Privacy-Preserving Machine Learning

Spring/Summer 2023

Instructor: Tamara Bonaci (t.bonaci@)

Office hours:

- TBD
- By appointment

Course website: TBD

Course assignment submission: TBD

Course gradebook: TBD

Course Piazza: TBD

Course Overview:

Driven by recent progress in data science, machine learning and AI, our data is being collected at an ever-increasing pace. Collectors, aggregators, and processors of such data are various private for-profit and non-profit entities, as well as public organizations. While machine learning models, trained on our data, can be beneficial to us personally, as well as to societies at large, they can also lead to a slew of undesirable, negative, and occasionally catastrophic privacy incidents.

We therefore must balance out two conflicting objectives: the need to maximize accuracy, utility and efficiency of the used machine learning models, while at the same time protecting the privacy of the data used for and by those models, as well as the privacy of the models themselves.

Privacy Preserving Machine Learning (PPML) is an important and very active research area, that focuses on this question exactly – how to benefit from machine learning techniques while preserving the privacy of training data and learned models.

In this course we will explore the variety of topics related to privacy preserving machine learning, focusing on theoretical and applied aspects of PPML. We will start by considering statistical and information-theoretic notion of privacy. We will then consider privacy attacks against machine learning models. From there, we will examine a variety of topics focused on preventing and mitigating such privacy attacks, including multi-party secure computation (MPC), differential privacy (DP), federated learning, robust federated learning, and split learning.

Course Prerequisites:

This course assumes a basic familiarity with stochastic mathematics (notion of a random variable and a random process, expectation, variance, independence, Markov process).

Course Goals:

By taking this course, the students will:

- Gain familiarity with statistical and information-theoretic notion of privacy.
- Gain familiarity with privacy attacks against machine learning models.
- Gain familiarity with privacy engineering techniques, including secure multi-party computation (MPC), and differential privacy.
- Gain knowledge of privacy preserving approaches, including federated learning and split learning.
- Gain familiarity with federated learning in adversarial attacks.
- Gain knowledge of state-of-the-art Python libraries and tools used for PPML, such as PySyft.

Course Outcomes:

Upon taking this course, students will:

- Be familiar with the different privacy attacks against machine learning systems.
- Be familiar with privacy engineering techniques, including secure multi-party computation (MPC), and differential privacy, and their applications to machine learning systems.
- Be familiar with privacy preserving approaches, including federated learning and split learning.

Course Logistics

Lectures: lecture slides, relevant reading and additional material will be made available in the following way:

  - Lecture notes and relevant reading material will be posted on Canvas, under the appropriate weekly module.
  - Pre-recorded video material (if available for the current module) will be posted on Canvas, under the appropriate weekly module.
- Every **XYZ PT,** we will meet ***in person in XYZ, and on Zoom, using our recurring Zoom link.*** Our lectures will focus on the outline weekly topics. Additionally, we will use the lecture time to discuss weekly readings. We will do our best to record lectures, and make it available through Canvas + Panopto.

Important note: **if you are unable to attend lectures in person, please reach out to Dr. Bonaci, and I will do my best to accommodate you, and allow you to access lectures remotely.**

Office hours: XYZ from xx-yy PT using UW Zoom meeting. **Office hours will not be recorded**, but may be attended by multiple students at the same time. If you would like to talk to me in private, please send me a note, and we can schedule a different time to meet.

Classroom recordings: This course, or parts of this course, may be recorded for educational purposes. These recordings will be made available only to students enrolled in the course, instructor of record, and any teaching assistants assigned to the course.

Course Progression:

The following is a preliminary class progression covering the 14 weeks of the course. It is subject to changes.

Week 1:

Lecture 1: Course overview. Why Machine Learning Needs Privacy-Preserving Manner?

Week 2:

Lecture 2: Crash Course to Machine Learning

Week 3:

Lecture 3: Machine Learning in Adversarial Setting – Privacy Attacks. Statistical and Information-Theoretic Notion of Privacy.

