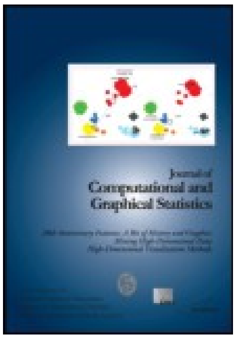

Journal of computational and graphical statistics, ISSN 1061-8600
Volume 25, number 4 (december 2016)

Convex Modeling of Interactions With Strong Heredity P. 981-1004

Asad Haris, Daniela Witten & Noah Simon

Abstract

We consider the task of fitting a regression model involving interactions among a potentially large set of covariates, in which we wish to enforce strong heredity. We propose FAMILY, a very general framework for this task. Our proposal is a generalization of several existing methods, such as VANISH, hierNet, the all-pairs lasso, and the lasso using only main effects. It can be formulated as the solution to a convex optimization problem, which we solve using an efficient alternating directions method of multipliers (ADMM) algorithm. This algorithm has guaranteed convergence to the global optimum, can be easily specialized to any convex penalty function of interest, and allows for a straightforward extension to the setting of generalized linear models. We derive an unbiased estimator of the degrees of freedom of FAMILY, and explore its performance in a simulation study and on an HIV sequence dataset. Supplementary materials for this article are available online.

Fused Lasso Additive Model P. 1005-1025

Ashley Petersen, Daniela Witten & Noah Simon

Abstract

We consider the problem of predicting an outcome variable using p covariates that are measured on n independent observations, in a setting in which additive, flexible, and interpretable fits are desired. We propose the *fused lasso additive model* (FLAM), in which each additive function is estimated to be piecewise constant with a small number of adaptively chosen knots. FLAM is the solution to a convex optimization problem, for which a simple algorithm with guaranteed convergence to a global optimum is provided. FLAM is shown to be consistent in high dimensions, and an unbiased estimator of its degrees of freedom is proposed. We evaluate the performance of FLAM in a simulation study and on two datasets. Supplemental materials are available online, and the R package flam is available on CRAN.

Sparse Partially Linear Additive Models P. 1126-1140

Yin Lou, Jacob Bien, Rich Caruana & Johannes Gehrke

Abstract

The generalized partially linear additive model (GPLAM) is a flexible and interpretable approach to building predictive models. It combines features in an additive manner, allowing each to have either a linear or nonlinear effect on the response. However, the choice of which features to treat as linear or nonlinear is typically assumed known. Thus, to make a GPLAM a viable approach in situations in which little is known *a priori* about the features, one must overcome two primary model selection challenges: deciding which features to include in the model and determining which of these features to treat nonlinearly. We introduce the sparse partially linear additive model (SPLAM), which combines model fitting and *both* of these model selection challenges into a single convex optimization problem. SPLAM provides a bridge between the lasso and sparse additive models. Through a statistical oracle inequality and thorough simulation, we demonstrate that SPLAM can outperform other methods across a broad spectrum of statistical regimes, including the high-dimensional ($p \gg N$) setting. We develop efficient algorithms that are applied to real datasets with half a million samples and over 45,000 features with excellent predictive performance. Supplementary materials for this article are available online.

Lina Liao, Cheolwoo Park, Jan Hannig & Kee-Hoon Kang

Abstract

The work revisits the autocovariance function estimation, a fundamental problem in statistical inference for time series. We convert the function estimation problem into constrained penalized regression with a generalized penalty that provides us with flexible and accurate estimation, and study the asymptotic properties of the proposed estimator. In case of a nonzero mean time series, we apply a penalized regression technique to a differenced time series, which does not require a separate detrending procedure. In penalized regression, selection of tuning parameters is critical and we propose four different data-driven criteria to determine them. A simulation study shows effectiveness of the tuning parameter selection and that the proposed approach is superior to three existing methods. We also briefly discuss the extension of the proposed approach to interval-valued time series. Supplementary materials for this article are available online.

