Syllabus

1 Logistics

- Instructor: Tselil Schramm (tselil@stanford.edu).
- Course website: www.tselilschramm.org/mltheory/stats214-fall23.html
- Teaching Assistants: Yash Nair and Asher Spector
- Lectures: Mondays & Wednesdays, 13:30-14:50, 200-205.
- Office Hours: listed on Canvas.

The best way to contact me is by email. Please be sure to include "STATS214" in the subject line.

Prerequisites: Students will need a background in calculus (MATH19/MATH20/MATH21), linear algebra (MATH51 or CS205), probability theory (STATS116, CS109, or MATH151), and machine learning (STATS229/CS229 or STATS315A). If you have not taken the courses listed but believe that you have sufficient background to take the course, please contact me.

2 Overview

The goal of this course is to give students a mathematical foundation for understanding machine learning. We will explore the following questions: How do machine learning algorithms work? How can we quantify their success? How much data do we need in order to successfully learn? This is a theoretical, proof-based course, and our focus will be on algorithms with rigorous guarantees wherever they are to be had.

Course Description (as it appears in the course catalogue): How do we use mathematical thinking to design better machine learning methods? This course focuses on developing mathematical tools for answering these questions. This course will cover fundamental concepts and principled algorithms in machine learning, particularly those that are related to modern large-scale non-linear models. The topics include concentration inequalities, generalization bounds via uniform convergence, non-convex optimization, implicit regularization effect in deep learning, and unsupervised learning and domain adaptations

Topics We will cover the following topics, roughly partitioned into units:

1. **Intro to supervised learning**. The theory of learning from labeled examples. Empirical risk minimization, asymptotic analysis.
2. **Generalization theory**. When can we guarantee that what we learn from our training data will generalize to new observations? Uniform convergence, Rademacher complexity, VC dimension.
3. **Optimization**. The framework of empirical risk minimization requires us to *optimize* a loss function. We'll briefly introduce the basics of convex optimization, then see a few topics selected to give an overview of these methods, including: stochastic gradient descent, online learning, regularization, non-convex optimization, and kernel methods.

4. **Unsupervised learning**. The theory of learning structure of unlabeled data. We'll focus on clustering unlabeled data, covering k-means clustering, spectral clustering in graphs, the method of moments in the context of Gaussian mixtures, and (time permitting) tensor decomposition.
5. **Special Topic: Deep learning**. Deep learning is revolutionizing machine learning, but practice remains far ahead of theory. We'll give some highlights of the nascent theory of deep learning, and explore some of the challenges in developing the theory more broadly.

The course schedule may be found on the website.

3 Materials & Resources

Course website. The course website is www.tselilschramm.org/mltheory/stats214-fall23.html. There you will find the course schedule, a list of texts and resources, and additional relevant readings.

Cavnas. Homeworks, solution keys, and lecture notes will be made available on Canvas. We'll also use Canvas to post Zoom links for remote lectures, if necessary.

Gradescope. You will turn in your homework solutions on Gradescope. You will receive a Gradescope invitation in the first week of the quarter.

Campuswire. We will use Campuswire as an online class forum, where you may ask questions and discuss with your fellow students. You will receive a Campuswire invitation in the first week of the quarter.

4 Coursework & Evaluation

Homework (50%) We will have four homework assignments in total. The first homework assignment (homework 0) is a warm-up assignment, and will be worth 10 points. Each of the three subsequent homeworks is worth 30 points. We will drop your lowest homework score out of homeworks 1,2,3 (not homework 0). Collaboration is encouraged, but you are responsible for your own understanding and you must independently write your own solutions.

Problem Set Reflections (optional) Within a week of submitting your problem set, you may submit a brief discussion of whether/how your solutions differ from the answer key (an example will be posted on Canvas). If you demonstrate an understanding of your mistake, you may earn back up to 50% of any deducted points (for problems on which you made a good faith attempt the first time around).

Course Project: research proposal (50%) For the course project, students will read 2 or more papers on a topic of their choice, then write a research proposal that identifies an interesting, concrete follow-up question for future research. Your proposal should include a brief summary of the papers read, a discussion of their significance, strengths, and weaknesses, as well as an explanation of why the proposed follow-up work is interesting, promising, and tractable. An outstanding proposal might also contain some preliminary research (a theoretical approach or supporting experiments).

The expectation is that your proposal contain original insights, connections, and critiques—do not simply regurgitate what is written in the papers verbatim. The goal is to prepare you for research in machine learning theory: this includes reading challenging papers critically, and identifying interesting and promising open questions. A list of suggested papers will be provided. You may work in teams of up to 3. In total, the proposal should be no more than 5 pages long. More details will be provided in the early weeks of the quarter.

Course grades As a rough guideline, students whose grade falls in the range

- $85\% - 100\%$ will receive between A- and A+.
- $70\% - 85 - \varepsilon\%$ will receive between B- and B+.
- $55\% - 70 - \varepsilon\%$ will receive between C- and C+.
- $45\% - 55 - \varepsilon\%$ will receive between D- and D+.
- $0\% - 45\%$ will receive a NP.

These are not the final percentages; any differences from this scale will be in students' favor (that is, your grade might be higher than the percentage here indicates).

To receive credit for the class with CR/NC grading, you must obtain a minimum of C-.

Campuswire participation (up to 5% bonus) Students may receive up to 5% extra credit for meaningful participation on Campuswire. Meaningful participation constitutes the asking or answering of questions in a way that has a positive impact on the learning of other students in the class.

5 Policies

The Honor Code. It is expected that you and I will follow Stanford's Honor Code in all matters relating to this online course. You are encouraged to meet and exchange ideas with your classmates while studying and working on homework assignments, but you are individually responsible for your own work and for understanding the material. You are not permitted to copy or otherwise reference another student's homework or computer code.

Late Work Policy. Late work will not be accepted. Instead, your lowest homework score (out of homeworks 1,2,3) will be dropped.

Accommodations. I am happy to provide accommodations, understanding that they may be necessary for student success. Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations.

Course Privacy Statement. As noted in the University's policy on recording and broadcasting courses, students may not audio or video record class meetings without permission from the instructor (and guest speakers, when applicable).