Modern Challenges in Data Decentralization: Federated Learning, Differential Privacy and Communication Constraints

Workshop Two on Data Privacy

Abstracts

13–17 July 2026

July 1, 2026

Permutation Testing under Local Differential Privacy

13 July
10.30 am

Tom Berrett
University of Warwick, UK

In this talk I will discuss recent work on two-sample testing under a local differential privacy constraint where a permutation procedure is used to calibrate the tests. While permutation testing is a classical statistical resampling technique, popular due to its ease of implementation and uniform Type I error control, its use under local privacy constraints is complicated by the fact that access to the data is limited. Building on previous work from private and non-private literatures, in this work we design appropriate privacy mechanisms, both interactive and non-interactive, that allow for permutation tests. Our analysis shows that these lead to minimax optimal separation rates in both discrete and continuous settings, with interactive procedures being significantly more powerful.

Federated Learning for Functional Mean Estimation: Optimality under Distributed Privacy Constraints

14 July
10.30 am

Tony Cai
University of Pennsylvania, USA

Federated learning (FL) is a distributed machine learning technique designed to preserve data privacy and security, and it has gained significant importance due to its broad range of applications. In this talk, we discuss the problem of optimal functional mean estimation from discretely sampled data in a federated differential privacy setting.

We consider both common and independent design settings. In the common design setting, the same design points are measured for each individual, whereas in the independent design, each individual has their own random collection of design points. Within this framework, we establish minimax upper and lower bounds for the estimation error of the underlying mean function, highlighting the nuanced differences between common and independent designs under distributed privacy constraints.

We propose algorithms that achieve the optimal trade-off between privacy and accuracy and provide optimality results that quantify the fundamental limits of private functional mean estimation across diverse distributed settings. These results characterize the cost of privacy and offer practical insights into the potential for privacy-preserving statistical analysis in federated environments.

The Blessing of Heterogeneity in Federated Reinforcement Learning

13 July
9 am

Yuejie Chi
Yale University, USA

Reinforcement learning (RL), concerning decision making in uncertain environments, lies at the heart of modern artificial intelligence. Due to the high dimensionality, training of RL agents typically requires a significant amount of computation and data to achieve desirable performance. However, data collection can be extremely time-consuming with limited access in real-world applications, especially when performed by a single agent. On the other hand, it is plausible to leverage multiple agents to collect data simultaneously, under the premise that they can learn collaboratively without the need of sharing local data in a federated manner. This talk addresses the fundamental statistical complexity in the algorithmic designs of federated RL algorithms, in the presence of data and task heterogeneities across the agents.

SMART Fine-tuning Factor Augmented Neural Lasso

14 July
9 am

Jianqing Fan
Princeton University, USA

Fine-tuning is a widely used strategy for adapting pre-trained models to new tasks, yet its methodology and theoretical properties in high-dimensional nonparametric settings with variable selection have not yet been developed. We propose a source-model-augmented residual tuning (SMART) framework, which incorporates the pre-trained source model into the target learner and estimates only the residual target-specific component. The approach is widely applicable, from parametric and sparse models to neural networks and blackbox machine learning models. We focus on the development of fine-tuning factor-augmented neural Lasso, resulting in SMART-FAN-Lasso. This transfer-learning framework for high-dimensional nonparametric regression with variable selection simultaneously handles covariate and posterior shifts. We use a low-rank factor structure to manage high-dimensional dependent covariates and a residual tuning decomposition in which the target function is expressed as a function of the source model and other target-specific variables, thereby reducing the effective complexity of the target task. We derive minimax-optimal excess risk bounds for SMART-FAN-Lasso, characterizing the precise conditions, in terms of relative sample sizes and function complexities, under which fine-tuning yields statistical acceleration over single-task learning. Extensive numerical experiments across diverse covariate- and posterior-shift scenarios demonstrate that SMART-FAN-Lasso consistently outperforms standard baselines and achieves near-oracle performance even under severe target sample size constraints, empirically

validating the derived rates.

Joint work with Jinhang Chai, Cheng Gao, and Qishuo Yin

Statistical Inference for Decentralized Federated Learning

17 July
3.30 pm

Jia Gu

Zhejiang University, China

This paper considers decentralized Federated Learning (FL) under heterogeneous distributions among distributed clients or data blocks for the M-estimation. The mean squared error and consensus error across the estimators from different clients via the decentralized stochastic gradient descent algorithm are derived. The asymptotic normality of the Polyak–Ruppert (PR) averaged estimator in the decentralized distributed setting is attained, which shows that its statistical efficiency comes at a cost as it is more restrictive on the number of clients than that in the distributed M-estimation. To overcome the restriction, a one-step estimator is proposed which permits a much larger number of clients while still achieving the same efficiency as the original PR-averaged estimator in the nondistributed setting. The confidence regions based on both the PR-averaged estimator and the proposed one-step estimator are constructed to facilitate statistical inference for decentralized FL.

