IFDS DRDS Workshop 2022: Abstracts

For the **full schedule**, please see the webpage.

Yao Xie

Date & Time: 9:00 to 9:45 am on Thursday, August 4, 2022

Title: Hypothesis tests via distributionally robust optimization, and more

Abstract: We consider a general data-driven robust hypothesis test formulation to find an optimal test (function) that minimizes the worst-case performance regarding distributions that are close to the empirical distributions with respect to some divergence, in particular, the Wasserstein and the sink horn divergences. The robust tests are beneficial, for instance, for cases with limited or unbalanced samples - such a scenario often arises from applications such as health care, online change-point detection, and anomaly detection. We present a distributionally robust optimization framework to solve such a problem and study the computational and statistical properties of the proposed test by presenting a tractable convex reformulation of the original infinite-dimensional variational problem. Finally, I will present the generalization of the approach to other related problems, including domain adaptation.

This is joint work with Rui Gao (UT Austin), Liyan Xie (CUHK-Shenzhen), and Jie Wang (Georgia Tech).

Samory Kpotufe

Date & Time: 9:45 to 10:30 am on Thursday, August 4, 2022

Title: Tracking Most Significant Arm Switches in Bandits

Abstract: In bandit with distribution shifts, one aims to automatically adapt to unknown changes in reward distribution, and restart exploration when necessary. While this problem has received attention for many years, no adaptive procedure was known till a recent breakthrough of Auer et al (2018, 2019) which guarantees an optimal (dynamic) regret (LT)^{1/2}, for T rounds and L stationary phases.

However, while this rate is tight in the worst case, it leaves open whether faster rates are possible, adaptively, if few changes in distribution are actually severe, e.g., involve no change in best arm. We provide a positive answer, showing that in fact, a much weaker notion of change can be adapted to, which can yield significantly faster rates than previously known, whether as expressed in terms of number of best arm switches--for which no adaptive procedure was

known, or in terms of total variation. Finally, our parametrization captures at once, both stochastic and non-stochastic adversarial settings.

This is joint work with Joe Suk.

Ludwig Schmidt

Date & Time: 11:00 to 11:45 am on Thursday, August 4, 2022

Title: A data-centric view on robustness

Abstract: Over the past few years, researchers have proposed many ways to measure the robustness of machine learning models. In the first part of the talk, we will survey the current robustness landscape based on a large-scale experimental study involving more than 200 different models and test conditions. Despite the large variety of test conditions, common trends emerge: (i) robustness to natural distribution shift and synthetic perturbations are distinct phenomena, (ii) current algorithmic techniques have little effect on robustness to natural distribution shifts, (iii) training on more diverse datasets offers robustness gains on several natural distribution shifts.

In the second part of the talk, we then leverage the aforementioned insights to improve OpenAI's CLIP model. CLIP achieved unprecedented robustness on several natural distribution shifts, but only when used as a zero-shot model. The zero-shot evaluation precludes the use of extra data for fine-tuning and hence leads to lower performance when there is a specific task of interest. To address this issue, we introduce a simple yet effective method for fine-tuning zero-shot models that leads to large robustness gains on several distribution shifts without reducing in-distribution performance.

Jamie Morgenstern

Date & Time: 2:00 to 2:45 pm on Thursday, August 4, 2022

Title: Endogeneous distribution shifts in competitive environments

Abstract: In this talk, I'll describe recent work exploring the dynamics between ML systems and the populations they serve, particularly when the models deployed impact what populations a system will have as future customers.

Hongseok Namkoong

Date & Time: 9 to 9:45 am on Friday, August 5, 2022

Title: Assessing the external validity (a.k.a. distributional robustness) of causal findings

Abstract: Causal inference—analyzing the ceteris paribus effect of an intervention—is key to reliable decision-making. Due to population shifts over time and underrepresentation of marginalized groups, standard causal estimands that measure the average treatment effect often lose validity outside the study population. To guarantee that causal findings remain valid over population shifts, we propose the worst-case treatment effect (WTE) across all subpopulations of a given size. We develop an optimal estimator for the WTE based on flexible prediction methods, which allows analyzing the external validity of the popular doubly robust estimator. On real examples where external validity is of core concern, our proposed framework successfully guards against brittle findings that are invalidated under unanticipated population shifts.

