B9145: Reliable Statistical Learning

Course information and syllabus

Instructor: Hong Namkoong (namkoong@gsb.columbia.edu)
Lectures: Thursdays, 2:00–5:15pm, Geffen 430
Office hours: By appointment
TA: Tiffany Cai (tc3100@columbia.edu)

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.

I. Stochastic optimization methods

II. Generalization bounds: concentration, symmetrization, chaining

III. M-estimation theory: asymptotics

IV. 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. Domain adaptation and covariate shift

II. Certifiable defenses against adversarial attacks

III. Distributional robustness and causal learning

IV. Ethics, fairness, and subpopulation performance

V. Causal inference

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 3 problem sets in the class; they will count for 57% of the grade. You will be required to scribe for a lecture, which will count as 3% of the grade.

Students taking the course for a grade will complete a final project for the course, which will count for 40% 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. Ideally, projects will have a chance to 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