B9145: Reliable Statistical Learning

Course information and syllabus

Instructor: Hong Namkoong (namkoong@gsb.columbia.edu)
Lectures: Tuesdays, 3:50PM-7:05PM, Uris 333
Format: HyFlex model; lectures will take place in Uris 333 and broadcasted via Zoom.
TA: Chao Qin (CQin22@gsb.columbia.edu)
Office hours: TBD

Description: As ML systems increasingly affect high-stakes decisions, it is critical that they maintain a reliable level of performance under operation. However, traditional modeling assumptions rarely hold in practice due to noisy inputs, shifts in environment, omitted variables, and even adversarial attacks. The standard machine learning paradigm that optimize average performance is brittle to even small distributional shifts, exhibiting poor performance on minority groups and tail inputs. Even performance of heavily engineered state-of-the-art models degrades significantly on domains that are slightly different from what the model was trained on. Lack of understanding of their failure modes highlights the need for models that reliably work, and rigorous safety tests to evaluate them.

This course surveys a range of emerging topics on reliability and robustness in machine learning. Most of the topics discussed in this class are active research areas, and relevant reading materials will draw upon recent literature (to be posted on the website). The goal of this class is to foster discussion on new research questions. This will encompass theoretical and methodological developments, modeling considerations, novel application areas, and other concerns rising out of practice.

Outline: The course will comprise of pedagogical lectures and seminar-style guided discussions. We will begin with an overview of foundational tools in statistical learning. In the first third of the class, the focus will be on how these technical tools give basic theoretical results in statistical learning and stochastic optimization. We will study standard stochastic optimization methods, generalization bounds (concentration, symmetrization, chaining), M-estimation theory (asymptotics), and fundamental hardness results (information theoretic lower bounds).

Then, we will cover the recent set of works on improving reliability in machine learning. Since reliability is a loosely defined term with many connotations, we will explore various aspects of this concept, alongside a discussion of future directions. The following is a selection of topics that will be covered in the course (subject to change).

I. Distributional robustness

- Connections to risk-aversion: Chapter 6 of Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM and Mathematical Programming Society, 2009
- Regularization perspective: Jose Blanchet, Yang Kang, and Karthyek Murthy. Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, 56(3):830–857, 2019,
  John C. Duchi and Hongseok Namkoong. Variance-based regularization with convex objectives. *Journal of Machine Learning Research*, 20(68):1–55, 2019,

Henry Lam. Robust sensitivity analysis for stochastic systems. *Mathematics of Operations Research*, 41(4):1248–1275, 2016

- Connections to causality: Nicolai Meinshausen. Causality from a distributional robustness point of view. In *2018 IEEE Data Science Workshop (DSW)*, pages 6–10. IEEE, 2018,
  Dominik Rothenhäusler, Peter Bühlmann, Nicolai Meinshausen, and Jonas Peters. Anchor regression: heterogeneous data meets causality. *arXiv:1801.06229 [stat.ME]*, 2018,
  Christina Heinze-Deml and Nicolai Meinshausen. Conditional variance penalties and domain shift robustness. *arXiv:1710.11469 [stat.ML]*, 2017

II. Ethics, fairness, and subpopulation performance

- Ethics-informed ML research practices: Timnit Gebru and Emily Denton. Tutorial on fairness accountability transparency and ethics in computer vision. `https://sites.google.com/view/f` 2020. Accessed August 2020,
  Ruha Benjamin. Reimagining the default settings of technology and society. `https://bit.ly/32kSbLp`, 2020,
  Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229, 2019
- Fairness notions: Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2):153–163, 2017,
  Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. In *Proceedings of the 8th Conference on Innovations in Theoretical Computer Science (ITCS)*, 2016,
  Sam Corbett-Davies and Sharad Goel. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv:1808.00023 [cs.CY]*, 2018
- Uniform performance on subpopulations and distributional robustness: John C. Duchi and Hongseok Namkoong. Learning models with uniform performance via distributionally robust optimization. *Annals of Statistics*, 2020,
  John Duchi, Tatsunori Hashimoto, and Hongseok Namkoong. Distributionally robust losses for latent covariate mixtures. *arXiv:2007.13982 [stat.ML]*, 2020

III. Certifiable defenses against adversarial attacks

- Certifiable relaxations: Aditi Raghunathan, Jacob Steinhardt, and Percy Liang. Certified defenses against adversarial examples. In *Proceedings of the Sixth International Conference on Learning Representations*, 2018,
  Eric Wong and Zico Kolter. Provable defenses against adversarial examples via the convex outer adversarial polytope. In *Proceedings of the 35th International Conference on Machine Learning*, pages 5286–5295, 2018
- Distributional robustness perspective: Aman Sinha, Hongseok Namkoong, Riccardo Volpi, and John C. Duchi. Certifiable distributional robustness with principled adversarial training. *arXiv:1710.10571 [stat.ML]*, 2020

IV. Domain adaptation and covariate shift

V. Causal inference

- Semiparametrics: Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018
  Keisuke Hirano, Guido Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003
- Sensitivity analysis: Steve Yadlowsky, Hongseok Namkoong, Sanjay Basu, John Duchi, and Lu Tian. Bounds on the conditional and average treatment effect in the presence of unobserved confounders. *arXiv:1808.09521 [stat.ME]*, 2018

Prerequisites: There are no formal prerequisites, but the class will be fast-paced and will assume a strong background in machine learning, statistics, and optimization. This is a class intended for PhD students conducting research in related fields. Although some materials are of applied interest, this course has significant theoretical content that require mathematical maturity. The ability to read, write, and think rigorously is essential to understanding the material.

Grading and Evaluation: There will be 2-3 problem sets in the class; they will count for 50% of the grade.

Students taking the course for a grade will complete a final project for the course, which will count for 50% of the grade. Students are expected to work on an original research topic related to the content of the course, and at the end of the course the student(s) will present a brief writeup to the course staff detailing their work. Projects will likely turn into publishable work.

In the case that progress on a research project prove difficult (and only when this turns out to be the case), students will have the option to do a pedagogical project. This can take the form of surveying the literature on a particular topic from a critical viewpoint, replicating the empirical work in a paper, or developing exercises from a few papers around topics in the class.

More information about the course project will be posted in the first few weeks of class. The project can be done individually, or in pairs. Students are expected to meet with the instructor during office hours to discuss their project ideas.

References There is no textbook for the class. The following are useful references for different parts of the course.

- Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998
- A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York, 1996
- Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009
- Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: a Nonasymptotic Theory of Independence*. Oxford University Press, 2013

- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004
- Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM and Mathematical Programming Society, 2009
- John C. Duchi. Introductory lectures on stochastic convex optimization. In *The Mathematics of Data*, IAS/Park City Mathematics Series. American Mathematical Society, 2018
- Percy Liang. Statistical learning theory. Lecture Notes for CS 229T, Stanford University, 2016. URL `http://web.stanford.edu/class/cs229t/notes.pdf`. Accessed August 2020
- Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2019
- Guido Imbens and Donald Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015