Power-Conditional-Expected Priors: Using g-Priors With Random Imaginary Data for Variable Selection

Dimitris Fouskakis & Ioannis Ntzoufras

Abstract

The Zellner's g-prior and its recent hierarchical extensions are the most popular default prior choices in the Bayesian variable selection context. These prior setups can be expressed as power-priors with fixed set of imaginary data. In this article, we borrow ideas from the power-expected-posterior (PEP) priors to introduce, under the g-prior approach, an extra hierarchical level that accounts for the imaginary data uncertainty. For normal regression variable selection problems, the resulting power-conditional-expected-posterior (PCEP) prior is a conjugate normal-inverse gamma prior that provides a consistent variable selection procedure and gives support to more parsimonious models than the ones supported using the g-prior and the hyper-g prior for finite samples. Detailed illustrations and comparisons of the variable selection procedures using the proposed method, the g-prior, and the hyper-g prior are provided using both simulated and real data examples.

Bayesian Sparse Group Selection

Ray-Bing Chen, Chi-Hsiang Chu, Shinsheng Yuan & Ying Nian Wu

Abstract

This article proposes a Bayesian approach for the sparse group selection problem in the regression model. In this problem, the variables are partitioned into different groups. It is assumed that only a small number of groups are active for explaining the response variable, and it is further assumed that within each active group only a small number of variables are active. We adopt a Bayesian hierarchical formulation, where each candidate group is associated with a binary variable indicating whether the group is active or not. Within each group, each candidate variable is also associated with a binary indicator, too. Thus, the sparse group selection problem can be solved by sampling from the posterior distribution of the two layers of indicator variables. We adopt a group-wise Gibbs sampler for posterior sampling. We demonstrate the proposed method by simulation studies as well as real examples. The simulation results show that the proposed method performs better than the sparse group Lasso in terms of selecting the active groups as well as identifying the active variables within the selected groups.

A Practical Sequential Stopping Rule for High-Dimensional Markov Chain Monte Carlo

Lei Gong & James M. Flegal

Abstract

A current challenge for many Bayesian analyses is determining when to terminate high-dimensional Markov chain Monte Carlo simulations. To this end, we propose using an automated sequential stopping procedure that terminates the simulation when the computational uncertainty is small relative to the posterior uncertainty. Further, we show this stopping rule is equivalent to stopping when the effective sample size is sufficiently large. Such a stopping rule has previously been shown to work well in settings with posteriors of moderate dimension. In this article, we illustrate its utility in high-dimensional

simulations while overcoming some current computational issues. As examples, we consider two complex Bayesian analyses on spatially and temporally correlated datasets. The first involves a dynamic space-time model on weather station data and the second a spatial variable selection model on fMRI brain imaging data. Our results show the sequential stopping rule is easy to implement, provides uncertainty estimates, and performs well in high-dimensional settings.

Toward Automatic Model Comparison: An Adaptive Sequential Monte Carlo Approach

Yan Zhou, Adam M. Johansen & John A.D. Aston

Abstract

Model comparison for the purposes of selection, averaging, and validation is a problem found throughout statistics. Within the Bayesian paradigm, these problems all require the calculation of the posterior probabilities of models within a particular class. Substantial progress has been made in recent years, but difficulties remain in the implementation of existing schemes. This article presents adaptive sequential Monte Carlo (SMC) sampling strategies to characterize the posterior distribution of a collection of models, as well as the parameters of those models. Both a simple product estimator and a combination of SMC and a path sampling estimator are considered and existing theoretical results are extended to include the path sampling variant. A novel approach to the automatic specification of distributions within SMC algorithms is presented and shown to outperform the state of the art in this area. The performance of the proposed strategies is demonstrated via an extensive empirical study. Comparisons with state-of-the-art algorithms show that the proposed algorithms are always competitive, and often substantially superior to alternative techniques, at equal computational cost and considerably less application-specific implementation effort.

