specific genes or groups of genes, in terms of biological pathways or crowds of genes that are neighbors in the protein–protein interaction network. For such cases, a plain l_1-norm or other l_α-norm regularizers will fail to encode these spatially local constraints [40].

Due to the above considerations, we further present a structural sparse locality preserving projection (SSLPP) method, which takes into account both "locality" and "structural sparsity". Unlike the structured sparse PCA (SSPCA) method [41] that preserves the global Euclidean structure of the data, SSLPP considers the local geometric information. So it can better explore the nonlinear characters of the data structure. In addition, by taking advantage of the structural-sparsity induced norms that analyzed recently in [40], SSLPP can derive structural sparse subspaces by imposing structural sparse constraints on the objective function. As a result, the obtained subspaces are not merely with reduced cardinalities, but also exhibit structural sparse patterns that characterize the interrelation and structure information of the variables. This enables SSLPP to find more discriminating variables or groups of variables.

To verify our approaches, we carry out experiments in data classification, face recognition, and pixel-corrupted face recognition. The main contributions of our work are as follows:

(1) α-SLPP and SSLPP take into account both "locality" and "sparsity". This is different from many existing sparse subspace learning methods that preserve the global Euclidean structure of the data. As a result, the present methods can efficiently exploit the nonlinear characters inhered in the data structure. Also, they can facilitate the interpretation of the projection results.
(2) α-SLPP applies the non-convex l_α-regularizations in sparse subspace learning process, which is different from most sparse subspace learning methods that use l_1-regularizations. Actually, a number of recent researches have demonstrated that l_α $(0<\alpha<1)$ quasi-norms are much closer to l_0-norm and can penalize more aggressively for small coefficients.
(3) SSLPP induces structural sparsity by taking the interrelation and structure information of the variables into account. This differs from most sparse subspace learning methods that merely decrease the cardinality of the projection vectors. As a result, SSLPP can achieve better results in cases where variables share interrelations with each other.

The paper is organized as follows. Section 2 reviews some existing dimensionality reduction methods and useful preliminaries. Sections 3 and 4 present the α-SLPP and SSLPP methods, respectively. A discussion about these methods is conducted in Section 5, followed by several experiments included in Section 6. Finally, we summarize in Section 7.

2. Related works

In this section, we give a brief introduction of the dimensionality reduction methods that are related to our work, and present some useful preliminaries that will be used in the following discussion.

Let the data matrix be given as $X=[\boldsymbol{x_1},\boldsymbol{x_2},\dots,\boldsymbol{x_m}]\in\mathbb{R}^{n\times m}$, with columns $\boldsymbol{x_i}$'s as the training samples for $i=1,\dots,m$.[1] The problem

[1] It is possible that $n>m$. For example, in face recognition, the original image space dimensionality may be much larger than the amount of the training samples.