Fairness, Privacy, and Trustworthiness in Machine Learning (Winter 2023)

Course: MAT 280
CRN: 30935
Title: Fairness, Privacy, and Trustworthiness in Machine Learning

Instructor: Thomas Strohmer
Email:"my last name" at math.ucdavis.edu
Office Hours: W: 3:10pm-4pm

Lectures: MW 1:40 - 3:00pm. SOCSCI 90

Course Objective:

As machine learning increasingly pervades more and more aspects of our life, hitherto often ignored questions regarding social responsibility, transparency, and trustworthiness of machine learning and AI emerge as mainstream topics that pose deep and challenging mathematical problems. How do we take fairness and transparency into account while developing machine-learned models and systems? How do we protect the privacy of users when building large-scale, AI-based systems? How can we develop a rigorous understanding of the vulnerabilities inherent to machine learning? Due to their real-world implications, these topics are now at the forefront of ML and AI research. This course will introduce and analyze the mathematical concepts behind fairness, privacy, and transparency of ML. Benefits and shortcomings of existing mathematical tools, metrics, and methods will be investigated and open problems will be discussed.

Prerequisite:

Linear algebra and a basic background in probability as well as basic experience in programming (preferably Matlab) will be required. Some basic knowledge in optimization and machine learning is helpful.

Textbooks:

The topics of this course are very recent and not covered in textbooks yet (at least not from a mathematical viewpoint). Here are a few books that will be useful (and may are available freely as pdf):

- S. Barocas, M. Hardt, and A. Narayanan. Fairness and machine learning Limitations and Opportunities. https://fairmlbook.org
- M. Kearns and A. Roth. The Ethcial Algorithm. The Science of Socially Aware Algorithm Design. Oxford University Press.
- C. Dwork and A. Roth. The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science 9:3-4. https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf
- R. Vershynin. High-Dimensional Probability: An Introduction with Applications in Data Science. https://www.math.uci.edu/~rvershyn

/papers/HDP-book/HDP-book.pdf

- A. Bandeira, A. Singer, T. Strohmer. Mathematics of Data Science. Book draft, downloadable from https://people.math.ethz.ch/~abandeira/BandeiraSingerStrohmer-MDS-draft.pdf
- M. Hardt and B. Recht. Patterns, predictions, and actions. A story about machine learning. https://arxiv.org/pdf/2102.05242

Grading Scheme:

- 50% Homework
- 50% Final Project

Homework:

I will assign homework about every other week. A subset of these problems will be graded. Late homework will not be accepted.

Final Project:

For the Final Project you need to write a report (length about 8 pages) on one of the following topics:

- Describe how some of the methods you learned in this course will be used in your research.
- Find a practical application yourself (not copying from papers/books) using the methods you learned in this course; describe how to use them; describe the importance of that application; what impact would you expect if you are successful?
- A report describing a thorough numerical comparison of existing algorithms related to one of the topics of this couse for a specific application or problem.
- More details about the Final Project will be discussed in class.

UC Davis is committed to educational equity in the academic setting, and in serving a diverse student body. I encourage all students who are interested in learning more about the Student Disability Center (SDC) to contact them directly at sdc.ucdavis.edu, sdc@ucdavis.edu or 530-752-3184. If you are a student who currently receives academic accommodation(s), please submit your SDC Letter of Accommodation to me as soon as possible, ideally within the first two weeks of this course. Also, the campus offers various helpful resources in case the covid-caused conditions really start getting to you. Do not hesitate to make use of them, this is what theses resources are for. See https://sdps.ucdavis.edu/departments/asap/resources/ucd-covid-19