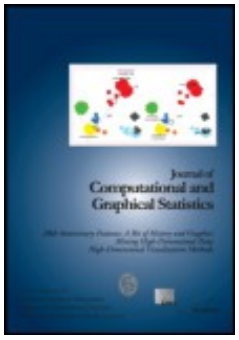

Journal of computational and graphical statistics, ISSN 1061-8600
Volume 26, number 4 (december 2017)

Abstract
The joint optimization of fidelity and commensurability (JOFC) manifold matching methodology embeds an omnibus dissimilarity matrix consisting of multiple dissimilarities on the same set of objects. One approach to this embedding optimizes the preservation of fidelity to each individual dissimilarity matrix together with commensurability of each given observation across modalities via iterative majorization of a raw stress error criterion by successive Guttman transforms. In this article, we exploit the special structure inherent to JOFC to exactly and efficiently compute the successive Guttman transforms, and as a result we are able to greatly speed up the JOFC procedure for both in-sample and out-of-sample embedding. We demonstrate the scalability of our implementation on both real and simulated data examples.

Abstract
Feature screening plays an important role in dimension reduction for ultrahigh-dimensional data. In this article, we introduce a new feature screening method and establish its sure independence screening property under the ultrahigh-dimensional setting. The proposed method works based on the nonparanormal transformation and Henze–Zirkler's test, that is, it first transforms the response variable and features to Gaussian random variables using the nonparanormal transformation and then tests the dependence between the response variable and features using the Henze–Zirkler's test. The proposed method enjoys at least two merits. First, it is model-free, which avoids the specification of a particular model structure. Second, it is condition-free, which does not require any extra conditions except for some regularity conditions for high-dimensional feature screening. The numerical results indicate that, compared to the existing methods, the proposed method is more robust to the data generated from heavy-tailed distributions and/or complex models with interaction variables. The proposed method is applied to screening of anticancer drug response genes. Supplementary material for this article is available online.

Abstract
In multivariate regression models, a sparse singular value decomposition of the regression component matrix is appealing for reducing dimensionality and facilitating interpretation. However, the recovery of such a decomposition remains very challenging, largely due to the simultaneous presence of orthogonality constraints and co-sparsity regularization. By delving into the underlying statistical data-generation mechanism, we reformulate the problem as a supervised co-sparse factor analysis, and develop an efficient computational procedure, named sequential factor extraction via co-sparse unit-rank estimation (SeCURE), that completely bypasses the orthogonality requirements. At each step, the problem reduces to a sparse multivariate regression with a unit-rank constraint. Nicely, each

sequentially extracted sparse and unit-rank coefficient matrix automatically leads to co-sparsity in its pair of singular vectors. Each latent factor is thus a sparse linear combination of the predictors and may influence only a subset of responses. The proposed algorithm is guaranteed to converge, and it ensures efficient computation even with incomplete data and/or when enforcing exact orthogonality is desired. Our estimators enjoy the oracle properties asymptotically; a non-asymptotic error bound further reveals some interesting finite-sample behaviors of the estimators. The efficacy of SeCURE is demonstrated by simulation studies and two applications in genetics. Supplementary materials for this article are available online.

Bayesian Dimensionality Reduction With PCA Using Penalized Semi-Integrated Likelihood P. 826-839

Piotr Sobczyk, Małgorzata Bogdan & Julie Josse

Abstract

We discuss the problem of estimating the number of principal components in principal components analysis (PCA). Despite the importance of the problem and the multitude of solutions proposed in literature, it comes as a surprise that there does not exist a coherent asymptotic framework, which would justify different approaches depending on the actual size of the dataset. In this article, we address this issue by presenting an approximate Bayesian approach based on Laplace approximation and introducing a general method of developing criteria for model selection, called PEnalized SEmi-integrated Likelihood (PESEL). Our general framework encompasses a variety of existing approaches based on probabilistic models, like the Bayesian Information Criterion for Probabilistic PCA (PPCA), and enables the construction of new criteria, depending on the size of the dataset at hand and additional prior information. Specifically, we apply PESEL to derive two new criteria for datasets where the number of variables substantially exceeds the number of observations, which is out of the scope of currently existing approaches. We also report results of extensive simulation studies and real data analysis, which illustrate the desirable properties of our proposed criteria as compared to state-of-the-art methods and very recent proposals. Specifically, these simulations show that PESEL-based criteria can be quite robust against deviations from the assumptions of a probabilistic model. Selected PESEL-based criteria for the estimation of the number of principal components are implemented in the R package *pesel*, which is available on github (*https://github.com/psobczyk/pesel*). Supplementary material for this article, with additional simulation results, is available online.

