Workshop on Sparsity in Applied Mathematics and Statistics Brussels, June 1-2, 2017

SAMS 2017, Workshop program

Thursday, June 1, 2017

8h30–9h00	Registration	
9h00–10h00	Göran Kauermann	Statistical Models for Network Data Analysis – A Gentle Introduction
10h00–11h00	Laure Blanc-Féraud	Continuous approximation of sparse least-square (l2-l0) criterion : application to super-resolution microscopy
Coffee break		
11h30–12h30	Francesco Stingo	Bayesian approaches for array-variate graphical models
Lunch break		
14h00–14h30	Dominik Mueller	Selection of Sparse Vine Copulas in High Dimensions with the Lasso
14h30–15h00	Vahe Avagyan	D-trace Estimation of a Precision Matrix Using Adaptive Lasso Penalties
15h00–15h30	Sander Devriendt	A computational framework for sparse modeling with different types of predictors
15h30–16h00	Eugen Pircalabelu	Estimating multiple directed and undirected graphs using zoom in/out penalties
Coffee break		
16h30–17h00	M. Remedios Sillero Denamiel	A proposal of sparse Naïve Bayes
17h00–17h30	Sandra Benítez-Peña	Cost-sensitive feature selection in Support Vector Machines
17h30–18h00	Robin Bruyndonckx	Simulation-based evaluation of the impact of singletons on the fixed effects in a linear mixed model

Friday, June 2, 2017

8h30–9h00	Registration	
9h00–10h00	Richard Samworth	
10h00–11h00	Ivan Markovsky	Sparsity in system identification and data-driven control
Coffee break		
11h30–12h00	Nicolas Gillis	A Fast Gradient Method for Nonnegative Sparse Regression with Self Dictionary
12h00–12h30	Jixin Chen	On the convergence of nested primal-dual iterations
12h30–13h00	Meabh McCurdy	Investigating Krylov subspace methods for the calculation of sparse matrix exponentials
Lunch break		
14h00–14h30	Pepa Ramirez-Cobo	Enhancing interpretability by tightening linear regression methods
14h30–15h00	Tomas Masak	An Iteratively Reweighted Least Squares Algorithm for Sparse Principal Component Analysis with an Application to Voting Records
15h00–15h30	Stéphanie Aerts	Cellwise robust regularized discriminant analysis

SAMS 2017, Abstracts of talks

Continuous approximation of sparse least-square (l2-l0) criterion : application to super-resolution microscopy

LAURE BLANC-FÉRAUD

CNRS, I3S lab, Sophia Antipolis, France

Joint work with Emmanuel Soubies and Gilles Aubert

In the first part of the talk, theoretical results on continuous approximations of sparse least-square (l2-l0) criterion are described. We exhibit necessary and sufficient conditions for smooth separable approximations of l2-l0 problem such that the approximation has the same global minimizers and such that local minimizers of the approximation are also minimizers of the original l2-l0 criterion. Numerical simulations, using several relaxed functionals proposed in the literature, reveal that the inferior limit of the proposed class of continuous approximations is the one which removes the largest number of minimizers of l2-l0. Then results will be shown on a single molecule super-resolution microscopy problem, in the dense molecule case using data of a challenge of the IEEE ISBI conference.

Bayesian approaches for array-variate graphical models

FRANCESCO STINGO

Università degli Studi di Firenze

Multi-dimensional data constituted by measurements along multiple axes have emerged across many scientific areas such as genomics and cancer surveillance. Traditional multivariate approaches are unsuitable for such highly structured data due to inefficiency, loss of power and lack of interpretability. I will illustrate a novel class of multi-dimensional graphical models that includes both directed and undirected graphs as well as arbitrary combinations of these.

Statistical Models for Network Data Analysis – A Gentle Introduction

GÖRAN KAUERMANN

Ludwig-Maximilians-Universität München

We give an introduction to statistical models for network data. Starting from 'simple' models we concentrate on Exponential Random Graph Models (ERGM) as a common tool for network data analyses. The ERGM describes the distribution of a network graph with an exponential family distribution, where the statistics are counts of edges, k-stars or triangles, for instance. This allows for meaningful interpretations of network data, as demonstrated in the talk.

Though the model class mirrors welcome properties of exponential families, the fitting of ERGMs is numerical a burden due to a non-feasible normalization constant in the exponential model. We discuss the state of the art of estimation routines and extend classical models towards stable models based on non-linear smoothly fitted statistics.

Two data example accompany the presentation. First, we investigate the network of international arms trade, second, we look at facebook data and explore friendship networks.