Zero Expectile Processes and Bayesian Spatial Regression

Anandamayee Majumdar & Debashis Paul

Abstract

We introduce new classes of stationary spatial processes with asymmetric, sub-Gaussian marginal distributions using the idea of expectiles. We derive theoretical properties of the proposed processes. Moreover, we use the proposed spatial processes to formulate a spatial regression model for point-referenced data where the spatially correlated errors have skewed marginal distribution. We introduce a Bayesian computational procedure for model fitting and inference for this class of spatial regression models. We compare the performance of the proposed method with the traditional Gaussian process-based spatial regression through simulation studies and by applying it to a dataset on air pollution in California.

Bayesian Ensemble Trees (BET) for Clustering and Prediction in Heterogeneous Data

Leo L. Duan, John P. Clancy & Rhonda D. Szczesniak

Abstract

We propose a novel "tree-averaging" model that uses the ensemble of classification and regression trees (CART). Each constituent tree is estimated with a subset of similar data. We treat this grouping of subsets as Bayesian ensemble trees (BET) and model them as a Dirichlet process. We show that BET determines the optimal number of trees by adapting to the data heterogeneity. Compared with the other ensemble methods, BET requires much fewer trees and shows equivalent prediction accuracy using weighted averaging. Moreover, each tree in BET provides variable selection criterion and interpretation for each subset. We developed an efficient estimating procedure with improved estimation strategies in both CART and mixture models. We demonstrate these advantages of BET with simulations and illustrate the approach with a real-world data example involving regression of lung function measurements obtained from patients with cystic fibrosis.

GPU-Powered Shotgun Stochastic Search for Dirichlet Process Mixtures of Gaussian Graphical Models

Chiranjit Mukherjee & Abel Rodriguez

Abstract

Gaussian graphical models (GGMs) are popular for modeling high-dimensional multivariate data with sparse conditional dependencies. A mixture of GGMs extends this model to the more realistic scenario where observations come from a heterogenous population composed of a small number of homogeneous subgroups. In this article, we present a novel stochastic search algorithm for finding the posterior mode of high-dimensional Dirichlet process mixtures of decomposable GGMs. Further, we investigate how to harness the massive thread-parallelization capabilities of graphical processing units to accelerate computation. The computational advantages of our algorithms are demonstrated with various simulated data examples in which we compare our stochastic search with a Markov chain Monte Carlo (MCMC) algorithm in moderate dimensional data examples. These experiments show that our stochastic search largely outperforms the MCMC algorithm in terms of computing-times and in terms of the quality of the posterior mode discovered. Finally, we analyze a gene expression dataset in which MCMC algorithms are too slow to be practically useful.

Parallel Resampling in the Particle Filter

Lawrence M. Murray, Anthony Lee & Pierre E. Jacob

Abstract

Modern parallel computing devices, such as the graphics processing unit (GPU), have gained significant traction in scientific and statistical computing. They are particularly well-suited to data-parallel algorithms such as the particle filter, or more generally sequential Monte Carlo (SMC), which are increasingly used in statistical inference. SMC methods carry a set of weighted *particles* through repeated propagation, weighting, and resampling steps. The propagation and weighting steps are straightforward to parallelize, as they require only independent operations on each particle. The resampling step is more difficult, as standard schemes require a collective operation, such as a sum, across particle weights. Focusing on this resampling step, we analyze two alternative schemes that do not involve a collective operation (Metropolis and rejection resamplers), and compare them to standard schemes (multinomial, stratified, and systematic resamplers). We find that, in certain circumstances, the alternative resamplers can perform significantly faster on a GPU, and to a lesser extent on a CPU, than the standard approaches. Moreover, in single precision, the standard approaches are numerically biased for upward of hundreds of thousands of particles, while the alternatives are not. This is particularly important given greater single- than double-precision throughput on modern devices, and the consequent temptation to use single precision with a greater number of particles. Finally, we provide auxiliary functions useful for implementation, such as for the permutation of ancestry vectors to enable in-place propagation.