Inference for General Linear Functionals in High-Dimensional Sparse Regression

16 July
2 pm

Dongming Huang

National University of Singapore, Singapore

We study the problem of testing whether a general linear functional of the regression vector equals a specified value in high-dimensional sparse linear regression with Gaussian random design and unknown design covariance. The loading vector is arbitrary, and the exact sparsity level is unknown but bounded by a known upper bound. Tests are required to control Type I error uniformly over the sparse null parameter space determined by this upper bound, while power is evaluated against alternatives with the true sparsity level. We construct a computationally efficient mixed test that gives an upper bound on the adaptive separation distance, and we establish an information-theoretic lower bound calibrated to the magnitude profile of the loading vector. In the ultra-sparse case, these bounds characterize the adaptive separation rate up to logarithmic factors for arbitrary loading vectors. In the

moderately sparse case, these bounds match for several classes of loading vectors but may differ in general. In this case, we further prove a low-degree lower bound that matches the upper bound up to logarithmic factors. This provides evidence that improving on the rate of the mixed test, when statistically possible, may be computationally hard. For flat sparse loadings, we complement this evidence with a polynomial-time reduction from sparse canonical correlation analysis. Finally, we examine how information about the design covariance affects the adaptive separation rate in two settings. Under a sparse signed-spiked covariance model, the information-theoretic lower bound is attainable up to logarithmic factors by a computationally inefficient procedure, while the low-degree lower bound and sparse-canonical-correlation-analysis reduction continue to apply, providing evidence for a statistical-computational gap. When the design covariance is known and diagonal, the adaptive separation rate takes the same form as in the ultra-sparse case.

Contextual Dynamic Pricing: Algorithms, Optimality, and Local Differential Privacy Constraints

16 July
3.30 pm

Feiyu Jiang
Fudan University, China

We study contextual dynamic pricing problems where a firm sells products to T sequentially-arriving consumers, behaving according to an unknown demand model. The firm aims to minimize its regret over a clairvoyant that knows the model in advance. The demand follows a generalized linear model (GLM), allowing for stochastic feature vectors in $\mathbb{R}^d$ encoding product and consumer information. We first show the optimal regret is of order $\sqrt{dT}$, up to logarithmic factors, improving existing upper bounds by a $\sqrt{d}$ factor. This optimal rate is materialized by two algorithms: a confidence bound-type algorithm and an explore-then-commit (ETC) algorithm. A key insight is an intrinsic connection between dynamic pricing and contextual multi-armed bandit problems with many arms with a careful discretization. We further study contextual dynamic pricing under local differential privacy (LDP) constraints. We propose a stochastic gradient descent-based ETC algorithm achieving regret upper bounds of order $d\sqrt{T}/\epsilon$, up to logarithmic factors, where $\epsilon > 0$ is the privacy parameter. The upper bounds with and without LDP constraints are matched by newly constructed minimax lower bounds, characterizing costs of privacy. Moreover, we extend our study to dynamic pricing under mixed privacy constraints, improving the privacy-utility tradeoff by leveraging public data. This is the first time such setting is studied in the dynamic pricing literature and our theoretical results seamlessly bridge dynamic pricing with and without LDP. Extensive numerical experiments and real data applications are conducted to illustrate the efficiency and practical value of our algorithms.

Learning, Approximation and Control

14 July
2 pm

Qianxiao Li

National University of Singapore, Singapore

We discuss some interesting problems and recent results on the interface of deep learning, approximation theory and control theory. Through a dynamical system viewpoint of deep residual architectures, the study of model complexity in deep learning can be formulated as approximation or interpolation problems that can be studied using control theory, but with a mean-field twist. In a similar vein, training deep architectures can be formulated as optimal control problems in the mean-field sense. We provide some basic mathematical results on these new control problems that so arise, and discuss some applications in improving efficiency, robustness and adaptability of deep learning models.

Locally Private Estimation with Public Features

16 July
9 am

Yuheng Ma

East China Normal University, China

We initiate the study of locally differentially private (LDP) learning with public features. We define semi-feature LDP, where some features are publicly available while the remaining ones, along with the label, require protection under local differential privacy. Under semifeature LDP, we consider three fundamental estimation problems: non-parametric density estimation, classification, and regression. Given the smoothness assumption, we show that the minimax convergence rate is significantly improved compared to classical LDP. Then, we propose HistOfTree, an estimator that fully leverages the information contained in both public and private features. Theoretically, HistOfTree reaches the mini-max optimal convergence rate. Empirically, HistOfTree achieves superior performance on both synthetic and real data. We also explore scenarios where users have the flexibility to select features for protection manually. In such cases, we propose an estimator and a data-driven parameter tuning strategy, leading to analogous theoretical and empirical results.

