Neural Networks 46 (2013) 190–198

Contents lists available at ScienceDirect

Neural Networks

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

2DPCA with L1-norm for simultaneously robust and sparse modelling

Haixian Wang *, Jing Wang

Key Laboratory of Child Development and Learning Science of Ministry of Education, Research Center for Learning Science, Southeast University, Nanjing, Jiangsu 210096, PR China

ARTICLE INFO

Article history:
Received 19 May 2012
Received in revised form 3 April 2013
Accepted 4 June 2013

Keywords:
Dimensionality reduction
2DPCA-L1
Robust modelling
Lasso regularization

ABSTRACT

Robust dimensionality reduction is an important issue in processing multivariate data. Two-dimensional principal component analysis based on L1-norm (2DPCA-L1) is a recently developed technique for robust dimensionality reduction in the image domain. The basis vectors of 2DPCA-L1, however, are still dense. It is beneficial to perform a sparse modelling for the image analysis. In this paper, we propose a new dimensionality reduction method, referred to as 2DPCA-L1 with sparsity (2DPCAL1-S), which effectively combines the robustness of 2DPCA-L1 and the sparsity-inducing lasso regularization. It is a sparse variant of 2DPCA-L1 for unsupervised learning. We elaborately design an iterative algorithm to compute the basis vectors of 2DPCAL1-S. The experiments on image data sets confirm the effectiveness of the proposed approach.

1. Introduction

Dimensionality reduction (DR) is of great importance for multivariate data analysis. For classifying typically high-dimensional patterns in practice, DR can relieve the "curse of dimensionality" effectively (Jain, Duin, & Mao, 2000). Principal component analysis (PCA) (Jolliffe, 1986) is perhaps the most popular DR technique. It seeks a few basis vectors such that the variances of projected samples are maximized. In the domain of image analysis, two-dimensional PCA (2DPCA) (Yang, Zhang, Frangi, & Yang, 2004) is more efficient, due to its direct formulation based on raw two-dimensional images.

Although PCA and 2DPCA have been widely applied in many fields, they are vulnerable at the presence of atypical samples because of the employment of the L2-norm in the variance formulation. As a robust alternative, L1-norm-based approaches were developed. Specifically, the L1-norm-based PCA variants include L1-PCA (Ke & Kanade, 2005), R1-PCA (Ding, Zhou, He, & Zha, 2006), PCA-L1 (Kwak, 2008), and non-greedy PCA-L1 (Nie, Huang, Ding, Luo, & Wang, 2011). Li, Pang, and Yuan (2009) developed the L1-norm-based 2DPCA (2DPCA-L1), which demonstrated encouraging performance for the image analysis.

A limitation of the above methods is that the basis vectors learned are still dense, which makes it difficult to explain the resulting features. It is desirable to select the most relevant or salient elements from a large number of features. To address this issue, sparse modelling has been developed and received increasing attention in the community of pattern classification (Wright et al., 2010). The sparsity was achieved by regularizing objective variables with a lasso penalty term using the L1-norm (Chen, Donoho, & Saunders, 1998; Tibshirani, 1996). Mathematically, the classic PCA approach could be reformulated as a regression-type optimization problem, and then the sparsity-inducing lasso penalty was imposed, resulting in sparse PCA (SPCA) (Zou, Hastie, & Tibshirani, 2006). The sparsity was further generalized to structured version, producing structured sparse PCA (Jenatton, Obozinski, & Bach, 2010). With the graph embedding platform (Yan et al., 2007), various DR approaches were endowed with a unified sparse framework by the L1-norm penalty (Cai, He, & Han, 2007; Wang, 2012; Zhou, Tao, & Wu, 2011). Recently, the robustness of SPCA was improved by the L1-norm maximization (Meng, Zhao, & Xu, 2012).

The sparse modelling for 2DPCA-L1, however, is still not addressed. Note that the L1-norm used in 2DPCA-L1 works as a robust measure of sample dispersion rather than regularizing basis vectors. A common way of enforcing sparsity is to fix the L2-norm and minimize the L1-norm with a length constraint.