Reinforced Angle-Based Multicategory Support Vector Machines

Chong Zhang, Yufeng Liu, Junhui Wang & Hongtu Zhu

Abstract

The support vector machine (SVM) is a very popular classification tool with many successful applications. It was originally designed for binary problems with desirable theoretical properties. Although there exist various multicategory SVM (MSVM) extensions in the literature, some challenges remain. In particular, most existing MSVMs make use of k classification functions for a k-class problem, and the corresponding optimization problems are typically handled by existing quadratic programming solvers. In this article, we propose a new group of MSVMs, namely, the reinforced angle-based MSVMs (RAMSVMs), using an angle-based prediction rule with $k - 1$ functions directly. We prove that RAMSVMs can enjoy Fisher consistency. Moreover, we show that the RAMSVM can be implemented using the very efficient coordinate descent algorithm on its dual problem. Numerical experiments demonstrate that our method is highly competitive in terms of computational speed, as well as classification prediction performance.

Sparse Distance Weighted Discrimination

Boxiang Wang & Hui Zou

Abstract

Distance weighted discrimination (DWD) was originally proposed to handle the data piling issue in the support vector machine. In this article, we consider the sparse penalized DWD for high-dimensional classification. The state-of-the-art algorithm for solving the standard DWD is based on second-order cone programming, however such an algorithm does not work well for the sparse penalized DWD with high-dimensional data. To overcome the challenging computation difficulty, we develop a very efficient algorithm to compute the solution path of the sparse DWD at a given fine grid of regularization parameters. We implement the algorithm in a publicly available R package sdwd. We conduct extensive numerical experiments to demonstrate the computational efficiency and classification performance of our method.

Fast and Flexible ADMM Algorithms for Trend Filtering

Aaditya Ramdas & Ryan J. Tibshirani

Abstract

This article presents a fast and robust algorithm for trend filtering, a recently developed nonparametric regression tool. It has been shown that, for estimating functions whose derivatives are of bounded variation, trend filtering achieves the minimax optimal error rate, while other popular methods like smoothing splines and kernels do not. Standing in the way of a more widespread practical adoption, however, is a lack of scalable and numerically stable algorithms for fitting trend filtering estimates. This article presents a highly efficient, specialized alternating direction method of multipliers (ADMM) routine for trend filtering. Our algorithm is competitive with the specialized interior point methods that are currently in use, and yet is far more numerically robust. Furthermore, the proposed ADMM implementation is very simple, and, importantly, it is flexible enough to extend to many interesting related problems, such as sparse trend filtering and isotonic trend filtering. Software for our method is freely available, in both the C and R languages.

Supervised Sparse and Functional Principal Component Analysis

Gen Li, Haipeng Shen & Jianhua Z. Huang

Abstract

Principal component analysis (PCA) is an important tool for dimension reduction in multivariate analysis. Regularized PCA methods, such as sparse PCA and functional PCA, have been developed to incorporate special features in many real applications. Sometimes additional variables (referred to as supervision) are measured on the same set of samples, which can potentially drive low-rank structures of the primary data of interest. Classical PCA methods cannot make use of such supervision data. In this article, we propose a supervised sparse and functional principal component (SupSFPC) framework that can incorporate supervision information to recover underlying structures that are more interpretable. The framework unifies and generalizes several existing methods and flexibly adapts to the practical scenarios at hand. The SupSFPC model is formulated in a hierarchical fashion using latent variables. We develop an efficient modified expectation-maximization (EM) algorithm for parameter estimation. We also implement fast data-driven procedures for tuning parameter selection. Our comprehensive simulation and real data examples demonstrate the advantages of SupSFPC.

SHAH: SHape-Adaptive Haar Wavelets for Image Processing

Piotr Fryzlewicz & Catherine Timmermans

Abstract

We propose the shape-adaptive Haar (SHAH) transform for images, which results in an orthonormal, adaptive decomposition of the image into Haar-wavelet-like components, arranged hierarchically according to decreasing importance, whose shapes reflect the features present in the image. The decomposition is as sparse as it can be for piecewise-constant images. It is performed via a stepwise bottom-up algorithm with quadratic computational complexity; however, nearly linear variants also exist. SHAH is rapidly invertible. We show how to use SHAH for image

denoising. Having performed the SHAH transform, the coefficients are hard- or soft-thresholded, and the inverse transform taken. The SHAH image denoising algorithm compares favorably to the state of the art for piecewise-constant images. A clear asset of the methodology is its very general scope: it can be used with any images or more generally with any data that can be represented as graphs or networks.