Bayesian Fused Lasso Regression for Dynamic Binary Networks P. 840-850

Brenda Betancourt, Abel Rodríguez & Naomi Boyd

Abstract

We propose a multinomial logistic regression model for link prediction in a time series of directed binary networks. To account for the dynamic nature of the data, we employ a dynamic model for the model parameters that is strongly connected with the fused lasso penalty. In addition to promoting sparseness, this prior allows us to explore the presence of change points in the structure of the network. We introduce fast computational algorithms for estimation and prediction using both optimization and Bayesian approaches. The performance of the model is illustrated using simulated data and data from a financial trading network in the NYMEX natural gas futures market. Supplementary material containing the trading network dataset and code to implement the algorithms is available online.

Large-Scale Structured Sparsity via Parallel Fused Lasso on Multiple GPUs P. 851-864

Taehoon Lee, Joong-Ho Won, Johan Lim & Sungroh Yoon

Abstract

We present a massively parallel algorithm for the fused lasso, powered by a multiple number of graphics processing units (GPUs). Our method is suitable for a class of large-scale sparse regression problems on which a two-dimensional lattice structure among the coefficients is imposed. This structure is important in many statistical applications, including image-based regression in which a set of images are used to locate image regions predictive of a response

variable such as human behavior. Such large datasets are increasingly common. In our study, we employ the split Bregman method and the fast Fourier transform, which jointly have a high data-level parallelism that is distinct in a two-dimensional setting. Our multi-GPU parallelization achieves remarkably improved speed. Specifically, we obtained as much as 433 times improved speed over that of the reference CPU implementation. We demonstrate the speed and scalability of the algorithm using several datasets, including 8100 samples of 512 × 512 images. Compared to the single GPU counterpart, our method also showed improved computing speed as well as high scalability. We describe the various elements of our study as well as our experience with the subtleties in selecting an existing algorithm for parallelization. It is critical that memory bandwidth be carefully considered for multi-GPU algorithms. Supplementary material for this article is available online.

Large-Scale Structured Sparsity via Parallel Fused Lasso on Multiple GPUs P. 851-864

Taehoon Lee, Joong-Ho Won, Johan Lim & Sungroh Yoon

Abstract

We present a massively parallel algorithm for the fused lasso, powered by a multiple number of graphics processing units (GPUs). Our method is suitable for a class of large-scale sparse regression problems on which a two-dimensional lattice structure among the coefficients is imposed. This structure is important in many statistical applications, including image-based regression in which a set of images are used to locate image regions predictive of a response variable such as human behavior. Such large datasets are increasingly common. In our study, we employ the split Bregman method and the fast Fourier transform, which jointly have a high data-level parallelism that is distinct in a two-dimensional setting. Our multi-GPU parallelization achieves remarkably improved speed. Specifically, we obtained as much as 433 times improved speed over that of the reference CPU implementation. We demonstrate the speed and scalability of the algorithm using several datasets, including 8100 samples of 512 × 512 images. Compared to the single GPU counterpart, our method also showed improved computing speed as well as high scalability. We describe the various elements of our study as well as our experience with the subtleties in selecting an existing algorithm for parallelization. It is critical that memory bandwidth be carefully considered for multi-GPU algorithms. Supplementary material for this article is available online.

Improving the Graphical Lasso Estimation for the Precision Matrix Through Roots of the Sample Covariance Matrix P. 865-872

Vahe Avagyan, Andrés M. Alonso & Francisco J. Nogales

Abstract

In this article, we focus on the estimation of a high-dimensional inverse covariance (i.e., precision) matrix. We propose a simple improvement of the graphical Lasso (glasso) framework that is able to attain better statistical performance without increasing significantly the computational cost. The proposed improvement is based on computing a root of the sample covariance matrix to reduce the spread of the associated eigenvalues. Through extensive numerical results, using both simulated and real datasets, we show that the proposed modification improves the glasso procedure. Our results reveal that the square-root improvement can be a reasonable choice in practice. Supplementary material for this article is available online.

Variational Bayes With Intractable Likelihood P. 873-645

Minh-Ngoc Tran, David J. Nott & Robert Kohn

Abstract

Variational Bayes (VB) is rapidly becoming a popular tool for Bayesian inference in statistical modeling. However, the existing VB algorithms are restricted to cases where the likelihood is tractable, which precludes their use in many interesting situations such as in state–space models and in approximate Bayesian computation (ABC), where application of VB methods was previously impossible. This article extends the scope of application of VB to cases where the likelihood is intractable, but can be estimated unbiasedly. The proposed VB method therefore makes it

possible to carry out Bayesian inference in many statistical applications, including state–space models and ABC. The method is generic in the sense that it can be applied to almost all statistical models without requiring too much model-based derivation, which is a drawback of many existing VB algorithms. We also show how the proposed method can be used to obtain highly accurate VB approximations of marginal posterior distributions. Supplementary material for this article is available online.