Sparsity in system identification and data-driven control

IVAN MARKOVSKY

VUB, Brussels

Sparsity is a key underlying assumption in many applications. A challenge in using sparse approximation methods in new applications is finding bases that allow sparse representation of the data. This talk shows how sparsity appears naturally in problems in systems identification. The underlying assumption then is that the data is generated by a low complexity model. The system identification problem is then a structured low-rank approximation problem. The link between sparsity and model complexity allows us to use effective methods from the field of compressive sensing in systems and control. A specific example considered in this talk is data-driven control, i.e., obtaining a control signal directly from observed data without explicit model identification.

Cellwise robust regularized discriminant analysis

STÉPHANIE AERTS

Université de Liège, Belgium

Quadratic and Linear Discriminant Analysis (QDA/LDA) are the most often applied classification rules under normality. In QDA, a separate covariance matrix is estimated for each group. If there are more variables than observations in the groups, the usual estimates are singular and cannot be used anymore. Assuming homoscedasticity, as in LDA, reduces the number of parameters to estimate. This rather strong assumption is however rarely verified in practice. Regularized discriminant techniques that are computable in high-dimension and cover the path between the two extremes QDA and LDA have been proposed in the literature. However, these procedures rely on sample covariance matrices. As such, they become inappropriate in presence of cellwise outliers, a type of outliers that is very likely to occur in high-dimensional datasets. We propose cellwise robust counterparts of these regularized discriminant techniques by inserting cellwise robust covariance matrices. Our methodology results in a family of discriminant methods that are robust against outlying cells, cover the gap between LDA and QDA and are computable in high-dimension.

D-trace Estimation of a Precision Matrix Using Adaptive Lasso Penalties

VAHE AVAGYAN

UGent, Belgium

An accurate estimation of a precision matrix has a crucial role in the current age of high-dimensional data explosion. To deal with this problem, one of the prominent and commonly used techniques is the ℓ_1 norm (Lasso) penalization for a given loss function. This approach guarantees the sparsity of the precision matrix estimator for properly selected penalty parameters. However, the ℓ_1 norm penalization often fails to control the bias of the obtained estimator because of its overestimation behavior. In this paper, we introduce two adaptive extensions of the recently proposed ℓ_1 norm penalized D-trace loss minimization method. The proposed approaches intend to diminish the produced bias in the estimator. Extensive numerical results, using both simulated and real datasets, show the advantage of our proposed estimators.

Cost-sensitive feature selection in Support Vector Machines

SANDRA BENÍTEZ-PEÑA

University of Seville, Spain

Feature Selection (FS) is a crucial procedure in Data Science tasks such as Classification, since it identifies the relevant variables, making thus the classification procedures more interpretable and more effective by reducing noise and data overfit. The relevance of features in a classification procedure is linked to the fact that misclassifications costs are frequently asymmetric, since false positive and false negative cases may have very different consequences. However, off-the-shelf FS procedures seldom take into account such cost-sensitivity of errors.

In this paper we propose an FS procedure embedded in one of the most popular classification procedures, namely, Support Vector Machines (SVM), accommodating asymmetric misclassification costs. The key idea is to replace the traditional margin maximization by minimizing the number of features selected, but imposing upper bounds on the false positive and negative rates on an independent test set. The problem is written as a mixed integer linear problem for SVM with a linear kernel. Moreover, an alternating procedure is proposed for SVM with arbitrary kernels.

The numerical experience reported demonstrates the usefulness of the FS procedure proposed. Indeed, our results on benchmark data sets show that a substantial decrease of the number of features is obtained, whilst the desired trade-off between false positive and false negative rates is achieved.

Simulation-based evaluation of the impact of singletons on the fixed effects in a linear mixed model

ROBIN BRUYNDONCKX

Universiteit Hasselt - Censtat & Universiteit Antwerpen - Labo Medische Microbiologie, Belgium

Data that are collected in medical sciences often have a hierarchical structure. Regardless of sparseness caused by the presence of a large number of small units or a small number of large units, linear mixed models are used to account for within-unit correlation. Using a simulation study, we assess the impact of an increasing proportion of singletons (i.e. one subunit in a unit) at the highest or lowest level of the hierarchy on the fixed effects in a linear mixed model.

Additionally, we assess whether, when high proportions of singletons are present, performance improves by removing or grouping singletons, splitting singletons at the highest level or ignoring the dependency within units at the lowest level. We show that, in the presence of singletons at the lowest level, the model is quite stable. Ignoring clustering and dropping the singletons come with biased standard error estimates. Grouping the singletons does not improve the models performance. In the presence of singletons at the highest level, the model is unstable. Grouping, dropping and splitting the singletons either decrease type I error rate or increase power, while worsening the other. The likelihood ratio test and Wald test perform poorly. The performance of the permutation test however is superior to that of the F test. In conclusion, the linear mixed model is stable in the presence of singletons at the lowest level, but care should be taken in the presence of singletons at the highest level. In that case, the permutation test rather than the F test should be used to assess the significance of included fixed effects.

