MATRIX Research Program: Computational Mathematics for High-Dimensional Data in Statistical Learning

Feb 6–17, 2023, Creswick

ABSTRACT BOOK

Week 1 Presentations

Monday, Feb 6, 2023

11:00-12:00

Title: Speeding up the lasso feature selection method in high dimensions
Author: Zdravko Botev, UNSW Sydney

Abstract: The lasso estimator is a popular feature selection approach in statistics. In this talk we describe a novel algorithm for computing the lasso estimator in high-dimensional situations where the signal is sparse. The algorithm uses a set of auxiliary variables and orthogonal transformations to achieve a fast local convergence. We make comparisons of the novel algorithm against current methodology and consider extensions of the method to feature selection methods beyond the lasso.

12:00-12:30

Title: Some ideas on Cox models with image predictors
Author: Jun Ma, Macquarie University

Abstract: Recent research activities on deep learning Cox models focused on survival times with only right censoring, and thus only the partial likelihood method was employed to develop neuron network algorithms. In this talk, we will discuss some of our thoughts on extending deep learning Cox models to more general and realistic partly interval censoring, which means the observed survival times can include event times or left, right and interval censoring times. The objective function we wish to maximize is a penalized log-likelihood. This is a challenge problem as both the regression coefficients and the non-parametric baseline hazard function need to estimated.

14:00 –15:00

Title: Mathematical theory of structured deep neural networks
Author: Dingxuan Zhou, The University of Sydney

Abstract: Deep learning has been widely applied and brought breakthroughs in speech recognition, computer vision, natural language processing, and many other domains. The involved deep neural network architectures and computational issues have been well studied in machine learning. But there lacks a theoretical foundation for understanding the modelling, approximation or generalization ability of deep learning models with network architectures. An important family of structured deep neural networks is deep convolutional neural networks (CNNs) with convolutional structures. The convolutional architecture gives essential differences between deep CNNs and fully-connected neural networks, and the classical approximation theory for fully-connected networks developed around 30 years ago does not apply. This talk describes approximation and generalization analysis of deep CNNs and related structured deep neural networks.

Tuesday, Feb 7, 2023

09:00-10:00

Title: Bayesian calibration for multivariate posterior distributions
Author: Kate Lee, University of Auckland

Abstract: When we carry out Bayesian inference it is often convenient, even when not strictly necessary, to make approximations. If we use an approximate posterior distribution, the resulting inference and summary are likely to be damaged. The question is whether the damage is acceptable, possibly fixable. One calibration procedures is checking the coverage of approximate Bayesian credible sets including intervals estimated using Monte Carlo methods. Typically Highest Posterior Density (HPD) credible sets for univariate marginal posterior distributions are reported for ease of computation and visualisation however, they can mislead. A joint credible set can be used to extract a wealth of information about the structure of high dimensional posterior distribution. In this talk, I will present Bayesian calibration procedures, recalibration and how to obtain a joint credible set for multivariate posterior distributions.

10:00-10:30

Title: Best subset selection in linear and non-linear regressions via continuous optimization
Author: Sarat Badu Moka, UNSW Sydney

Abstract: We consider the problem of best subset selection in linear and non-linear regression models, where the goal is to find for every model size k, a subset of k features that best fit the response. This problem is particularly challenging when the total available number of features is very large compared to the number of data samples. I will present a novel continuous optimization method that identifies the best suitable subset for each k in both linear and non-linear regression models. Our approach is fast and more efficient than the well-known subset selection methods and thus making subset selection possible even when the number of features is well in excess of thousands. I will also present simulation results to show the outstanding performance of our method in comparison to the existing popular methods, including Lasso, Mixed Integer Optimization, and Forward Stepwise.

14:00-15:00

Title: Machine learning meets numerical simulation
Author: Jochen Garcke, University Bonn and Fraunhofer SCAI

Abstract: We present a conceptual overview to bridge the knowledge gap between the two separate communities of machine learning and numerical simulation, to identify and illustrate potential combined approaches, and to promote the development of hybrid systems.

As concrete examples of a combination in each direction, we highlight our work on the development of data analysis methods for damage detection in wind turbine gearboxes or wind turbine rotor blades using simulated data, and (time allowing) on machine learning-assisted simulation evaluation in the automotive industry. In the end we discuss where we see particular further potential for hybrid system.

Wednesday, Feb 8, 2023

09:00-10:00 ***(Online talk, 22:00-23:00 Tuesday, Feb 7, UK)***

Title: Convergence analysis of non-stationary and deep Gaussian process regression
Author: Aretha Teckentrup, University of Edinburgh

Abstract: We are interested in the task of estimating an unknown function from data, given as a set of point evaluations. In this context, Gaussian process regression is often used as a Bayesian inference procedure, and we are interested in the convergence as the number of data points goes to infinity. Using results from scattered data approximation, we provide a convergence analysis of the method applied to a given, unknown function of interest. We are particularly interested in the case of non-stationary covariance kernels, and the extension of the results to deep Gaussian processes.

