Ethical Statistical Learning
Syllabus

September 8, 2023

1 Course Description

As Data Science (DS) and Machine learning (ML) algorithms are increasingly used in consequential domains such as health care, mortgage applications, college admissions, and bail decisions, there are growing concerns about DS/ML ethics. When and how do the ML algorithms exacerbate the human bias in the data? How can we mitigate such biases and ensure algorithmic fairness? When do ML algorithms reveal sensitive individual information? What protections should be in place for data privacy? Why do deep neural networks make ML algorithms vulnerable to adversaries? How to design adversarially robust algorithms? This course will bring together advanced Ph.D. students in statistics, computer science, and related fields to tackle these and related questions.

2 Course Objective

This course is designed for Ph.D. students in Statistics, Computer Science, Electrical and Computer Engineering, Information Sciences, or related fields. A limited number of Master students and high-level DS undergraduate majors are also welcome. Students will learn to analyze and mitigate privacy loss and algorithmic bias, and enhance adversarially robustness in data analysis. Principal techniques will come from differential privacy, adversarial robustness, and the newly emerging area of algorithmic fairness.

3 Prerequisites

Probability theory, statistical inference, basic machine learning, mathematical analysis, linear algebra, and statistical programming ability in Python or R.

4 Grading

The grading is based on the course project.

5 Tentative Schedule

1. Week 1 (Intro and Overview): Introduction, and basic concepts of privacy and algorithmic fairness
2. Week 2 (Privacy 1): Reconstruction theorem; intro to differential privacy (DP); properties; randomized response
3. Week 3 (Privacy 2): DP Primitives: Laplace and Exponential Mechanisms
4. Week 4 (Privacy 3): DP Primitives: Approximate DP; Gaussian mechanism; Advanced Composition
5. Week 5 (Privacy 4): Statistical learning with DP; Score attacks

6. Week 6 (Privacy 5): Adaptive Data Analysis
7. Week 7 (Fairness 1): Individual and group fairness; testing group fairness; impossibility results
8. Week 8 (Fairness 2): Composition; definition of calibration
9. Week 9 (Fairness 3): Multi-group fairness notions
10. Week 10 (Fairness 4): Fairness in uncertainty quantification
11. Week 11 (Fairness 5): Spurious correlation in statistical learning
12. Week 12 (Large Language Models): New privacy and fairness challenges in the area of LLMs
13. Week 13 (Presentation): Student Presentation