On the convergence of nested primal-dual iterations

JIXIN CHEN

Universit libre de Bruxelles, Belgium

We show convergence of an iterative optimization algorithm consisting of nested primal-dual proximal-gradient iterations. The algorithm can be used for the numerical solution of convex optimization problems defined by the sum of a differentiable term and a possibly non-differentiable term. The latter should take the form of the composition of a linear map and a proximable function, while the former needs an accessible gradient. The algorithm reduces to the usual proximal gradient algorithm in certain special cases and generalizes another existing algorithm.

A Fast Gradient Method for Nonnegative Sparse Regression with Self Dictionary

NICOLAS GILLIS

University of Mons, Belgium

Nonnegative matrix factorization (NMF) can be computed efficiently under the separability assumption, which asserts that all the columns of the input data matrix belong to the convex cone generated by only a few of its columns. The provably most robust methods to identify these basis columns are based on nonnegative sparse regression and self dictionary, and require the solution of large-scale convex optimization problems. In this talk, we present a particular nonnegative sparse regression model with self dictionary. As opposed to previously proposed models, it is a smooth optimization problem where sparsity is enforced through appropriate linear constraints. We show that the Euclidean projection on the set defined by these constraints can be computed efficiently, and propose a fast gradient method to solve our model. We show the effectiveness of the approach compared to state-of-the-art methods on several synthetic data sets and real-world hyperspectral images. This is joint work with Robert Luce (EPFL).

A computational framework for sparse modeling with different types of predictors

SANDER DEVRIENDT

KU Leuven, Belgium

Insurance companies use predictive models for a variety of analytic tasks, including pricing, marketing campaigns, claims handling, fraud detection and reserving. Typically, these predictive models use a selection of continuous, ordinal, nominal and spatial variables to differentiate risks. Such models should not only be competitive, but also interpretable by stakeholders (including the policyholder and the regulator) and easy to implement and maintain in a production environment. That is why current actuarial literature puts focus on generalized linear models where risk cells are constructed by binning variables up front, using ad hoc techniques or professional expertise. In statistical literature penalized regression is often used to encourage the selection and fusion of variables in predictive modeling. Most penalization strategies work for data where predictors are of the same type, such as LASSO for continuous variables and Fused LASSO for ordered variables. We design an estimation strategy for generalized linear models which includes variable selection and the binning of parameters through L1-type penalties. We consider the joint presence of different types of covariates and a specific penalty for each type of predictor. Using the theory of proximal operators, our estimation procedure is computationally efficient since it splits the overall optimization problem into easier to solve sub-problems per predictor and its associated penalty. As such, we are able to simultaneously select, estimate and group, in a statistically sound way, any combination of continuous, ordinal, nominal and spatial variables. We illustrate the approach with simulation studies, an analysis of Munich rent data, and a case-study on motor insurance pricing.

An Iteratively Reweighted Least Squares Algorithm for Sparse Principal Component Analysis with an Application to Voting Records

TOMAS MASAK

Charles University in Prague, Czechia

Principal component analysis (PCA) is a popular dimensionality reduction and data visualization method, and sparse PCA (SPCA) is its extensively studied and NP-hard-to-solve modification. In the past decade, many different algorithms were proposed to perform SPCA. We build upon the work of Zou et al. (2006) who recast the SPCA problem into the regression framework and proposed to induce sparsity with the ℓ_1 penalty. We propose to drop the ℓ_1 penalty, and instead promote sparsity by re-weighting the ℓ_2-norm. Our algorithm thus consists mainly of solving weighted ridge regression problems. We show that the algorithm basically attempts to find a solution to a penalized least squares problem with a non-convex penalty that more closely resembles the ℓ_0-norm. We also apply the algorithm to analyze the voting records of the Chamber of Deputies of the Parliament of the Czech Republic. We show not only why the SPCA is more appropriate to analyze this type of data, but we also discuss whether the variable selection property can be utilized as an additional piece of information, for example to create voting calculators automatically.

