Journal of Multivariate Analysis 128 (2014) 165–185

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Graphical model selection and estimation for high dimensional tensor data

Shiyuan He [a], Jianxin Yin [a,*], Hongzhe Li [b], Xing Wang [a]

[a] Center for Applied Statistics and School of Statistics, Renmin University of China, No. 59 Zhongguancun Street, Haidian District, Beijing 100872, China

[b] Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6021, USA

ARTICLE INFO

Article history:
Received 12 December 2012
Available online 31 March 2014

AMS subject classifications:
62F10
62F12

Keywords:
Gaussian graphical model
Gene networks
l_1 penalized likelihood
Oracle property
Tensor normal distribution

ABSTRACT

Multi-way tensor data are prevalent in many scientific areas such as genomics and biomedical imaging. We consider a K-way tensor-normal distribution, where the precision matrix for each way has a graphical interpretation. We develop an l_1 penalized maximum likelihood estimation and an efficient coordinate descent-based algorithm for model selection and estimation in such tensor normal graphical models. When the dimensions of the tensor are fixed, we drive the asymptotic distributions and oracle property for the proposed estimates of the precision matrices. When the dimensions diverge as the sample size goes to infinity, we present the rates of convergence of the estimates and sparsistency results. Simulation results demonstrate that the proposed estimation procedure can lead to better estimates of the precision matrices and better identifications of the graph structures defined by the precision matrices than the standard Gaussian graphical models. We illustrate the methods with an analysis of yeast gene expression data measured over different time points and under different experimental conditions.

1. Introduction

An increasing number of statistical and data mining problems involves analysis of data that are indexed by more than one way. This type of data is often called the multidimensional matrix, multi-way array or tensor [2]. Recently high-dimensional tensor data have become prevalent in many scientific areas such as genomics, biomedical imaging, remote sensing, bibliometrics, chemometrics and internet. Take a two-way $n \times p$ data matrix as an example, if n samples are not independent, their correlations should be taken into consideration in statistical modeling, which leads to a transposable matrix [1]. In genomic experiments, gene expression data are often collected at different time points during the cell cycle process and under varying experimental conditions. This gives rise to a 3-way tensor data [8]. In social-economics studies, export of commodity k from country i to country j at year t [4] defines a three-way tensor data.

Statistical methods for tensor data analysis are limited. Omberg et al. [8] developed tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. Tucker and parallel factor analysis (PARAFAC) are useful methods for tensor decomposition [5]. When modeling high dimensional tensor data, a separable covariance matrix structure is often assumed. Such a separable structure on the covariance matrix can dramatically reduce the dimension of the parameter space. Consider a four-way tensor data, suppose that the dimensions are $m_1 = m_2 = m_3 = 100$

* Corresponding author.
E-mail addresses: jyin@ruc.edu.cn, jianxinyin@gmail.com (J. Yin).

http://dx.doi.org/10.1016/j.jmva.2014.03.007
0047-259X/

and $m_4 = 10$. The nonseparable model requires a joint covariance matrix of $10^7 \times 10^7$ entries, while the separable model requires only three 100×100 matrices and one 10×10 matrix for each way. The joint covariance matrix is simply the Kronecker product of the matrices over all dimensions. The ratio of dimension between two models is almost of the order of 10^{10}.

In this paper, we consider sparse modeling of the precision matrices of K-way tensor data, assuming a separable covariance matrix structure. The corresponding precision matrices define graphical models for tensor data. In many applications, sparsity in each of the corresponding precision matrices can be assumed to facilitate the interpretation. In addition, tensor normality is a natural assumption for the data distribution when the data are continuous [4]. With the separability assumption on the covariance matrix, the joint covariance matrix of the vectorization of the tensor can be obtained by a Kronecker product of K covariance matrices.

When $K = 2$, the 2-way normal tensor data are also called matrix normal data. Yin and Li [13] discussed the sparse model selection and estimation for the matrix normal distribution using a penalized likelihood approach with Lasso and adaptive Lasso penalties. In their work, the dimensions for row and column can diverge to infinity when the sample size goes to infinity. Other related works in modeling matrix-normal data include [1,16,15,12].

In this paper, we generalize the work by Yin and Li [13] to K-way tensor data and focus our work on graphical model selection and estimation. We develop a penalized maximum likelihood estimation with an adaptive Lasso penalty. The consistency and oracle property are obtained when the tensor dimensions hold fixed. In addition, we derive the rate of convergence and prove sparsistency of the estimates when the dimensions diverge with sample size going to infinity. We further show that the effective sample size for estimating the covariance matrix in each way of the tensor is the product of the number of independent samples and the dimensions of the other $K-1$ matrices. It is worth noting that this effective sample size is usually very large, hence the convergence is quite fast and the high dimension is actually a bless. Our simulation study demonstrates the high accuracy in estimating the precision matrices with small sample size N.