In this paper, we limit our attention to the image analysis, and consider extending 2DPCA-L1 with sparsity, referred to as 2DPCAL1-S. On account of the L1-norm used as the lasso penalty in the sparsity-inducing modelling, we propose incorporating the L1-norm lasso penalty, together with the fixed L2-norm, onto the basis vectors of 2DPCA-L1. Consequently, 2DPCAL1-S maximizes the L1-dispersion of samples subject to the elastic net (i.e., L2-norm and L1-norm) (Zou et al., 2006) constraint onto the basis vectors. Formally, we combine the L1-dispersion and the elastic net constraint onto the objective function. As can be seen, we use the L1-norm for both robust and sparse modelling simultaneously. Due to the involvement of the L1-norm in the two aspects, the optimization of 2DPCAL1-S is not straightforward. We design an elegant iterative algorithm to solve 2DPCAL1-S.

* Corresponding author. Tel.: +86 25 83795664; fax: +86 25 83795929.
E-mail addresses: hxwang@seu.edu.cn, haixian_wang@hotmail.com (H. Wang).

0893-6080/$ – see front matter
http://dx.doi.org/10.1016/j.neunet.2013.06.002

The remainder of this paper is organized as follows. The conventional 2DPCA-L1 method is briefly reviewed in Section 2. The formulation of 2DPCAL1-S is proposed in Section 3. Section 4 reports experimental results. And Section 5 concludes the paper.

2. Brief review of 2DPCA-L1

The 2DPCA-L1 approach, proposed by Li et al. (2009), finds basis vectors that maximize the dispersion of projected image samples in terms of the L1-norm. Suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are a set of training images with size $q \times p$, where n is the number of images. These images are assumed to be mean-centred.

Let $\mathbf{v} \in \mathbb{R}^p$ be the first basis vector of 2DPCA-L1. It maximizes the L1-norm-based dispersion of projected samples

$$g(\mathbf{v}) = \sum_{i=1}^{n} \|\mathbf{X}_i \mathbf{v}\|_1 \tag{1}$$

subject to $\|\mathbf{v}\|_2 = 1$, where $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the L1-norm and the L2-norm, respectively. In this paper, for a vector $\mathbf{z} = (z_1, \ldots, z_n)^T$, its Ld-norm is specified as $\|\mathbf{z}\|_d = \left(\sum_{i=1}^{n} |z_i|^d\right)^{1/d}$. Let $\mathbf{x}_{ji} \in \mathbb{R}^p$ be the jth row vector of $\mathbf{X}_i$, i.e.,

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{1i}^T \\ \vdots \\ \mathbf{x}_{qi}^T \end{bmatrix}. \tag{2}$$

Then $g(\mathbf{v})$ can be rewritten as

$$g(\mathbf{v}) = \sum_{i=1}^{n} \sum_{j=1}^{q} |\mathbf{v}^T \mathbf{x}_{ji}|. \tag{3}$$

The computation of $\mathbf{v}$ is implemented by an iterative algorithm as follows. Denote by t the iteration number. The basis vector $\mathbf{v}(t+1)$ at the $(t+1)$th-step is updated according to

$$\mathbf{v}(t+1) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{q} s_{ji}(t) \mathbf{x}_{ji}}{\left\| \sum_{i=1}^{n} \sum_{j=1}^{q} s_{ji}(t) \mathbf{x}_{ji} \right\|_2}, \tag{4}$$

where $s_{ji}(t)$ is defined as

$$s_{ji}(t) = \mathrm{sign}(\mathbf{v}^T(t) \mathbf{x}_{ji}) \tag{5}$$

for $j = 1, \ldots, q;\ i = 1, \ldots, n$, where $\mathrm{sign}(\cdot)$ is the sign function. This iterative procedure was theoretically shown to converge to a local maximum value of $g(\mathbf{v})$ (Li et al., 2009). The reminder basis vectors are computed likewise by using the deflated samples with previously obtained basis vectors.