Investigating Krylov subspace methods for the calculation of sparse matrix exponentials

MEABH MCCURDY

Queen's University Belfast

The most widely used and studied matrix function is arguably the calculation of the exponential of a matrix. The interest in this matrix exponential stems from the significant role it plays in finding solutions to differential equations in many different application areas. However one of the major drawbacks with this matrix operation is the computational run time. This is particularly a problem when implementing the standard algorithms such as scaling and squaring and the Pade approximation on large sparse matrices. Krylov subspace methods can be seen as the latest and most efficient approach to calculating these sparse matrix exponentials. Krylov subspace methods project a matrix of dimension n onto a small Krylov subspace of dimension m, where m¡n. The main advantage for using these methods are that they replace matrix-matrix operations with matrix-vector operations, with the aim of improving the issue of computational time. The expokit is a type of software that implements Krylov subspace methods. This software, initially in a Fortran and Matlab environment, can now be used within R using the package Rexpokit. The Rexpokit has issues with both accuracy and efficiency. This research aims to improve these issues through the use of a newly proposed R package, KEXPMV, which utilises Krylov subspace methods. These improvements will be demonstrated through the use of large sparse matrices.

Selection of Sparse Vine Copulas in High Dimensions with the Lasso

DOMINIK MUELLER

Technical University of Munich

We propose a novel structure selection method for high dimensional (d ¿ 100) sparse vine copulas. Current sequential greedy approaches for structure selection require calculating spanning trees in hundreds of dimensions and fitting the pair copulas and their parameters iteratively throughout the structure selection process. Our method uses a connection between the vine and structural equation models (SEMs). The later can be estimated very fast using the Lasso, also in very high dimensions, to obtain sparse models. Thus, we obtain a structure estimate independently of the chosen pair copulas and parameters. Additionally, we define the novel concept of regularization paths for R-vine matrices. It relates sparsity of the vine copula model in terms of independence copulas to a penalization coefficient in the structural equation models. We illustrate our approach and provide many numerical examples. These include simulations and data applications in high dimensions, showing the superiority of our approach to other existing methods.

Estimating multiple directed and undirected graphs using zoom in/out penalties

EUGEN PIRCALABELU

KU Leuven, ORSTAT and Leuven Statistics Research Center

We propose a method for estimating brain networks from fMRI datasets that do not all contain measurements on the same set of regions. For certain datasets, some of the regions have been split in smaller subregions, while others have not been split. This gives rise to the framework of mixed scale measurements. The purpose is to estimate sparse directed and undirected graphical models from data obtained at K different coarseness scales.

Starting from a predefined scale $k^* \leq K$ the method zooms in or out over scales on particular edges, which results in the joint estimation of graphs with similar structures, but different levels of sparsity. We can evaluate the evolution of the graphs from the coarsest to the finest scale or vice-versa. We accomplish this by pooling information from all subjects in order to estimate a common undirected and directed graph at each coarseness scale and by accounting for time dependencies and multiple coarseness scales. The applicability of the method goes beyond fMRI data, to other areas where data on different scales are observed and where the joint estimation of graphs that resemble each other is desired. We select an optimal coarseness scale to be used for further analyses.

Empirical and theoretical evaluations illustrate the usefulness of the method.

Enhancing interpretability by tightening linear regression methods

PEPA RAMIREZ-COBO

University of Cadiz, Spain

One of the main problems researchers face is to understand the main interactions between the features conforming a pile of data and the variable to be predicted. In this work we propose to use tools based on Mathematical Optimization to enhance interpretability in linear regression. In particular, we aim to: (i) improve the sparsity of the solutions, (ii) allow for only significant features to be represented in the model, and (iii) force the sign of the coefficients to be consistent with the sign of the correlations between predictors. This will be addressed by modelling constraints and integrating them into the estimation procedure. In particular, the so-obtained tightened sparse models will become Mixed Integer Non-Linear Problems (MINLP henceforth), which have recently become a tractable tool to address statistical problems. The numerical experiments carried out on real and simulated datasets support this statement and also show that tightening the search space of some standard linear regression models may enhance the interpretability of the outcomes with competitive predictive quality.

A proposal of sparse Naïve Bayes

M. REMEDIOS SILLERO DENAMIEL

University of Sevilla, Spain

The Naïve Bayes has proven a tractable and efficient method for classification in multivariate analysis. However, as it is common in real contexts, datasets are often characterized by a large number of features, which may complicate the interpretation of the results as well as slow down the methods execution. In addition, features are usually correlated, a fact that violates the Nave Bayes assumption of conditional independence and may deteriorate the methods performance. In this paper we propose a sparse version of the Naïve Bayes in which a variable reduction approach is embedded into the classification algorithm. Unlike typical variable selection methods in classification, which are based on the correlation between the features with the response variable, our strategy for reducing the number of predictors takes into account the dependencies among features. Our findings show that, under a reasonable computational cost, not only the number of variables is significantly reduced, but also the performance measures are comparable or even better than those obtained under the classic version of the classifier.