TBA

Short Talk
17 July
10.30 am

Paul Rognon-Vael
Bocconi University, Italy

Pending

Anti-Concentration Inequalities for the Difference of Maxima of Gaussian Random Vectors

17 July
2 pm

Shuting Shen
National University of Singapore, Singapore

We derive novel anti-concentration bounds for the difference between the maximal values of two Gaussian random vectors under various settings. Our bounds are dimension-free, scaling with the dimension of the Gaussian vectors only through the smaller expected maximum of the Gaussian subvectors. Meanwhile, our bounds remain valid under degenerate covariance structures, which previous results do not cover. In addition, we show that our conditions are sharp under the homogeneous component-wise variance setting, while we only impose some mild assumptions on the covariance structures under the heterogeneous variance setting. We apply the new anti-concentration bounds to derive the central limit theorem for the maximizers of discrete empirical processes. Finally, we back up our theoretical findings with comprehensive numerical studies.

TBA

Short Talk
17 July
11 am

Tao Shen
National University of Singapore, Singapore

Pending

Muon Outperforms Adam in Tail-End Associative Memory Learning

15 July
9 am

Vincent Tan

National University of Singapore, Singapore

The Muon optimizer is consistently faster than Adam in training Large Language Models (LLMs), yet the mechanism underlying its success remains unclear. This paper demystifies this mechanism through the lens of associative memory. By ablating the transformer components optimized by Muon, we reveal that the associative memory parameters of LLMs, namely the Value and Output (VO) attention weights and Feed-Forward Networks (FFNs), are the primary contributors to Muon's superiority. Motivated by this associative memory view, we then explain Muon's superiority on real-world corpora, which are intrinsically heavy-tailed: a few 'head' classes are extremely frequent, while a vast number of 'tail' classes are individually rare. The superiority is explained through two key properties: (i) its update rule consistently yields a more isotropic singular spectrum than Adam; and as a result, (ii) on heavy-tailed data, it optimizes tail classes more effectively than Adam. Beyond empirical evidence, we theoretically confirm these findings by analyzing a one-layer associative memory model under class-imbalanced data. We prove that Muon consistently achieves balanced learning across classes regardless of feature embeddings, whereas Adam can induce large disparities in learning errors depending on embedding properties. In summary, our empirical observations and theoretical analyses reveal Muon's core advantage: its update rule aligns with the outer-product structure of linear associative memories, enabling more balanced and effective learning of tail classes in heavy-tailed distributions than Adam.

Joint Work with Shuche Wang, Fengzhuo Zhang, Jiaxiang Li, Cunxiao Du, Chao Du, Tianyu Pang, Zhuoran Yang, Mingyi Hong

---

Optimal Estimation, Adaptation and Inference for Linear Functionals under Differential Privacy

14 July
3.30 pm

Lasse Vuursteen

Duke University, USA

Differential privacy protects individual-level data, but it can sharply change the statistical difficulty of nonparametric estimation. In this talk I consider the problem of estimating linear functionals of an unknown function under differential privacy, including adaptive estimation when the underlying function class is not known in advance. We work in a federated framework in which data are distributed across m servers, each holding n observations, thereby encompassing both central and local

differential privacy as special cases.

Our main results give sharp minimax and adaptive risk characterizations through moduli of continuity. In the non-adaptive setting, the private risk is governed by a joint modulus involving total variation at the privacy-driven radius $(\sqrt{m}\, n\varepsilon)^{-1}$ and Hellinger or L_2-type geometry at the classical statistical radius $(mn)^{-1/2}$. For adaptation over multiple function classes, a between-class modulus determines whether adaptation is free, costs only logarithmic factors, or incurs a genuine polynomial penalty. A key message is that privacy changes not only rates but the relevant geometry of the problem: in Gaussian white noise the privacy and classical branches collapse to a single L_2-modulus, whereas in density estimation the private branch is governed by an L_1-modulus and the classical branch by an L_2-modulus, producing new phase transitions and privacy-induced adaptation costs. I will also describe constructive procedures attaining these rates, including clipping-and-inversion methods for unbounded observations and sensitivity-adapted private queries for density estimation.