3. 2DPCA-L1 with sparsity

3.1. Basic idea

Sparse modelling has been receiving exploding attention in computer vision and pattern classification (Wright et al., 2010). The obtained basis vectors of 2DPCA-L1, however, are still dense (Li et al., 2009). In other words, the projection procedure involves all the original features. As we know, a typical image usually has a large number of features. There may exist irrelevant or redundant features for classification. It is important to find a few salient features, which correspond to specific parts of the image such as eyes or mouth of a face image. To select a set of representative features, the projection vectors are expected to have very sparse elements with respect to such features. Such sparse projection vectors, if learned correctly, could encode semantic information and thus deliver valuable discriminative information. The sparse modelling has been successfully applied to many classification problems (Wright et al., 2010).

It is desirable to learn sparse basis vectors for the purpose of classification. In light of the advantage of the L1-norm penalty in the sparse modelling (Chen et al., 1998; Tibshirani, 1996), we propose regularizing the basis vectors of 2DPCA-L1 using the L1-norm penalty together with the fixed L2-norm. We refer to the proposed approach as 2DPCAL1-S. It results in sparse basis vectors. Note that the L1-norm used in 2DPCAL1-S takes effect in two different perspectives: measuring dispersion and regularizing basis vectors. Computationally, we elaborately design an iterative algorithm to implement 2DPCAL1-S.

3.2. Objective function

We impose the sparsity-inducing L1-norm penalty, as well as the fixed L2-norm, onto the basis vector $\mathbf{v}$. Specifically, we integrate the elastic net into the objective function. The elastic net generalizes the L1-norm lasso penalty by combining the ridge penalty and can circumvent potential limitations of the lasso (Zou et al., 2006). Consequently, we wish to select a vector $\mathbf{v}$ such that the objective function

$$h(\mathbf{v}) = \sum_{i=1}^{n} \sum_{j=1}^{q} |\mathbf{v}^T \mathbf{x}_{ji}| - \frac{\eta}{2} \|\mathbf{v}\|_2^2 - \gamma \|\mathbf{v}\|_1, \tag{6}$$

is maximized, where η and γ are positive tuning parameters which are usually selected by cross validation. Due to the absolute value operation, it is not a direct issue to solve the optimization problem (6). We thus derive an iterative algorithm for optimization and show its monotonicity in the following two subsections.

3.3. Iterative algorithm

An iterative algorithm for 2DPCAL1-S is formally presented as follows. Let $\mathbf{v}(0)$ be the initial basis vector.

1. Let $t = 0$, and initialize $\mathbf{v}(t)$ as any p-dimensional vector.
2. Compute the quantity $s_{ji}(t)$ as in (5), which results in value 1, 0, or -1 depending on $\mathbf{v}^T(t)\mathbf{x}_{ji}$ larger than zero, equal to zero, or less than zero, respectively.
3. Let

   $$\mathbf{y}(t) = \sum_{i=1}^{n} \sum_{j=1}^{q} s_{ji}(t) \mathbf{x}_{ji}, \tag{7}$$

   and

   $$\mathbf{w}(t) = \left(\frac{|v_1(t)|}{\gamma + \eta |v_1(t)|}, \ldots, \frac{|v_p(t)|}{\gamma + \eta |v_p(t)|} \right)^T, \tag{8}$$

   where $v_k(t)$ is the kth entry of $\mathbf{v}(t)$ for $k = 1, \ldots, p$. Then, the basis vector $\mathbf{v}(t)$ is updated as

   $$\mathbf{v}(t+1) = \mathbf{y}(t) \circ \mathbf{w}(t), \tag{9}$$

   where $\circ$ denotes the element-wise product between two vectors.
4. If the objective function $h(\mathbf{v}(t+1))$ does not grow significantly, then stop the iterative procedure and set $\mathbf{v}^* = \mathbf{v}(t+1)$. Otherwise, set $t \leftarrow t+1$, and go to Step 2.
5. Output $\mathbf{v}^*$ as the basis vector.

The computational complexity of the above algorithm is $O(nqp)$ per iteration. Note that the update formula (9) can be further