Week 4:

Lecture 4: Introduction to Secure Multi-Party Computation (MPC).

Week 5:

Lecture 5: MPC and Machine Learning. Introduction to PySyft.

Week 6:

Lecture 6: Introduction to Decentralized Privacy-Preserving Machine Learning Algorithms. Federated Learning.

Week 7:

Lecture 7: Federated Learning II.

Week 8:

Lecture 8: Federated Learning in Adversarial Environment.

Week 9:

Lecture 9: Federated Learning in Adversarial Environment II. Differential privacy and federated learning.

Week 10:

Lecture 10: Introduction to split learning.

Final exam week:

Project presentations.

About the Course:

The course will consist of ***readings and discussion, classroom presentations, labs and a project.***

Readings and Discussion:

Readings and discussions of assigned papers are an important part of this course. The goal of this exercise is to work together on understanding broader implications of the foundational privacy-preserving machine learning research. Additionally, we will also try to keep up with the state-of-the-art PPML research.

Every week, a single research paper will be assigned, and you will be expected to post to the class discussion board about it. Your post should contain something original beyond what others have already posted. You may consider posting:

- A summary of the paper,
- Evaluation of the paper's strengths and weaknesses,
- Open research question related to the topic of the paper, or
- Question you would like to discuss in class.

All posts are due by **11:59pm PT on a XYZ day,** and they will be graded on the scale of 0-2, where:

- 0 means missed or irrelevant post,
- 1 means a relevant post submitted, and
- 2 means a good and interesting post.

Post submitted after the deadline will receive no credit. There will be a total of 11 reading assignments, and we will take 8 best scores when determining your grade.

Labs:

We will have up to four Python-based lab assignments, intended to give you experience with some simple privacy-preserving techniques. The labs will utilize publicly available data sets.

Project:

The final component of this course is a project, and its goal is to give you a deeper understanding of how to think about, and how to solve a real-life problem related to privacy-preserving machine learning.

For the project, you will choose a topic related to privacy-preserving machine learning. You can work on the project either individually, or in groups of two. When working in a group, your end-result should reflect the fact that it is a multi-person effort.

Your work on the project will consists of several milestones:

- Project pitch – due in Week 3,
- Project proposal – due in Week 5,
- Project update – due in Week 10

- Project presentation – due in Week 14
- Final report – due in Week 14

Grading:

Your grade in this course will be based on readings and discussion, homework assignments, and project. The expected grade breakdown is:

- Readings and discussion – 15%
- Labs: 25%
- Project – 60%

Course Material:

There is no required textbook for this course, but some recommended books that you might want to consider include:

- Jin Li, Ping Li, Zheli Liu, Xiaofeng Chen, Tong Li – Privacy Preserving Machine Learning, Springer Briefs on Cyber Security Systems and Networks, Springer 2022
- Muhammad Habib ur Rehman, Mohamed Mdhat Gaber – Federated Learning Systems, Towards Next-Generation AI, Springer, 2021.
- [Research Papers to be added before this syllabus is finalized]

Assignment Turn-In and Late Submission Policy:

Collaboration: In this course, we want you to learn from each other. Therefore, you are allowed (and encouraged!) to talk to your classmates and other students about all course assignments. You may also consult outside reference materials, and the instructor. ***However, all material that you decide to turn in should reflect your own understanding of the subject matter at the time of writing.*** This means that you should write your own posts about the assigned papers, and code up your own lab assignments, and prepare your own course project. If you work with someone else on any assignment, please include their names on the material that you turn in.

Assignment Turn-in: Posts about the assigned papers should be submitted using the Canvas discussion board. All other material (lab and project assignments) should be submitted using course website on Canvas. Please, ***do not use*** email for assignment submissions.

Late Assignment Turn-in: All assignments are due **by 11:59pm PT on the assigned date,** but we understand that you may have to sometimes turn them in late. The grading penalty is 5% of the grade that you would otherwise receive for each day, or part of the day, that you are late. No submissions will be accepted after two weeks.

Checking grades: Grades will be posted to the course gradebook.