Fast Nonparametric Density-Based Clustering of Large Datasets Using a Stochastic Approximation Mean-Shift Algorithm

Ollivier Hyrien & Andrea Baran

Abstract

Mean-shift is an iterative procedure often used as a nonparametric clustering algorithm that defines clusters based on the modal regions of a density function. The algorithm is conceptually appealing and makes assumptions neither about the shape of the clusters nor about their number. However, with a complexity of $O(n^2)$ per iteration, it does not scale well to large datasets. We propose a novel algorithm which performs density-based clustering much quicker than mean shift, yet delivering virtually identical results. This algorithm combines subsampling and a stochastic approximation procedure to achieve a potential complexity of $O(n)$ at each step. Its convergence is established. Its performances are evaluated using simulations and applications to image segmentation, where the algorithm was tens or hundreds of times faster than mean shift, yet causing negligible amounts of clustering errors. The algorithm can be combined with existing approaches to further accelerate clustering.

Testing for Hermite Rank in Gaussian Subordination Processes

Jan Beran, Sven Möhrle & Sucharita Ghosh

Abstract

Statistical inference for time series with long-range dependence is often based on the assumption of Gaussian subordination $X_t = G(Z_t)$. Although the Hermite rank m of G plays an essential role for statistical inference in these situations, the question of estimating m or of testing hypotheses about the Hermite rank has not been addressed in the literature. In this article, a method is introduced for testing H_0: $m = 1$ against H_1: $m > 1$. This allows for deciding whether inference based on the usual assumption of $m = 1$ is appropriate. Simulations and data examples illustrate the method.

Parsimonious and Efficient Likelihood Composition by Gibbs Sampling

Davide Ferrari, Guoqi Qian & Tane Hunter

Abstract

The traditional maximum likelihood estimator (MLE) is often of limited use in complex high-dimensional data due to the intractability of the underlying likelihood function. Maximum composite likelihood estimation (McLE) avoids full likelihood specification by combining a number of partial likelihood objects depending on small data subsets, thus enabling inference for complex data. A fundamental difficulty in making the McLE approach practicable is the selection from numerous candidate likelihood objects for constructing the composite likelihood function. In this article, we propose a flexible Gibbs sampling scheme for optimal selection of sub-likelihood components. The sampled composite likelihood functions are shown to converge to the one maximally informative on the unknown parameters in equilibrium, since sub-likelihood objects are chosen with probability depending on the variance of the corresponding McLE. A penalized version of our method generates sparse likelihoods with a relatively small number of components when the data complexity is intense. Our algorithms are illustrated through numerical examples on simulated data as well as real genotype single nucleotide polymorphism (SNP) data from a case–control study.

RAPTT: An Exact Two-Sample Test in High Dimensions Using Random Projections

Radhendushka Srivastava, Ping Li & David Ruppert

Abstract

In high dimensions, the classical Hotelling's T^2 test tends to have low power or becomes undefined due to singularity of the sample covariance matrix. In this article, this problem is overcome by projecting the data matrix onto lower dimensional subspaces through multiplication by random matrices. We propose RAPTT (RAndom Projection T^2-Test), an exact test for equality of means of two normal populations based on projected lower dimensional data. RAPTT does not require any constraints on the dimension of the data or the sample size. A simulation study indicates that in high dimensions the power of this test is often greater than that of competing tests. The advantages of RAPTT are illustrated on a high-dimensional gene expression dataset involving the discrimination of tumor and normal colon tissues.

Displaying Variation in Large Datasets: Plotting a Visual Summary of Effect Sizes

Gregory B. Gloor, Jean M. Macklaim & Andrew D. Fernandes

Abstract

Displaying the component-wise between-group differences high-dimensional datasets is problematic because widely used plots such as Bland–Altman and Volcano plots do not show what they are colloquially *believed* to show. Thus, it is difficult for the experimentalist to grasp why the between-group difference of one component is "significant" while that of another component is not. Here, we propose a type of "Effect Plot" that displays between-group differences in relation to respective underlying variability for every component of a high-dimensional dataset. We use synthetic data to show that such a plot captures the essence of what determines "significance" for between-group differences in each component, and provide guidance in the interpretation of the plot. Supplementary online materials contain the code and data for this article and include simple R functions to produce an effect plot from suitable datasets.