Aditi Raghunathan

Date & Time: 9:45 to 10:30 am on Friday, August 5, 2022

Title: Estimating and improving the performance of machine learning under natural distribution shifts

Abstract: Machine learning systems often fail catastrophically under the presence of distribution shift—when the test distribution differs in some systematic way from the training distribution. If we can mathematically characterize a distribution shift, we could devise appropriate robust training algorithms that promote robustness to that specific class of shifts. However, the resulting robust models show limited gains on shifts that do not admit the structure they were specifically trained against. Naturally occurring shifts are both hard to predict a priori and intractable to mathematically characterize necessitating different approaches to addressing distribution shifts in the wild.

In this talk, we first discuss how to estimate the performance of models under natural distribution shifts—the shift could cause a small degradation or a catastrophic drop. Obtaining ground truth labels is expensive and requires the a priori knowledge of when and what kind of distribution shifts are likely to occur. We present a phenomenon that we call *agreement-on-the-line* that allows us to effectively predict performance under distribution shift from just unlabeled data. Next, we investigate a promising avenue for improving robustness to natural shifts—leveraging representations pre-trained on diverse data. Via theory and experiments, we find that the de facto fine-tuning of pre-trained representations does not maximally preserve robustness. Using

insights from our analysis, we provide two simple alternate fine-tuning approaches that substantially boost robustness to natural shifts.

Terry Rockafellar

Date & Time: 11 to 11:45 am on Friday, August 5, 2022

Title: Robustness from the Perspective of Coherent Measures of Risk

Abstract: Measures of risk seek to quantify the overall "risk" in a cost- or loss-oriented random variable by a single value, such as its expectation or worst outcome, or better, something in between.Common sense axioms were developed for this in the late 90s by mathematicians working in finance, and they used the term "coherent" for the risk measures that satisfied those axioms. Their work was soon broadened, and it was established by way of duality in convex analysis that coherent measures of risk correspond exactly to looking for robustness with respect to some collection of alternative probability distributions.

The theory has since become highly developed with powerful approaches to quantification like conditional value-at-risk and others based on that. Now also, it includes interesting connections with statistics and regression captured by the "fundamental quadrangle of risk", as will be explained in this talk, with examples.

Rediet Abebe

Date & Time: 2 to 2:45 pm on Friday, August 5, 2022

Title: Algorithms on Trial: Interrogating Evidentiary Statistical Software

Abstract: The U.S. criminal legal system increasingly relies on software output to convict and incarcerate people. In a large number of cases each year, the government makes these consequential decisions based on evidence from statistical software---such as probabilistic genotyping, environmental audio detection and toolmark analysis tools---that the defense counsel cannot fully cross-examine or scrutinize. This undermines the commitments of the adversarial criminal legal system, which relies on the defense's ability to probe and test the prosecution's case to safeguard individual rights.

Responding to this need to adversarially scrutinize output from such software, in this talk, we propose a novel framework for examining the validity of evidentiary statistical software called Robust Adversarial Testing. We define and operationalize this notion of robust adversarial testing for defense use by drawing on a large body of recent work in robust machine learning and algorithmic fairness. We demonstrate how this framework both standardizes the process for scrutinizing such tools and empowers defense lawyers to examine their validity for instances

most relevant to the case at hand. We further discuss existing structural and institutional challenges within the U.S. criminal legal system which may create barriers for implementing this framework and close with a discussion on policy changes that could help address these concerns. We close with an outline of research directions in this burgeoning area of adversarial ML and adversarial scrutiny in the law.

This talk is based on joint and ongoing work with Moritz Hardt, Angela Jin, John Miller, Ludwig Schmidt, Rebecca Wexler, as well as conversations with numerous public defenders and forensic scientists including Nathan Adams, Clinton Hughes, Daniel Krane, and Richard Torres.

