CAS 751

Information-Theoretic Methods in Trustworthy Machine Learning

Fall 2024

Course Outline

Dr. Shahab Asoodeh

McMaster University

Revised: September 4, 2024

Note: This course outline contains important information that may affect your grade. You should retain it and refer to it throughout the semester, as you will be assumed to be familiar with the rules specified in this document.

Instructor

Shahab Asoodeh
Email: asoodehs@mcmaster.ca
Web: https://www.cas.mcmaster.ca/~asoodehs
Office Hours:
Mondays (virtual): 11:30 – 12.30 through Zoom
Meeting ID: 943 0824 3533
Passcode: CAS751

Schedule

Lectures	Wednesdays	11.30am–2.30pm	Shahab Asoodeh	ITB 225
Break	12:45 - 1:00pm			

Course Web Site

The course materials will be posted the following page:

https://www.cas.mcmaster.ca/~asoodehs/cas751.html

It is the student's responsibility to be aware of the information on the course's webpage and to check regularly for announcements.

Mission

The *mission* of the course is to (1) introduce students to social aspects of ML and illustrate what would happen when training ML models with only

accuracy as a benchmark (why do we need *trustworthy* ML?), (2) provide students with a (rather) full depth exposition of differential privacy and its application in training deep models on private and sensitive datasets, and (3) discuss other aspects of trustworthy ML such as fairness and generalization, and how to train ML models with provably fairness and generalization guarantees.

Major Topics

1. Basics of information theory (f-divergence, mutual information, channel capacity, rate distortion, strong data processing inequality)
2. Central and local Differential privacy
3. Deep learning with central differential privacy
4. Statistical efficiency under local differential privacy
5. Fairness
6. (Time permitting) Generalization

Learning Objectives: Postcondition

A *learning objective* for a course is something the student is expected to know and understand or to be able to do. The *precondition* of a course is the set of learning objectives that the student is expected to have achieved before the start of the course. The *postcondition* of a course is the set of learning objectives that the student is expected to have achieved by the end of the course.

Course Precondition

1. Students should know and understand:
    a. Basic calculus such as integration, derivative, and limit
    b. Undergrad-level probability and statistics: random variable, CDF, PDF, Gaussian density, conditional distribution, etc.
    c. Undergrad-level (basic) machine learning: loss functions, risk minimization, gradient descent (will be revisited formally during the course)

Course Postcondition

1. Students should know and understand:
    a. What the umbrella term "trustworthy ML" refers to in terms of social aspects (or lack thereof) of modern data science.

b. Intuitive and mathematical definitions of differential privacy .

c. How to design and analyze privacy-preserving ML algorithms.

d. How to read and write Python codes (TensorFlow and PyTorch) for training private deep models.

e. How to mathematically define "fairness" and "discrimination" in ML.

f. How to make potentially discriminatory algorithms fair.

g. (time permitting) What is generalization and how to measure it.

Required Resources

1. *Course web site*: All course materials will be available on the course website.

Optional Resources

1. Y. Polyanskiy and Y. Wu. "Information Theory: From Coding to Learning," Cambridge University Press, 2022+. [available online]: An *advanced* resource for the information-theoretic background needed for the course. **We'll discuss Chapters 7, 31 (partially), and 33 (partially).**

2. Yihong Wu, "Information-Theoretic methods for High-Dimensional Statistics": An excellent resource for the classical statistics such as minimax and Bayesian estimation problems. **See Sec 1-3, 6-13**

3. John Duchi, "Information Theory and Statistics": Sec 6 provides a detailed (and advanced) exposition of statistical estimation under differential privacy.

Work Plan

There will be lectures, (approximately) 4 sets of assignments, a midterm test, and a final project. Students are expected to attend each lecture and engage by asking questions. With the exception of Friday Sep 4, each lecture will be given on the board. The first lecture, however, will be an informal presentation about the common pitfalls of modern ML/AI and how the materials learned in this course can help address them.

Assignment. There will be (approximately) 4 assignments during the course and together account for 30% of the final grade. Each assignment will be due in two weeks after they are posted to the course website. **Late submission is not accepted.** The submission will have to be in LaTeXwith the format provided later in the course. **Handwritten submission is not accepted.**

Midterm. The midterm exam will be on **Thursday October 23, 2024** during the lecture. A formula sheet (i.e., double-sided) will be permitted for the mid-term exam. (Although the exam will be on what is"understood" in the lectures and not what is memorized!) The midterm accounts for 30% of the final grade.

Final Project. The last component of the course is final project: Each student is expected to pick n research papers on topics related to what is covered in the course (for $n \geq 1$, or preferably $n \geq 2$) and present them in a cohesive and clear way. The deadline for choosing the research topic is **November 13th at 23:59pm**. In addition to the presentation, each student needs to submit a report that summarizes and critiques the paper(s) by a week after the last lecture in the LaTeXformat provided later in the website. The guidelines and a list of recommended projects will be announced later in the course. The final project accounts for 40% of the final grade (equally divided between report and presentation).

Marking Scheme. The course grade will be based on the student's performance on class participation (bonus), assignments, midterm exam, and the final project as follows:

Assignments	30%
Midterm exam	30%
Final project	40%
Total	**100%**
Class participation and engagement bonus	10%

Note that the Instructor reserves the right to adjust the marks for an assignment, midterm test, or final exam by increasing or decreasing every score by a fixed number of points (curving the grades).

Anonymous Comments on the Course/Instructor

If there is anything bothering you in the course and you would like to discuss it with me, please email me and we can chat about it. And if you prefer to let me know the issue **anonymously**, you can contact the Instructor through this Google Form.

End-of-Term Course Evaluation

Near the end of the term, each student will have the opportunity to evaluate the effectiveness of this course. The feedback that is received from the course evaluation is very valuable to the Instructor and will be used to improve the course in subsequent years.

Other Policy Statements

1. Significant study and reading outside of class is required.
2. Student are expected to engage in the lecture by asking questions.
3. If there is a problem with the marking of an assignment, students should immediately discuss the problem with the Instructor. An assignment mark will only be changed if the problem is reported within two weeks of the date that the mark was announced.
4. Email with a source address outside of McMaster University will not be read by the instructional staff.