Xiaoyue Cheng, Dianne Cook & Heike Hofmann

Abstract

Temporal data are information measured in the context of time. This contextual structure provides components that need to be explored to understand the data and that can form the basis of interactions applied to the plots. In multivariate time series, we expect to see temporal dependence, long term and seasonal trends, and cross-correlations. In longitudinal data, we also expect within and between subject dependence. Time series and longitudinal data, although analyzed differently, are often plotted using similar displays. We provide a taxonomy of interactions on plots that can enable exploring temporal components of these data types, and describe how to build these interactions using data transformations. Because temporal data are often accompanied other types of data we also describe how to link the temporal plots with other displays of data. The ideas are conceptualized into a data pipeline for temporal data and implemented into the R package cranvas. This package provides many different types of interactive graphics that can be used together to explore data or diagnose a model fit.

Richard A. Davis, Pengfei Zang & Tian Zheng

Abstract

The vector autoregressive (VAR) model has been widely used for modeling temporal dependence in a multivariate time series. For large (and even moderate) dimensions, the number of the AR coefficients can be prohibitively large, resulting in noisy estimates, unstable predictions, and difficult-to-interpret temporal dependence. To overcome such drawbacks, we propose a two-stage approach for fitting sparse VAR (sVAR) models in which many of the AR coefficients are zero. The first stage selects nonzero AR coefficients based on an estimate of the partial spectral coherence (PSC) together with the use of BIC. The PSC is useful for quantifying the conditional relationship between marginal series in a multivariate process. A refinement second stage is then applied to further reduce the number of parameters. The performance of this two-stage approach is illustrated with simulation and real data examples. Supplementary materials for this article are available online.

Clustering Multivariate Longitudinal Observations: The Contaminated Gaussian Hidden Markov Model P. 1097-1098

Antonio Punzo & Antonello Maruotti

Abstract

The Gaussian hidden Markov model (HMM) is widely considered for the analysis of heterogenous continuous multivariate longitudinal data. To robustify this approach with respect to possible elliptical heavy-tailed departures from normality, due to the presence of outliers, spurious points, or noise (collectively referred to as *bad points* herein), the contaminated Gaussian HMM is here introduced. The contaminated Gaussian distribution represents an elliptical generalization of the Gaussian distribution and allows for automatic detection of bad points in the same natural way as observations are typically assigned to the latent states in the HMM context. Once the model is fitted, each observation has a posterior probability of belonging to a particular state and, inside each state, of being a bad point or not. In addition to the parameters of the classical Gaussian HMM, for each state we have two more parameters, both with a specific and useful interpretation: one controls the proportion of bad points and one specifies their degree of atypicality. A sufficient condition for the identifiability of the model is given, an expectation-conditional maximization algorithm is outlined for parameter estimation and various operational issues are discussed. Using a large-scale simulation study, but also an illustrative artificial dataset, we demonstrate the effectiveness of the proposed model in comparison with HMMs of different elliptical distributions, and we also evaluate the performance of some well-known information criteria in selecting the true number of latent states. The model is finally used to fit data on criminal activities in Italian provinces. Supplementary materials for this article are available online

Mixture Modeling for Longitudinal Data P. 1117-1137

Xiwei Tang & Annie Qu

Abstract

In this article, we propose an unbiased estimating equation approach for a two-component mixture model with correlated response data. We adapt the mixture-of-experts model and a generalized linear model for component distribution and mixing proportion, respectively. The new approach only requires marginal distributions of both component densities and latent variables. We use serial correlations from subjects' subgroup memberships, which improves estimation efficiency and classification accuracy, and show that estimation consistency does not depend on the choice of the working correlation matrix. The proposed estimating equation is solved by an expectation-estimating-equation (EEE) algorithm. In the E-step of the EEE algorithm, we propose a joint imputation based on the conditional linear property for the multivariate Bernoulli distribution. In addition, we establish asymptotic properties for the proposed estimators and the convergence property using the EEE algorithm. Our method is compared to an existing competitive mixture model approach in both simulation studies and an election data application. Supplementary materials for this article are available online.