The rest of the paper is organized as follows. A brief summary of multi-way tensor data is presented in Section 2. Section 3 introduces the definition of the array normal distribution of [4] and its estimation in high dimensional settings. The convexity and optimization of the objective function is discussed in Section 4. In Section 5, the asymptotic properties are derived both for the case of fixed dimensions and the case of diverging dimensions when the sample size goes to infinity. A Monte Carlo simulation study is presented in Section 6. Finally, a 3-way tensor data set on gene expressions [8] is analyzed in Section 7.

2. Multi-way tensor data structure and operations

This section presents a brief summary of multi-way array data or high order tensor data [4,2]. Tensor data are higher order parallels of vector and matrix. Entries in a vector can be indexed by a single index set, while a matrix is indexed by two sets (row and column). In the following presentation, we use non-bold italic letters for scalars, bold-faced lower case letters for vectors, and bold-faced capitals for matrices and the multi-way tensor. For a matrix $\mathbf{A}$, we use $\mathbf{a}(j)$ to denote its j-th column, $\mathbf{a}[i]$ its i-th row, and $A(i,j)$ its (i,j)-th element. Standard matrix identities and inequalities used in this paper can be found in [9].

A K-way tensor is an arrangement of elements, which is indexed by K sets. Suppose $\mathbf{Y}$ is a K-way tensor with dimensions $\{m_1, m_2, \ldots, m_K\}$, then the total number of elements of $\mathbf{Y}$ is $m = m_1 \times m_2 \times \cdots \times m_K$. All the elements in Y are

$$\{y_{(i_1,\ldots,i_K)} : i_k = 1, 2, \ldots, m_k;\ k = 1, 2, \ldots, K\}.$$

Clearly, $\mathbf{Y}$ is a vector when $K = 1$ and a matrix when $K = 2$. We further introduce the notation $\mathbf{Y}_{(\ldots,i_k^0,\ldots)}$, which is a $(K-1)$-subarray of $\mathbf{Y}$. Specifically, $\mathbf{Y}_{(\ldots,i_k^0,\ldots)}$ has the same elements as $\mathbf{Y}$, except that its k-th sub-index is fixed at i_k^0. In other words, all the elements in $\mathbf{Y}_{(\ldots,i_k^0,\ldots)}$ are

$$\{y_{(i_1,\ldots,i_k^0,\ldots,i_K)} : i_h = 1, 2, \ldots, m_h;\ h = 1, 2, \ldots, k-1, k+1, \ldots, K\}.$$

To analyze the properties of the K-way tensor, it is helpful to relate the tensor with vector or matrix. The vectorization of $\mathbf{Y}$ is a vector of dimension m,

$$\begin{aligned}\mathrm{vec}(\mathbf{Y}) = \big(&y_{(1,1,1,\ldots,1)}, y_{(2,1,1,\ldots,1)}, \ldots, y_{(m_1,1,1,\ldots,1)},\\ &y_{(1,2,1,\ldots,1)}, y_{(2,2,1,\ldots,1)}, \ldots, y_{(m_1,2,1,\ldots,1)},\\ &\ldots,\\ &y_{(1,m_2,1,\ldots,1)}, y_{(2,m_2,1,\ldots,1)}, \ldots, y_{(m_1,m_2,1,\ldots,1)},\\ &\ldots,\\ &y_{(1,m_2,m_3,\ldots,m_K)}, y_{(2,m_2,m_3,\ldots,m_K)}, \ldots, y_{(m_1,m_2,\ldots,m_K)}\big)^T.\end{aligned}$$

To be explicit, $y_{(i_1,\ldots,i_K)}$ is the j-th element of $\mathrm{vec}(\mathbf{Y})$ with

$$j = \sum_{k=2}^{K}\Big[(i_k - 1)\Big(\prod_{l=1}^{k-1} m_l\Big)\Big] + i_1.$$

On the other hand, k-mode matrix unfolding results in a $m_k \times (m/m_k)$ matrix, $\mathbf{Y}_{(k)}$, whose i_k^0-th row is $[\mathrm{vec}(\mathbf{Y}_{(\ldots,i_k^0,\ldots)})]^T$ for $i_k^0 = 1, 2, \ldots, m_k$.