Depth-Based Recognition of Shape Outlying Functions P. 883-893

Stanislav Nagy, Irène Gijbels & Daniel Hlubinka

Abstract

A major drawback of many established depth functionals is their ineffectiveness in identifying functions outlying merely in shape. Herein, a simple modification of functional depth is proposed to provide a remedy for this difficulty. The modification is versatile, widely applicable, and introduced without imposing any assumptions on the data, such as differentiability. It is shown that many favorable attributes of the original depths for functions, including consistency properties, remain preserved for the modified depths. The powerfulness of the new approach is demonstrated on a number of examples for which the known depths fail to identify the outlying functions. Supplementary material for this article is available online.

Bayesian Registration of Functions With a Gaussian Process Prior P. 894-904

Yi Lu, Radu Herbei & Sebastian Kurtek

Abstract

We present a Bayesian framework for registration of real-valued functional data. At the core of our approach is a series of transformations of the data and functional parameters, developed under a differential geometric framework. We aim to avoid discretization of functional objects for as long as possible, thus minimizing the potential pitfalls associated with high-dimensional Bayesian inference. Approximate draws from the posterior distribution are obtained using a novel Markov chain Monte Carlo (MCMC) algorithm, which is well suited for estimation of functions. We illustrate our approach via pairwise and multiple functional data registration, using both simulated and real datasets. Supplementary material for this article is available online.

Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models P. 905-917

Gregor Kastner, Sylvia Frühwirth-Schnatter & Hedibert Freitas Lopes

Abstract

We discuss efficient Bayesian estimation of dynamic covariance matrices in multivariate time series through a factor stochastic volatility model. In particular, we propose two interweaving strategies to substantially accelerate convergence and mixing of standard MCMC approaches. Similar to marginal data augmentation techniques, the proposed acceleration procedures exploit nonidentifiability issues which frequently arise in factor models. Our new interweaving strategies are easy to implement and come at almost no extra computational cost; nevertheless, they can boost estimation efficiency by several orders of magnitude as is shown in extensive simulation studies. To conclude, the application of our algorithm to a 26-dimensional exchange rate dataset illustrates the superior performance of the new approach for real-world data. Supplementary materials for this article are available online.

Efficient Data Augmentation for Fitting Stochastic Epidemic Models to Prevalence Data P. 918-929

Jonathan Fintzi, Xiang Cui, Jon Wakefield & Vladimir N. Minin

Abstract

Stochastic epidemic models describe the dynamics of an epidemic as a disease spreads through a population. Typically, only a fraction of cases are observed at a set of discrete times. The absence of complete information about

the time evolution of an epidemic gives rise to a complicated latent variable problem in which the state space size of the epidemic grows large as the population size increases. This makes analytically integrating over the missing data infeasible for populations of even moderate size. We present a data augmentation Markov chain Monte Carlo (MCMC) framework for Bayesian estimation of stochastic epidemic model parameters, in which measurements are augmented with subject-level disease histories. In our MCMC algorithm, we propose each new subject-level path, conditional on the data, using a time-inhomogenous continuous-time Markov process with rates determined by the infection histories of other individuals. The method is general, and may be applied to a broad class of epidemic models with only minimal modifications to the model dynamics and/or emission distribution. We present our algorithm in the context of multiple stochastic epidemic models in which the data are binomially sampled prevalence counts, and apply our method to data from an outbreak of influenza in a British boarding school. Supplementary material for this article is available online.

Abstract

Recurrence relations for integrals that involve the density of multivariate normal distributions are developed. These recursions allow fast computation of the moments of folded and truncated multivariate normal distributions. Besides being numerically efficient, the proposed recursions also allow us to obtain explicit expressions of low-order moments of folded and truncated multivariate normal distributions. Supplementary material for this article is available online.

Abstract

Penalized quantile regression (PQR) provides a useful tool for analyzing high-dimensional data with heterogeneity. However, its computation is challenging due to the nonsmoothness and (sometimes) the nonconvexity of the objective function. An iterative coordinate descent algorithm (QICD) was recently proposed to solve PQR with nonconvex penalty. The QICD significantly improves the computational speed but requires a double-loop. In this article, we propose an alternative algorithm based on the alternating direction method of multiplier (ADMM). By writing the PQR into a special ADMM form, we can solve the iterations exactly without using coordinate descent. This results in a new single-loop algorithm, which we refer to as the QPADM algorithm. The QPADM demonstrates favorable performance in both computational speed and statistical accuracy, particularly when the sample size n and/or the number of features p are large. Supplementary material for this article is available online.