LSE Department of Statistics

Archive of Data Science Seminars - Michaelmas Term 2020

Monday 5 October 2020, 3-4pm - Joshua Loftus

Title: Algorithmic fairness: causal paths beyond impossibility

Abstract: For the past several years scholars, journalists, and the public have given more attention to the social impacts of data and information technology. Many are concerned that algorithmic decision systems may affect people negatively, especially in ways that reinforce harmful patterns in historic data related to attributes like race, gender, and other ethically or legally important attributes. Some of this recent work has rediscovered fundamental issues in political philosophy and economics, connecting the current state of research on fair algorithms to, for example, impossibility results in social choice theory. This talk will briefly survey the emerging field and then focus on causal modeling as a pathway to reach consensus.

…………………………………………………………………………………………………...

Monday 2 November 2020, 3-4pm - Moez Driaef

Title: Vignettes in Computational Wireless Networks

Abstract: Wireless communication networks are at the core of the infrastructure enabling the mobile internet. They rely on large and complex set of equipment known as base stations. These towers packed with antennas have hundreds of parameters that need to be tuned to deliver the quality of service that we require to watch videos online, browse the web etc. In this presentation, I will talk about some of my recent work on the use of contextual bandit approaches, popular for sponsored search, in the context of a problem known as cellular handover in wireless communication.

…………………………………………………………………………………………………

Monday 16 November 2020, 3-4pm - Arthur Gretton

Title: Generalized Energy-Based Models

Abstract: I will introduce Generalized Energy Based Models (GEBM) for generative modelling. These models combine two trained components: a base distribution (generally an implicit model), which can learn the support of data with low intrinsic dimension in a high dimensional space; and an energy function, to refine the probability mass on the learned support. Both the energy function and base jointly constitute the final model, unlike GANs, which retain only the base distribution (the "generator"). In particular, while the energy function is analogous to the GAN critic function, it is not discarded after training.

GEBMs are trained by alternating between learning the energy and the base, much like a GAN. Both training stages are well-defined: the energy is learned by maximising a generalized likelihood, and the resulting energy-based loss provides informative gradients for learning the base. Samples from the posterior on the latent space of the trained model can be obtained via MCMC, thus finding regions in this space that produce better quality samples. Empirically, the GEBM samples on image-generation tasks are of better quality than those from the learned generator alone, indicating that all else being equal, the GEBM will outperform a GAN of the same complexity. GEBMs also return state-of-the-art performance on density modelling tasks, and when using base measures with an explicit form.

……

Monday 30 November 2020, 3-4pm - Professor Afonso S. Bandeira

Title: Computational Hardness of Hypothesis Testing and Quiet Plantings

Abstract: When faced with a data analysis, learning, or statistical inference problem, the amount and quality of data available fundamentally determines whether such tasks can be performed with certain levels of accuracy. With the growing size of datasets however, it is crucial not only that the underlying statistical task is possible, but also that is doable by means of efficient algorithms. In this talk we will discuss methods aiming to establish limits of when statistical tasks are possible with computationally efficient methods or when there is a fundamental «Statistical-to-Computational gap›› in which an inference task is statistically possible but inherently computationally hard. We will focus on Hypothesis Testing and the ``Low Degree Method'' and also address hardness of certification via ``quiet plantings''. Guiding examples will include Sparse PCA, bounds on the Sherrington Kirkpatrick Hamiltonian, and lower bounds on Chromatic Numbers of random graphs.

Archive of Data Science Seminars – Lent Term 2021

Monday 18 January 2021, 3-4pm - Zhaoran Wang

Biography: Zhaoran Wang is an assistant professor at Northwestern University, working at the interface of machine learning, statistics, and optimization. He is the recipient of the AISTATS (Artificial Intelligence and Statistics Conference) notable paper award, ASA (American Statistical Association) best student paper in statistical learning and data mining, INFORMS (Institute for Operations Research and the Management Sciences) best student paper finalist in data mining, Microsoft Ph.D. fellowship, Simons-Berkeley fellowship, and NSF CAREER award.

Title: Demystifying (Deep) Reinforcement Learning: The Optimist, The Pessimist, and Their Provable Efficiency.

Abstract: Coupled with powerful function approximators such as deep neural networks, reinforcement learning (RL) achieves tremendous empirical successes. However, its theoretical understandings lag behind. In particular, it remains unclear how to provably attain the optimal policy with a finite regret or sample complexity. In this talk, we will present the two sides of the same coin, which demonstrates an intriguing duality between optimism and pessimism.

- In the online setting, we aim to learn the optimal policy by actively interacting with the environment. To strike a balance between exploration and exploitation, we propose an optimistic least-squares value iteration algorithm, which achieves a $\sqrt{T}$ regret in the presence of linear, kernel, and neural function approximators.

- In the offline setting, we aim to learn the optimal policy based on a dataset collected a priori. Due to a lack of active interactions with the environment, we suffer from the insufficient coverage of the dataset. To maximally exploit the dataset, we propose a pessimistic least-squares value iteration algorithm, which achieves a minimax-optimal sample complexity.

……………………………………………………………………………………………………….

