Theory of Replicable Machine Learning
EN.601.774

Course Information

Meeting Times: Tue/Thu 3:00 - 4:15pm
Location: Hodson 316
Instructor: Jess Sorrell (jess@jhu.edu)
TA: Iliana Maifield-Carucci (imaifel1@jhu.edu)
Office Hours: Jess - Friday 11:00am EST or by appointment.
Iliana - TBD

Course Description

Replicability is vital to ensuring scientific conclusions are reliable, but failures of replicability have been a major issue in nearly all scientific areas of study, and machine learning is no exception. In this course, we will study replicability as a property of learning and other statistical algorithms, developing a theory of replicable learning. We will cover recent formalizations of replicability and their relationships to other common stability notions such as differential privacy and adaptive generalization. We will survey replicable algorithms for fundamental learning tasks, and discuss the limitations of replicable algorithms. If time permits, we will discuss replicability in other settings, such as reinforcement learning and clustering, or other useful and related stability notions such as list replicability and global stability.

Course Prerequisites

EN.601.433/633 Intro Algorithms. In this course, we will prove and contextualize theoretical results related to the theory of machine learning, and mathematical maturity will be assumed.

Planned Coursework and Grading Breakdown

The following plan is tentative and subject to change.

Homework (5%). There will be one homework assignment for this course, worth 5% of the final grade, assigned in the first week. This assignment is primarily to assess students' familiarity with relevant material and so will be graded for completion only.

Course Project (95%). This course includes a semester long project with one of the following aims:

- answering a novel question related to the theory of replicable learning
- improving existing results related to course material
- summarizing a related area of recent research not covered in lecture

- empirically evaluating existing theoretical results to establish practical limitations

Students are encouraged to work in groups of 3 or 4. If you prefer to work alone or in a larger group, please contact the instructor for approval first.

- Project deliverable 1 (written proposal, week 4): 10%
- Project deliverable 2 (preliminary results write up, week 9): 25%
- Project milestone 3 (presentation, last week of lecture): 20%
- Project deliverable 4 (paper, finals week): 40%

Course Resources

The lectures for this course will at times cover material similar to that covered in a related course on adaptive data analysis taught by Aaron Roth (Penn) and Adam Smith (BU). Their course notes are available here: https://adaptivedataanalysis.com/about/

There is no required textbook for this course, but there are a number of papers and texts from which the material for this course is drawn. Below is a list of these sources, organized roughly by topic.

1. Introduction: motivation, formal definition of replicability, generalization guarantees
   - Reproducibility in Learning. https://arxiv.org/abs/2201.08430
2. Replicable algorithms: statistical queries, heavy hitters, amplification, learning finite hypothesis classes
   - Reproducibility in Learning. https://arxiv.org/abs/2201.08430
   - Stability is Stable: Connections between Replicability, Privacy, and Adaptive Generalization https://arxiv.org/abs/2303.12921
3. Lower bounds: statistical queries, mean estimation for bounded covariance distributions
   - Reproducibility in Learning. https://arxiv.org/abs/2201.08430
   - Stability is Stable: Connections between Replicability, Privacy, and Adaptive Generalization https://arxiv.org/abs/2303.12921
   - Replicability in High Dimensional Statistics https://arxiv.org/abs/2406.02628
4. Connections between replicability and related stability notions: differential privacy, adaptive generalization, max information
   - Stability is Stable: Connections between Replicability, Privacy, and Adaptive Generalization https://arxiv.org/abs/2303.12921
   - Generalization in Adaptive Data Analysis and Holdout Reuse https://arxiv.org/abs/1506.02629
   - The Algorithmic Foundations of Differential Privacy https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf
   - Max-Information, Differential Privacy, and Post-Selection Hypothesis Testing https://arxiv.org/abs/1604.03924
5. Correlated sampling
   - Stability is Stable: Connections between Replicability, Privacy, and Adaptive Generalization https://arxiv.org/abs/2303.12921
   - User-Level Privacy via Correlated Sampling https://arxiv.org/abs/2110.11208
6. Computational separations between replicability and privacy

- Stability is Stable: Connections between Replicability, Privacy, and Adaptive Generalization https://arxiv.org/abs/2303.12921

7. Replicability in reinforcement learning
   - Replicable Reinforcement Learning https://arxiv.org/abs/2305.15284
   - Replicability in Reinforcement Learning https://arxiv.org/abs/2305.19562
8. Replicable clustering
   - Replicable Clustering https://arxiv.org/abs/2302.10359
9. List replicability and global stability
   - Replicability and Stability in Learning https://arxiv.org/abs/2304.03757

Academic Integrity

Students are expected to adhere to the JHU policy on academic honesty. Instances of plagiarism or cheating will result in disciplinary action.

Use of Generative AI: Below is a list of specific cases in which generative AI may explicitly be used or not be used. If you would like to use an AI tool in this course for some case that is not covered below, please email the instructor for clarification. If it's reasonable and does not interfere with the learning objectives for the course, it is likely to be approved.

Use of generative models (e.g. chatGPT or Claude) is *permitted* in the following cases:

- Assisting with generating ideas and scoping for the final project
- Generating boilerplate code to be used for presenting the results of empirical evaluations (e.g. generating plots)
- Rephrasing student-generated text to improve presentation.

Use of generative models is *prohibited* in the following cases:

- Assisting with proving theorems (I'm not ready for AI to take my job)
- Generating an implementation of an algorithm that is being evaluated for the final project

Please also be aware that I am unlikely to approve a project that I suspect a generative model could receive a C or better on without significant human involvement.

Accommodations

Any student with a disability who may need accommodations in this class must obtain an accommodation letter from Student Disability Services, 385 Garland, (410) 516–4720, studentdisabilityservices@jhu.edu.