Understanding Partial Transfer in CNNs via Kronecker Product Regression

16 July
10.30 am

Junhui Wang

Chinese University of Hong Kong, Hong Kong SAR

Transfer learning (TL) has emerged as a powerful approach in modern machine learning. While traditional TL methods often presume global similarity between tasks, this assumption is frequently invalid, as many practical applications involve only partial similarities. Convolutional neural networks (CNNs) provide a notable example for exploring partial TL, where it is generally assumed that only the convolutional filters—not the entire network—are similar across different tasks. In this talk, we adopt a Kronecker product regression model to analyze a class of CNNs, recasting filter transfer as factor transfer, with the decision to retrain these factors distinguishing the two common transfer strategies in CNNs; i.e., fixed feature extraction and fine-tuning. Our framework supports transfer learning between tasks of varying types or dimensions, increasing its versatility and practical relevance. Theoretical analysis demonstrates that fixed feature extraction can substantially enhance estimation accuracy when source and target filters are well-aligned. In contrast, fine-tuning provides greater robustness to model mismatches and can lead to improved computational efficiency. Extensive simulations and real-world experiments further validate these results, highlighting the respective benefits and trade-offs of each transfer learning strategy.

NetPTR: Optimal Differentially Private Spectral Community Detection

13 July
2 pm

Wanjie Wang

National University of Singapore, Singapore

Spectral community detection estimates latent labels from the leading eigenspace of a network adjacency matrix, but releasing the resulting labels can disclose sensitive relational information.

We consider this problem under differential privacy for both ordinary and bipartite networks. For ordinary networks, the protected unit is a single edge, leading to edge differential privacy (edge-DP). For bipartite networks, the inferential target is the community structure of the left-side nodes, while the protected unit is an entire right-side incidence profile, leading to column-node-DP.

We propose NetPTR, a private spectral clustering procedure that releases a noisy empirical spectral embedding after a stability test. The algorithm requires perturbation bounds for empirical eigenspaces under neighboring-network changes, which yield computable stability certificates and local sensitivity bounds. For ordinary networks, we establish edge-DP and the error bound under the degree-corrected stochastic blockmodel, which separates the non-private spectral clustering error from the additional privacy-induced error. It therefore guarantees weak consistency in sparse networks and exact recovery in moderately sparse networks. A matching lower bound shows that the required privacy budget is sharp up to logarithmic factors.

We further develop a column-node-DP algorithm for bipartite networks and prove consistency under a bipartite degree-corrected block model. Simulations and real-data examples illustrate the resulting privacy–accuracy tradeoff.

Robust Multilayer Networks Estimation under Batch Contamination

17 July
9 am

Yi Yu

University of Warwick, UK

Pending

Quantized Signal Recovery at Optimal Rates

15 July
10.30 am

Ming Yuan
Columbia University, USA

Federated and privacy-aware learning often require clients to transmit compressed, randomized, or low-bit summaries rather than raw data. Quantization is therefore both a communication tool and an information bottleneck. This talk asks a basic statistical question: what can still be recovered after severe quantization? I will present recent results on signal recovery from quantized measurements, including one-bit and multi-bit compressed sensing with dithering, as well as one-bit phase retrieval from quantized magnitude-only observations. We establish near-optimal information-theoretic recovery rates for structured signals and develop efficient projected-gradient algorithms based on a one-sided loss that match these rates up to logarithmic factors. The results show that surprisingly little numerical precision is needed for accurate recovery.

When Averaging Fails: Statistical Structure in Federated Learning

13 July
3.30 pm

Qiong Zhang
Renmin University of China, China

Federated learning is often told as a simple story: each site learns from its own data, sends a summary to the server, and the server averages the pieces into a global answer. This talk starts from the places where that story fails. The first failure occurs in distributed learning of finite mixture models. Here, each local machine may discover the same latent subpopulations, but label switching and the non-Euclidean geometry of the parameter space make ordinary averaging statistically meaningless. This motivates mixture reduction, which treats aggregation as a problem of aligning and reducing local mixture distributions rather than averaging coordinates. The second failure occurs in heterogeneous federated learning. When clients represent different hospitals, populations, or environments, their differences are not always noise to be smoothed out; they may encode specialization. Instead of asking only how the server should average client models, we can ask a different question: when a new query arrives, which client is best suited to handle it? This leads to routing-enabled federated learning, where heterogeneity becomes a source of information for prediction and decision-making. I will also briefly discuss how these two directions extend to Byzantine-tolerant distributed learning and to dual heterogeneity, where variation appears both across clients and within each client. The main message is that federated learning should not be designed around averaging as a reflex; it should be designed

around the statistical structure of what is being learned.
