Computational Statistics and Data Analysis 62 (2013) 26–38

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

# A coordinate descent MM algorithm for fast computation of sparse logistic PCA

Seokho Lee[a,*], Jianhua Z. Huang[b]

[a] *Department of Statistics, Hankuk University of Foreign Studies, Republic of Korea*

[b] *Department of Statistics, Texas A&M University, United States*

ARTICLE INFO

*Article history:*
Received 4 February 2012
Received in revised form 18 December 2012
Accepted 3 January 2013
Available online 9 January 2013

*Keywords:*
Binary data
Coordinate descent algorithm
MM algorithm
Penalized maximum likelihood
Principal component analysis

ABSTRACT

Sparse logistic principal component analysis was proposed in Lee et al. (2010) for exploratory analysis of binary data. Relying on the joint estimation of multiple principal components, the algorithm therein is computationally too demanding to be useful when the data dimension is high. We develop a computationally fast algorithm using a combination of coordinate descent and majorization–minimization (MM) auxiliary optimization. Our new algorithm decouples the joint estimation of multiple components into separate estimations and consists of closed-form elementwise updating formulas for each sparse principal component. The performance of the proposed algorithm is tested using simulation and high-dimensional real-world datasets.

## 1. Introduction

Principal component analysis (PCA) is a dimension reduction method that is widely used in data analysis. The direct application of PCA to a binary dataset is not sensible because the standard PCA exploits the Euclidean distance which is not compatible with binary variables. The logistic PCA model, as an extension of PCA to multivariate binary variables, has been studied in recent literature and several algorithms to fit the model have been proposed (Collins et al., 2002; Schein et al., 2003; de Leeuw, 2006; Lee et al., 2010). As in the sparse PCA for continuous variables (Zou et al., 2006; Shen and Huang, 2008), the sparsity-inducing penalization on the principal component loadings improves stability and interpretability in the logistic PCA, especially in high-dimensional and low-sample-size settings. Lee et al. (2010) proposed sparse logistic PCA for binary data and illustrated its application using simulated and real-world data. The method in Lee et al. (2010) is particularly applicable for the multivariate binary datasets where the number of variables exceeds the number of samples. These kind of datasets are frequently encountered in many research areas such as bioinformatics.

Lee et al. (2010) proposed an iterative algorithm to estimate principal component scores and principal component loadings for sparse logistic PCA. Since the algorithm involves matrix inversion in each iteration, it can be computationally inefficient when dealing with high-dimensional data. Moreover, because tuning multiple penalty parameters requires a grid search in multi-dimension and it is often computationally infeasible, they considered a single common penalty parameter for all components to reduce the computational burden. However, using single penalty parameter may not be flexible enough when the number of zeros in different principal component loading vectors are significantly different.

To overcome the computational challenges caused by the high dimensionality and multiple penalty parameter tuning, we propose a computationally fast new algorithm that improves the existing algorithm in many ways. The new algorithm

[*] Corresponding author. Tel.: +81 031 3304562.
*E-mail addresses:* leesh12@gmail.com, lees@hufs.ac.kr (S. Lee).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2013.01.001

combines two ideas: the auxiliary optimization using majorization–minimization (Lange et al., 2000) and coordinate descent (Tseng, 1988, 2001). The combination of the two ideas enables us to derive updating formulas in the iterations that just consist of simple vector inner products and thresholding rules. By fitting the principal components one by one, we are also able to reduce the multi-dimensional search of penalty parameters to several one-dimensional searches. Our numerical studies using simulated and real-world data have demonstrated that the new algorithm is faster than our old algorithm.

The rest of the paper is organized as follows. Section 2 reviews the formulation of sparse logistic PCA and the existing algorithm. Section 3 develops our new algorithm. Section 4 discusses modification of the new algorithm to incorporate different penalty functions. Sections 5 and 6 uses simulated and real data to demonstrate the performance of the new algorithm.

## 2. Review of sparse logistic PCA

We first describe the probabilistic model for the logistic PCA. Suppose we have an $n \times p$ binary data matrix $\mathbf{Y} = (y_{ij})$ each row of which represents a vector of observations from binary variables. We assume that entries of $\mathbf{Y}$ are realizations of mutually independent random variables and that $y_{ij}$ follows the Bernoulli distribution with success probability $\pi_{ij}$. The canonical parameter, $\theta_{ij} = \log\{\pi_{ij}/(1-\pi_{ij})\}$, is the logit transformation of $\pi_{ij}$. Define the inverse logit transformation $\pi(\theta) = \{1+\exp(-\theta)\}^{-1}$. Then the success probabilities can be represented using the canonical parameters as $\pi_{ij} = \pi(\theta_{ij})$. The individual data generating probability becomes $\Pr(Y_{ij} = y_{ij}) = \pi(\theta_{ij})^{y_{ij}}\{1-\pi(\theta_{ij})\}^{1-y_{ij}} = \pi(q_{ij}\theta_{ij})$ with $q_{ij} = 2y_{ij} - 1$. This representation leads to the compact form of the log likelihood as