Monday 1 February 2021, 3-4pm - Nan Jiang

Biography: Nan Jiang is an Assistant Professor of Computer Science at University of Illinois at Urbana-Champaign. Prior to joining UIUC, he was a postdoc researcher at Microsoft Research

NYC. He received his PhD in Computer Science and Engineering at University of Michigan. His research interests lie in the theory of reinforcement learning (RL), mostly focusing on sample efficiency. Specific research topics include sample complexity of exploration under function approximation, off-policy evaluation and validation, batch RL with insufficient data coverage, and spectral learning of dynamical systems. He is the recipient of the Best Paper Award at AAMAS 2015 and Rackham Predoctoral Fellowship in 2016.

Title: Reinforcement learning with less demanding function approximation.

Abstract: When value-function approximation is used in reinforcement learning (RL), what kind of expressivity properties do we need from the function class? Perhaps surprisingly, the standard intuition from supervised learning, that the function class should simply contain (a close approximation of) the function to be learned (a.k.a. the so-called *realizability* assumption), often does not carry over to RL, and the theoretical guarantees of standard RL algorithms often demand substantially stronger assumptions. It was unclear whether algorithms that only require realizability even exist for general stochastic environments, and even more basic versions of the question---such has how to identify the optimal value function out of merely 2 candidate functions using batch data---was wildly open. In this talk, I will present recent positive progress on this topic, with a focus on BVFT (Xie & Jiang, 2020), an information-theoretic algorithm for batch RL that only imposes realizability on the function class. Time permitting I will also discuss algorithms that require less demanding function approximation under other RL settings, such as online exploration (e.g., OLIVE; Jiang et al., 2017) and off-policy evaluation.

…….

Monday 15 March 2021, 3-4pm - Samory K. Kpotufe

Biography: Samory is an Associate Professor of Statistics at Columbia University. He works in statistical machine learning, with an emphasis on common nonparametric methods (e.g., kNN, trees, kernel averaging). He is particularly interested in *adaptivity*, i.e., how to automatically leverage beneficial aspects of data as opposed to designing specifically for each scenario. This involves characterizing statistical limits, under modern computational and data constraints, and identifying favorable aspects of data that help circumvent these limits.

Title: Some Recent Insights on Transfer-Learning

Abstract: A common situation in Machine Learning is one where training data is not fully representative of a target population due to bias in the sampling mechanism or due to prohibitive

target sampling costs. In such situations, we aim to 'transfer' relevant information from the training data (a.k.a. source data) to the target application. How much information is in the source data about the target application? Would some amount of target data improve transfer? These are all practical questions that depend crucially on 'how far' the source domain is from the target. However, how to properly measure 'distance' between source and target domains remains largely unclear.

In this talk we will argue that much of the traditional notions of 'distance' (e.g. KL-divergence, extensions of TV such as D_A discrepancy, density-ratios, Wasserstein distance) can yield an over-pessimistic picture of transferability. Instead, we show that some new notions of 'relative dimension' between source and target (which we simply term 'transfer-exponents') capture a continuum from easy to hard transfer. Transfer-exponents uncover a rich set of situations where transfer is possible even at fast rates; they encode relative benefits of source and target samples, and have interesting implications for related problems such as 'multi-task or multi-source learning'.

In particular, in the case of transfer from multiple sources, we will discuss (if time permits) a strong dichotomy between minimax and adaptive rates: no adaptive procedure exists that can achieve the same rates as minimax (oracle) procedures.

The talk is based on earlier work with Guillaume Martinet, and ongoing work with Steve Hanneke.

……..

Monday 29 March 2021, 3-4pm - Weijie Su

Biography: Weijie is an Assistant Professor in the Wharton Statistics Department and in the Department of Computer and Information Science, at the University of Pennsylvania. He serves as a co-director of Penn Research in Machine Learning. Prior to joining Penn, he obtained a Ph.D. from Stanford University, under the supervision of Emmanuel Candès. His research interests include: statistical machine learning, high-dimensional inference, large-scale multiple testing, optimization, and privacy-preserving data analysis.

Title: Gaussian Differential Privacy

Abstract: Privacy-preserving data analysis has been put on a firm mathematical foundation since the introduction of differential privacy (DP) in 2006. This privacy definition, however, has

some well-known weaknesses: notably, it does not tightly handle composition. In this talk, we propose a relaxation of DP that we term "f-DP", which has a number of appealing properties and avoids some of the difficulties associated with prior relaxations. First, f-DP preserves the hypothesis testing interpretation of differential privacy, which makes its guarantees easily interpretable. It allows for lossless reasoning about composition and post-processing, and notably, a direct way to analyze privacy amplification by subsampling. We define a canonical single-parameter family of definitions within our class that is termed "Gaussian Differential Privacy", based on hypothesis testing of two shifted normal distributions. We prove that this family is focal to f-DP by introducing a central limit theorem, which shows that the privacy guarantees of any hypothesis-testing based definition of privacy (including differential privacy) converge to Gaussian differential privacy in the limit under composition. This central limit theorem also gives a tractable analysis tool. We demonstrate the use of the tools we develop by giving an improved analysis of the privacy guarantees of noisy stochastic gradient descent. This is joint work with Jinshuo Dong and Aaron Roth.