This is joint work with Conor Osborne (University of Edinburgh).

10:00-10:30

Title: Sparse group variable selection to leverage pleiotropic association
Author: Benoit Liquet-Weiland, Macquarie University

Abstract: Genome-wide association studies (GWAS) have identified genetic variants associated with multiple complex diseases. We can leverage this phenomenon, known as pleiotropy, to integrate multiple data sources in a joint analysis. Often integrating additional information such as gene pathway knowledge can improve statistical efficiency and biological interpretation. In this talk, I review recent several frequentist ad Bayesian statistical methods we developed which incorporate both gene pathway and pleiotropy knowledge to increase statistical power and identify important risk variants affecting multiple traits. Our methods are applied to identify potential pleiotropy in an application considering the joint analysis of thyroid and breast cancers.

14:00-15:00

Title: Cost-efficient Bayesian inference with expensive computer experiments
Author: Simon Mak, Duke University

Abstract: Data science is at a critical and defining crossroad. On one hand, with remarkable breakthroughs in mathematical modeling and experimental technology, reliable data is now obtainable for complex scientific systems that were previously unobservable. On the other hand, the generation of such high-fidelity data requires costly virtual experiments, which greatly limits the amount of data for scientific discovery – a critical bottleneck in modern scientific studies. My research aims to bridge this gap by developing Bayesian methods (supported by theory algorithms) that embed scientific knowledge as prior information. This fusing of "data" and "science" within a Bayesian framework allows for principled integration of scientific prior knowledge, which then enables more accurate and precise scientific findings on high-dimensional parameter spaces given a limited experimental cost budget. I will present a suite of recent Bayesian methods developed by our group that tackle this integration, motivated by ongoing collaborations in high-energy physics, aerospace engineering and bioengineering.

Thursday, Feb 9, 2023

09:00-10:00 ***(Online talk, 23:00-24:00 Wednesday, Feb 8, Austria)***

Title: Opportunities and limitations for deep learning in the sciences
Author: Philipp Grohs, University of Vienna

Abstract: In a recent effort to push modern tools from machine learning into several areas of science and engineering, deep learning based methods have emerged as a promising alternative to classical numerical schemes for solving problems in the computational sciences – example applications include fluid dynamics, computational finance, or computational chemistry.

This talk seeks to illuminate the limitations and opportunities of this approach, both on a mathematical and an empirical level. In a first part we present computational hardness results for deep learning based algorithms and find that the computational hardness of a deep learning problem highly depends on the specific norm in which the error is measured. In a second part we present a deep learning based numerical algorithm that outperforms the previous state of the art in solving the multi electron Schrödinger equation – one of the key challenges in computational chemistry.

10:30-11:30

Title: High dimensional approximation and the curse of dimensionality
Author: Ian Sloan, UNSW, Sydney

Abstract: High dimensional approximation problems commonly arise from parametric PDE problems in which an input random field depends on very many univariate random variables. Typically (for example, in the method of "generalized polynomial chaos", or GPC) the dependence on these variables is modelled by multivariate polynomials, leading to exponentially increasing difficulty and cost (often expressed as the "curse of dimensionality") as the dimension increases (which is why sparsity of coefficients is necessarily a major theme). This lecture makes the case, when the domain of the random variables is bounded, for using instead periodic random variables, as proposed in 2020 by Kaarnioja, Kuo and Sloan for integration, and for L_2 approximation in recent joint work with Kuo, Kaarnioja, Kazashi and Nobile. In that work the approximation is a linear combination of kernels, with the kernels located at lattice points, as advocated long ago by Hickernell and colleagues. The advantage is that the cost can grow merely linearly or quadratically with dimension, giving no cause to appeal to sparsity of coefficients. A cost comparison will be made with a GPC method based on multivariate Chebyshev polynomials of the first kind.

14:00-15:00

Title: Variable importance and explainable AI
Author: Art Owen, Stanford University

Abstract: In order to explain what a black box algorithm does we can start by studying which variables are important for its decisions. Variable importance is studied by making hypothetical changes to predictor variables. Changing parameters one at a time can produce input combinations that are outliers or very unlikely. They can be physically impossible, or even logically impossible. It is problematic to base an explanation on outputs corresponding to impossible inputs. We introduced the cohort Shapley (CS) measure to avoid this problem,

based on Shapley value from cooperative game theory. There are many tradeoffs in picking a variable importance measure, so CS is not the unique reasonable choice. One interesting property of CS is that it can detect 'redlining', meaning the impact of a protected variable on an algorithm's output when that algorithm was trained without the protected variable.

This talk is based on recent joint work with Masayoshi Mase and Ben Seiler. The opinions expressed are my own, and not those of Stanford, the National Science Foundation, or Hitachi, Ltd.