$$\ell = \sum_{i=1}^{n}\sum_{j=1}^{p} \log \pi(q_{ij}\theta_{ij}). \tag{1}$$

In the logistic PCA, the $p$-dimensional canonical parameter vectors $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{ip})^T$ are constrained to reside in a low dimensional manifold of $\mathbb{R}^p$ with the dimensionality $k$ for some integer $k \le p$. Precisely, the canonical parameters satisfy $\boldsymbol{\theta}_i = \boldsymbol{\mu} + a_{i1}\tilde{\mathbf{b}}_1 + \cdots + a_{ik}\tilde{\mathbf{b}}_k$ for $i = 1, \ldots, n$. We call vectors of length $p$, $\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_k$, the principal component loading vectors and the coefficients $\mathbf{a}_i = (a_{i1}, \ldots, a_{ik})^T$ the principal component scores for the $i$th observation. In matrix form, the canonical parameter matrix $\boldsymbol{\Theta} = (\theta_{ij}) = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)^T$ is represented as

$$\boldsymbol{\Theta} = \mathbf{1}_n\boldsymbol{\mu}^T + \mathbf{A}\mathbf{B}^T, \tag{2}$$

where $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)^T$ is the $n \times k$ principal component score matrix and $\mathbf{B} = (\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_k)$ is the $p \times k$ principal component loading matrix. Let $\mathbf{b}_j$ denote the $j$th row of $\mathbf{B}$ and $\mu_j$ denote the $j$th element of $\boldsymbol{\mu}$. Then (2) implies that $\theta_{ij} = \mu_j + \mathbf{a}_i^T\mathbf{b}_j$. The log likelihood can be written as

$$\ell(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}) = \sum_{i=1}^{n}\sum_{j=1}^{p} \log \pi\{q_{ij}(\mu_j + \mathbf{a}_i^T\mathbf{b}_j)\}. \tag{3}$$

The logistic PCA is performed by maximizing the log likelihood over the parameters $\boldsymbol{\mu}$, $\mathbf{A}$, and $\mathbf{B}$. This formulation of logistic PCA can be viewed as an extension of the standard PCA for continuous variables. In particular, if the Bernoulli likelihood (3) is replaced by a normal likelihood, one recovers the standard PCA (Lee et al., 2010).

Inspired by the lasso regression (Tibshirani, 1996), Lee et al. (2010) proposed to obtain a sparse principal component loading matrix by minimizing the negative penalized log likelihood criterion

$$-\ell_p(\boldsymbol{\mu}, \mathbf{A}, \mathbf{B}) = -\sum_{i=1}^{n}\sum_{j=1}^{p} \log \pi\{q_{ij}(\mu_j + \mathbf{a}_i^T\mathbf{b}_j)\} + \frac{1}{4}\sum_{m=1}^{k} \lambda_m \|\tilde{\mathbf{b}}_m\|_1, \tag{4}$$

where $\|\tilde{\mathbf{b}}_m\|_1 = \sum_{j=1}^{p} |b_{jm}|$ is the $L_1$ norm of the $m$th principal component loading vector. The constant factor in front of the penalties, 1/4, is intentionally introduced for obtaining cleaner formula in the computational algorithm. Note that penalties are imposed on the columns of matrix $\mathbf{B}$. The use of $L_1$ penalization in PCA has two advantages: First, it provides the regularization to ensure stable extraction of the principal components, especially in high-dimensional situations. Second, the $L_1$ penalty results in many zero elements in the principal component loading vectors and thus makes it easy to interpret the results (Zou et al., 2006; Shen and Huang, 2008). These benefits of using the $L_1$ penalization have been demonstrated in simulation studies and data analysis in Lee et al. (2010). The algorithm developed in this paper can incorporate general sparsity-inducing penalties other than the lasso penalty; see Section 4.

For stable minimization of the negative penalized log likelihood, Lee et al. (2010) proposed an Majorization–Minimization (MM) algorithm (Lange et al., 2000), extending an earlier algorithm by de Leeuw (2006) that does not consider the penalty term. In the framework of the MM algorithm, the minimization of (4) is turned into the minimization of a quadratic auxiliary function. To obtain a quadratic auxiliary function for the negative Bernoulli log likelihood, we use the relationship

$$-\log \pi(z) \le -\log \pi(z^o) + 2\{1-\pi(z^o)\}^2 + \frac{1}{8}[z - z^o - 4\{1-\pi(z^o)\}]^2 \tag{5}$$

for all $z$ at any given $z^o$. The quadratic upper bound on the right-hand side is tangent to $-\log \pi(z)$ at $z^o$. This enables us to construct a quadratic upper bound for the negative penalized log likelihood with the previous parameter estimates as a