Particle Approximations of the Score and Observed Information Matrix for Parameter Estimation in State–Space Models With Linear Computational Cost P. 1138-1157

Christopher Nemeth, Paul Fearnhead & Lyudmila Mihaylova

Abstract

Poyiadjis, Doucet, and Singh showed how particle methods can be used to estimate both the score and the observed information matrix for state–space models. These methods either suffer from a computational cost that is quadratic in the number of particles, or produce estimates whose variance increases quadratically with the amount of data. This article introduces an alternative approach for estimating these terms at a computational cost that is linear in the number of particles. The method is derived using a combination of kernel density estimation, to avoid the particle degeneracy that causes the quadratically increasing variance, and Rao–Blackwellization. Crucially, we show the method is robust to the choice of bandwidth within the kernel density estimation, as it has good asymptotic properties regardless of this choice. Our estimates of the score and observed information matrix can be used within both online and batch procedures for estimating parameters for state–space models. Empirical results show improved parameter estimates compared to existing methods at a significantly reduced computational cost. Supplementary materials including code are available.

Hannes Helgason, Stefanos Kechagias & Vladas Pipiras

Abstract

Circulant matrix embedding is one of the most popular and efficient methods for the exact generation of Gaussian stationary univariate series. Although the idea of circulant matrix embedding has also been used for the generation of Gaussian stationary random fields, there are many practical covariance structures of random fields where classical embedding methods break down. In this work, we propose a novel methodology that adaptively constructs feasible circulant embeddings based on convex optimization with an objective function measuring the distance of the covariance embedding to the targeted covariance structure over the domain of interest. The optimal value of the objective function will be zero if and only if there exists a feasible embedding for the a priori chosen embedding size.

Xiaoyu Liu, Serge Guillas & Ming-Jun Lai

Abstract

Gaussian fields (GFs) are frequently used in spatial statistics for their versatility. The associated computational cost can be a bottleneck, especially in realistic applications. It has been shown that computational efficiency can be gained by doing the computations using Gaussian Markov random fields (GMRFs) as the GFs can be seen as weak solutions to corresponding stochastic partial differential equations (SPDEs) using piecewise linear finite elements. We introduce a new class of representations of GFs with bivariate splines instead of finite elements. This allows an easier implementation of piecewise polynomial representations of various degrees. It leads to GMRFs that can be inferred efficiently and can be easily extended to nonstationary fields. The solutions approximated with higher order bivariate splines converge faster, hence the computational cost can be alleviated. Numerical simulations using both real and simulated data also demonstrate that our framework increases the flexibility and efficiency. Supplementary materials are available online.

Shangbang Rao, Joseph G. Ibrahim, Jian Cheng, Pew-Thian Yap & Hongtu Zhu

Abstract

High angular resolution diffusion imaging (HARDI) has recently been of great interest in mapping the orientation of intravoxel crossing fibers, and such orientation information allows one to infer the connectivity patterns prevalent among different brain regions and possible changes in such connectivity over time for various neurodegenerative and neuropsychiatric diseases. The aim of this article is to propose a penalized multiscale adaptive regression model (PMARM) framework to spatially and adaptively infer the orientation distribution function (ODF) of water diffusion in regions with complex fiber configurations. In PMARM, we reformulate the HARDI imaging reconstruction as a weighted regularized least-square regression (WRLSR) problem. Similarity and distance weights are introduced to account for spatial smoothness of HARDI, while preserving the unknown discontinuities (e.g., edges between white matter and gray matter) of HARDI. The L_1 penalty function is introduced to ensure the sparse solutions of ODFs, while a scaled L_1 weighted estimator is calculated to correct the bias introduced by the L_1 penalty at each voxel. In PMARM, we integrate the multiscale adaptive regression models, the propagation-separation method, and Lasso (least absolute shrinkage and selection operator) to adaptively estimate ODFs across voxels. Experimental results indicate that PMARM can reduce the angle detection errors on fiber crossing area and provide more accurate reconstruction than standard voxel-wise methods. Supplementary materials for this article are available online.

High-Order Composite Likelihood Inference for Max-Stable Distributions and Processes P. 1212-1129

Stefano Castruccio, Raphaël Huser & Marc G. Genton

Abstract

In multivariate or spatial extremes, inference for max-stable processes observed at a large collection of points is a very challenging problem and current approaches typically rely on less expensive composite likelihoods constructed from small subsets of data. In this work, we explore the limits of modern state-of-the-art computational facilities to perform full likelihood inference and to efficiently evaluate high-order composite likelihoods. With extensive simulations, we assess the loss of information of composite likelihood estimators with respect to a full likelihood approach for some widely used multivariate or spatial extreme models, we discuss how to choose composite likelihood truncation to improve the efficiency, and we also provide recommendations for practitioners. This article has supplementary material online.