A Perfect Reconstruction Property for PDE-constrained total variation minimization

DAVID VICENTE

University of Graz, Austria

In this talk we study the recovery of piecewise constant functions of finite bounded variation which are constrained to satisfy a linear partial differential equation with unknown boundary conditions. We prove an exact recovery result up to a global constant under a mild geometric assumption on the jump set of the function to reconstruct. For that, we establish some theoretical results about BV spaces which consist in the generalization of the notion of wavefront in this setting. To demonstrate the practical relevance of these results, we also discuss the application to quantitative susceptibility mapping (QSM), a recently established magnetic resonance imaging (MRI) technique which aims at providing the spatial susceptibility distribution of tissues inside the human body. This is a joint work with Kristian Bredies (University of Graz, Austria).

SAMS 2017, List of participants

Name		Affiliation
Stéphanie	Aerts	Université de Liège, Belgium
Mohamed	Amghar	ULB, Brussels, Belgium
Man Shun	Ang	University of Mons, Belgium
Vahe	Avagyan	UGent, Belgium
Sandra	Benítez-Peña	University of Seville, Spain
Laure	Blanc-Feraud	Université de Nice-Sophia Antipolis, France
Robin	Bruyndonckx	UHasselt & Universiteit Antwerpen, Belgium
Andrea	Callegaro	GSK vaccines
Robrecht	Cannoodt	UGent, Belgium
Valerio	Cambareri	UC Louvain, Belgium
Ali	Charkhi	KU Leuven, Belgium
Jixin	Chen	ULB, Brussels, Belgium
Gerda	Claeskens	KU Leuven, Belgium
Véronique	Delouille	STCE/Royal Observatory of Belgium
Sander	Devriendt	KU Leuven, Belgium
David	D'Haese	Johnson & Johnson
Shuyu	Dong	UC Louvain, Belgium
Shahla	Faisal	Ludwig-Maximilians-Universität München, Germany
Baptiste	Feraud	UC Louvain, Belgium
Nicolas	Gillis	University of Mons, Belgium
Irène	Gijbels	KU Leuven, Belgium
Bernadette	Govaerts	UC Louvain, Belgium
Pierre	Gramme	DNAlytics
Thomas	Gueuning	KU Leuven, Belgium
Hilmar	Gudmundsson	UGent, Belgium
Leonardo	Gutierrez	UC Louvain, Belgium
Daan	Huybrechs	KU Leuven, Belgium
Mohammed Abdulkerim	Ibrahim	UHasselt, Belgium
Laurent	Jacques	UC Louvain, Belgium
Maarten	Jansen	ULB, Brussels, Belgium
Goeran	Kauermann	Ludwig-Maximilians-Universität München, Germany
Renaud	Lambiotte	University of Namur, Belgium
Christophe	Ley	UGent, Belgium
Liran	Lerman	ULB, Brussels, Belgium
Ignace	Loris	ULB, Brussels, Belgium
Glenn	Magerman	ULB, Brussels, Belgium
Rebecca	Marion	UC Louvain, Belgium
Ivan	Markovsky	VUB, Brussels, Belgium

Bastien	Marquis	ULB, Brussels, Belgium
Tomas	Masak	Charles University, Prague, Czechia
Meabh	McCurdy	Queen's University Belfast, UK
Amirafshar	Moshtaghpour	UC Louvain, Belgium
Ilham	Moudou	
Dominik	Mueller	Technical University of Munich, Germany
Ulrike	Naumann	Janssen, Belgium
Jérôme	Paul	DNAlytics
Leto	Peel	UC Louvain, Belgium
Eugen	Pircalabelu	KU Leuven, Belgium
Dian M	Pratiwi	U. of Essex, UK
Gustavo	Quintana Carapia	VUB, Brussels, Belgium
Pepa	Ramirez-Cobo	University of Cadiz, Spain
Emilie	Renard	UC Louvain, Belgium
M. Remedios	Sillero	Denamiel University of Sevilla, Spain
Richard	Samworth	Cambridge University, UK
Punit	Sharma	University of Mons, Belgium
Francesco	Stingo	Università degli Studi di Firenze, Italy
Arnaud	Vandaele	University of Mons, Belgium
Edward	van der Meulen	KU Leuven, Belgium
Stijn	Vansteelandt	UGent, Belgium
David	Vicente	University of Graz, Austria
Yue	Zhao	KU Leuven, Belgium
Jing	Zhou	KU Leuven, Belgium