Biography: Rediet Abebe is an Assistant Professor of Computer Science at the University of California, Berkeley. She is a 2022 Andrew Carnegie Fellow and on leave as a Junior Fellow at Harvard University. Her research examines the interaction of algorithms and inequality, with a focus on contributing to the mathematical and computational foundations of this emerging research area. Abebe co-launched and served on the Executive Committee for the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO). She previously co-founded the related global research initiative, MD4SG. Abebe's work has received recognitions including the MIT Technology Reviews' 35 Innovators Under 35, the 2020 ACM SIGKDD Dissertation Award, an honorable mention by the ACM SIGEcom Dissertation Award, and the Bloomberg 50 list as a one to watch. Abebe co-founded and serves on the Board of Directors for Black in AI, a non-profit organization tackling equity issues in AI. Abebe holds a Ph.D. in computer science from Cornell University and masters degrees in mathematics from Harvard University and the University of Cambridge.

Stephen J. Wright

Date & Time: 9:00 to 9:45 am on Saturday, August 6, 2022

Title: Robust formulations and algorithms for learning problems under distributional ambiguity

Abstract: We discuss learning problems in which the empirical distribution represented by the training data is used to define an ambiguous set of distributions, and we seek the classifier or regressor that solves a min-max problem involving this set. Our focus is mainly on linear models. First, we discuss formulation of the robust min-max problem based on classification with the discontinuous "zero-one" loss function, using Wasserstein ambiguity, and describe properties of the resulting nonconvex problem, which is benignly nonconvex in a certain sense. Second, we discuss robust formulations of convex ERM problems involving linear models, where the distributional ambiguity is measured using either a Wasserstein metric or f-divergence. We show that such problems can be formulated as "generalized linear programs" and solved using a first-order primal-dual algorithm that incorporates coordinate descent in the dual variable and variance reduction. We present numerical results to illustrate properties of the robust optimization formulations and algorithms.

This talk covers joint work with Nam Ho-Nguyen, Jelena Diakonikolas, Chaobing Song, and Eric Lin.

Brian Ziebart

Date & Time: 9:45 to 10:30 am on Saturday, August 6, 2022

Title: Prediction Games: From Maximum Likelihood Estimation to Active Learning, Fair Machine Learning, and Structured Prediction

Abstract: A standard approach to supervised machine learning is to choose the form of a predictor and to then optimize its parameters based on training data. Approximations of the predictor's performance measure are often required to make this optimization problem tractable. Instead of approximating the performance measure and using the exact training data, this talk explores a distributionally robust approach using game-theoretic approximations of the training data while optimizing the exact performance measures of interest. Though the resulting "prediction games" reduce to maximum likelihood estimation in simple cases, they provide new methods for more complicated prediction tasks involving covariate shift, fairness constraint satisfaction, and structured data.

Biography: Brian Ziebart is an Associate Professor in the Department of Computer Science at the University of Illinois at Chicago and a Software Engineer at Aurora Innovation. He earned his PhD in Machine Learning from Carnegie Mellon University where he was also a postdoctoral fellow. His interests lie in the intersections between machine learning, game theory, and decision theory. He has published over 35 articles in leading machine learning and artificial intelligence venues, including a Best Paper at the International Conference on Machine Learning.

Lillian Ratliff

Date & Time: 11:00 to 11:45 am on Saturday, August 6, 2022

Title: Learning from Strategic and Decision Dependent Data

Abstract: Learning-based systems are increasingly being deployed in contexts where the environment reacts to the decision in a strategic or even adversarial fashion, thereby creating a feedback loop in which the data distribution depends on the decision or algorithm. The challenge is that this decision-dependence often arises due to conflicting objectives between the environment and the decision-maker, and comes with inherent information asymmetries due to reactive elements in the environment with preferences, unknown apriori to the decision-maker, that drive the reaction. In this talk, I will discuss some recent problem formulations in this area and corresponding promising results. I will also highlight interesting open questions and directions for future research.