Friday, Feb 10, 2023

09:00-10:00 *(Online talk, 23:00-24:00 Thursday, Feb 9, Germany)*

Title: Computational Mathematics: From Artificial Intelligence to Quantum-Computing
Author: Gitta Kutyniok, LMU Munich

Abstract: Artificial intelligence is currently impacting computational mathematics in an unprecedented manner, in particular, the areas of inverse problems and partial differential equations. After an introduction into this vibrant research area, we will focus on fundamental limitations of deep neural networks and related approaches in terms of computability. Using a Turing machine as the canonical model for today's hardware, we will show that for many problem settings deep neural networks are in fact not Turing-computable, which seriously affects their reliability. Even worse, this is a more general obstacle for computational mathematics, not only for applications of artificial intelligence. Intriguingly, it will turn out that certain analog models do ensure computability, pointing towards the necessity for novel future hardware platforms such as quantum computing for reliable computations.

10:00-10:30

Title: Sparse grid approximation of stochastic micromagnetic equations
Author: Andrea Scaglioni, Vienna University of Technology

Abstract: We consider the stochastic Landau-Lifschitz-Gilbert problem, an SPDE model for dynamic micromagnetism. We first convert the problem to a (highly nonlinear) PDE with parametric coefficients using the Doss-Sussmann transform and the Lévy-Ciesielsky parametrization of the Brownian motion. We prove analytic regularity of the parameter-to-solution map and estimate its derivatives. These estimates are used to prove dimension-independent algebraic convergence of a family of piecewise-polynomial sparse grid methods. If time permits, we discuss space and time discretization and efficient multilevel strategies for large-scale simulation.

MATRIX

Week 2

More information on informal open problem sessions will be provided in due course.

14:00-14:30, Monday, Feb 13

Title: Convergence of Randomized Kaczmarz Algorithms in Hilbert Spaces
Author: Xin Guo, The University of Queensland

Abstract: The Kaczmarz algorithm was first introduced in 1937 to solve large systems of linear equations. Existing works on the convergence analysis of the randomized Kaczmarz algorithm typically provide exponential rates of convergence, with the base tending to one as the condition number of the system increases. Results of this kind do not work well for large systems of linear equations, and do not apply to the online algorithms on Hilbert spaces for machine learning. In this talk, we provide a condition number-free analysis, which leads to polynomial rates of weak convergence for the randomized Kaczmarz algorithm. We also show the applications to kernel-based machine learning.

Information on MATRIX Annals

Overview

Each year Springer will publish a book, entitled the MATRIX Annals, dedicated to articles related to MATRIX activities. The editorial board is:

The organisers of each MATRIX program appoint a guest editor, who will organise appropriate peer review and will ensure the scientific quality of articles. Each article will be of one of the following two types:

Peer-reviewed articles

These articles contain original results or be review articles on a topic related to the MATRIX program. They should contain relevant background material and references. Articles stemming from collaborations at MATRIX or discussing open problems worked on at MATRIX are particularly encouraged. These articles can be in their final form, but this is not essential. It is possible to write a short (2-5 page) paper, and then at a later date, publish a full paper of the same results in a regular journal. Articles can be single-authored or collaborative, possibly including co-authors who did not attend the MATRIX program. Previous editions of the MATRIX Annals have been indexed by MathSciNet and zbmath.

Other contributed articles

These articles will not be peer-reviewed, can be short or long, and will most likely be expository lecture notes based on talks or activities at MATRIX, containing references to results.

Copyright and timeline

As participants in our research program, please consider to contribute articles to MATRIX Annals 2023. Note that authors are required to sign a Springer "consent to publish" form, but may post a pre-publication version of their article on arXiv or on a personal web page. Articles should be produced according to the following timetable (starting at the end of the program):

2 months: peer-reviewed articles to be submitted to guest-editor, who organises refereeing

4 months: non-peer-reviewed articles to be submitted to guest-editor

4 months: referee reports to be submitted to guest-editor

5 months:

- articles to be finalised by guest-editor
- guest-editor writes a summary of the program, and introducing the articles, for inclusion in the preface of the book
- guest editor completes online spreadsheet with information about the articles
- guest-editor sends latex and pdf files to editor-in-chief and office@matrix-inst.org.au

- MATRIX staff process the articles
- articles are checked and approved by editorial board
- MATRIX staff upload articles to MATRIX website, and sends them to Springer

7 months: Springer processes files, and produces galley proofs

9 months: corresponding authors check galley proofs

11 months: Springer publishes articles online, and later in printed book form

Our guest editor will contact you for further details regarding the submission and review process after the program. Thank you for considering submitting your articles to MATRIX Annals 2023.

Program Organisers

Josef Dick, The University of New South Wales, Sydney

Michael Feischl, TU Wien

Peter Kritzer, Austrian Academy of Sciences

Houying Zhu, Macquarie University