A Data-Adaptive Principal Component Analysis: Use of Composite Asymmetric Huber Function P. 1230-1247

Yaeji Lim & Hee-Seok Oh

Abstract

This article considers a new type of principal component analysis (PCA) that adaptively reflects the information of data. The ordinary PCA is useful for dimension reduction and identifying important features of multivariate data. However, it uses the second moment of data only, and consequently, it is not efficient for analyzing real observations in the case that these are skewed or asymmetric data. To extend the scope of PCA to non-Gaussian distributed data that cannot be well represented by the second moment, a new approach for PCA is proposed. The core of the methodology is to use a composite asymmetric Huber function defined as a weighted linear combination of modified Huber loss functions, which replaces the conventional square loss function. A practical algorithm to implement the data-adaptive PCA is discussed. Results from numerical studies including simulation study and real data analysis demonstrate the promising empirical properties of the proposed approach. Supplementary materials for this article are available online.

Pair-Copula Bayesian Networks P. 1248-1271

Alexander Bauer & Claudia Czado

Abstract

Pair-copula Bayesian networks (PCBNs) are a novel class of multivariate statistical models, which combine the distributional flexibility of pair-copula constructions (PCCs) with the parsimony of conditional independence models associated with directed acyclic graphs (DAGs). We are first to provide generic algorithms for random sampling and likelihood inference in arbitrary PCBNs as well as for selecting orderings of the parents of the vertices in the underlying graphs. Model selection of the DAG is facilitated using a version of the well-known PC algorithm that is based on a novel test for conditional independence of random variables tailored to the PCC framework. A simulation study shows the PC algorithm's high aptitude for structure estimation in non-Gaussian PCBNs. The proposed methods are finally applied to modeling financial return data. Supplementary materials for this article are available online.

Accelerated Path-Following Iterative Shrinkage Thresholding Algorithm With Application to Semiparametric Graph Estimation P. 1272-1296

Tuo Zhao & Han Liu

Abstract

We propose an accelerated path-following iterative shrinkage thresholding algorithm (APISTA) for solving high-dimensional sparse nonconvex learning problems. The main difference between APISTA and the path-following iterative shrinkage thresholding algorithm (PISTA) is that APISTA exploits an additional coordinate descent subroutine to boost the computational performance. Such a modification, though simple, has profound impact: APISTA not only enjoys the same theoretical guarantee as that of PISTA, that is, APISTA attains a linear rate of convergence to a unique sparse local optimum with good statistical properties, but also significantly outperforms PISTA in empirical benchmarks. As an application, we apply APISTA to solve a family of nonconvex optimization problems motivated by estimating sparse semiparametric graphical models. APISTA allows us to obtain new statistical recovery results that do not exist in the existing literature. Thorough numerical results are provided to back up our theory.

Nonparametric Conditional Density Estimation in a High-Dimensional Regression Setting P. 1297-1316

Rafael Izbicki & Ann B. Lee

Abstract

In some applications (e.g., in cosmology and economics), the regression $\mathbb{E}[Z|\mathbf{x}]$ is not adequate to represent the association between a predictor $\mathbf{x}$ and a response Z because of multi-modality and asymmetry of $f(z|\mathbf{x})$; using the full density instead of a single-point estimate can then lead to less bias in subsequent analysis. As of now, there are no effective ways of estimating $f(z|\mathbf{x})$ when $\mathbf{x}$ represents high-dimensional, complex data. In this article, we propose a new nonparametric estimator of $f(z|\mathbf{x})$ that adapts to sparse (low-dimensional) structure in $\mathbf{x}$. By directly expanding $f(z|\mathbf{x})$ in the eigenfunctions of a kernel-based operator, we avoid tensor products in high dimensions as well as ratios of estimated densities. Our basis functions are orthogonal with respect to the underlying data distribution, allowing fast implementation and tuning of parameters. We derive rates of convergence and show that the method adapts to the intrinsic dimension of the data. We also demonstrate the effectiveness of the series method on images, spectra, and an application to photometric redshift estimation of galaxies. Supplementary materials for this article are available online.