Syllabus

Theoretical Statistics and Machine Learning

SDS 391P.6, Spring 2026

Instructor: Pratik Patil, pratikpatil@utexas.edu

Lectures: Tuesdays and Thursdays, 2-3.30pm, ECJ 1.306

Office hours: Tuesdays and Thursdays, 3.30-5pm, WEL 5.216H

Website: https://pratikpatil.io/teaching/sds391p6-s26

Canvas: https://utexas.instructure.com/courses/1444652

Overview

This course provides a guided tour of modern methods in statistics and machine learning, with a primary focus on high-dimensional and non-parametric settings. We will explore both classic foundations and modern techniques that allow us to understand these methods through a rigorous theoretical lens. The main emphasis is *non-asymptotic* analysis. Unlike traditional statistical theory, which often relies on limit theorems as the sample size n tends to infinity, we will focus on deriving sharp guarantees that hold for any fixed (sample size n, feature size d) or when both grow together. This perspective is essential for understanding contemporary data science practice, where the number of features often rivals or exceeds the number of observations.

Outline

In slightly more detail, the topics that we will cover will likely cover are divided into two main parts:

1. Theoretical tools and techniques:
   - Concentration inequalities: Understanding how random variables deviate from their expectations (e.g., Hoeffding, Chernoff, and Bernstein bounds).
   - Uniform laws of large numbers: Moving beyond point-wise convergence to understand the behavior of empirical processes over classes of functions.
   - Metric entropy and chaining: Quantifying the "size" of infinite or large hypothesis spaces through covering and packing numbers.
   - Information-theoretic lower bounds: Establishing the fundamental limits of estimation using tools like Fano's and Le Cam's lemmas to determine what is statistically possible.
2. High-dimensional and non-parametric applications:
   - Structured estimation: Mean and covariance estimation in high dimensions and the role of sparse estimation.
   - Dimension reduction: Theoretical guarantees for principal component analysis in the high-dimensional regime.
   - Learning theory: Analyzing empirical risk minimization and understanding generalization error.
   - Function estimation: Exploring non-parametric regression, where the underlying model lives in an infinite-dimensional space.

(The schedule of topics, along with all the lectures materials, can be found on the course website.)

Learning objectives

By the end of this course, students will be equipped to read current research papers in theoretical statistics and machine learning, derive their own non-asymptotic bounds, and understand the trade-offs between sample complexity, dimensionality, and structural assumptions like sparsity.

Prerequisites

The course is targeted at graduate students in statistics. It is also potentially useful for graduate students in computer science, electrical engineering, and econometrics. As such, we will assume a strong undergraduate background in basic aspects of mathematics (in particular, linear algebra and real analysis) and statistics (in particular, probability theory and mathematical statistics).

(Various resources to review and refresh the prerequisite material will be provided on the course website.)

Resources

The course is designed to be self-contained and will often provide references for further details on various topics covered. The following references are wonderful resources for the material covered in this course:

- High-dimensional probability, Vershynin
- High-dimensional statistics, Wainwright
- Concentration inequalities, Boucheron, Lugosi, Massart
- Probability in high dimensions, van Handel
- Probability in high dimensions, Tropp
- An introduction to matrix concentration inequalities, Tropp
- Topics in random matrix theory, Tao
- Empirical processes in M-estimation, van de Geer
- Introduction to nonparametric estimation, Tsybakov
- Mathematical foundations of infinite-dimensional statistical models, Gine, Nickl

Evaluation

Evaluation will be based on five homeworks, one midterm, and one final exam. The grading breakdown is as follows (each homework is worth an equal amount):

- Homeworks: 50%
- Midterm: 20%
- Final: 30%

Class participation is highly encouraged. We will reserve 10% bonus points for class participation.

(Details on the dates for the homework and exams will be provided on the course website.)

Homework

The homeworks are structured to give you experience in primarily written mathematical exercises, and to a lesser extent, computational exercises for additional intuition. Homework will be assigned biweekly and due via Canvas. Details submission instructions will be provided in the homework.

In general, late homework will not be accepted. In the case you are busy preparing for an important deadline, you must give us at least -3 days of notice if you are requesting an extension and we can give upto $+3$ days of extension. In the case of an emergency, no notice is needed, and we can work with you to give you a reasonable extension. But please use this option only in the case of true emergencies.

We encourage discussion and collaboration on homework assignments with other students. However, such collaboration must be clearly acknowledged by listing the names of students with whom you collaborated. We will assume that you are taking full responsibility in terms of writing your own solutions and implementing your own code.

Accommodations

We are committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let us know if you experience any barriers to learning so we can work with you to ensure you have an equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations, please contact Disability and Access (D&A): `https://disability.utexas.edu/`. If you are already registered with D&A, please deliver your Accommodation Letter to us as early as possible in the semester so we can discuss your approved accommodations and needs in this course.

Take care of yourself

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus, including:

- Sanger Learning Center: `https://undergraduates.utexas.edu/tutoring-academic-assistance/sanger-learning-center`
- Healthyhorns: `https://healthyhorns.utexas.edu`
  - University Health Services (UHS): `https://www.healthyhorns.utexas.edu/uhs/`
  - Longhorn Wellness Center (LWC): `https://healthyhorns.utexas.edu/lwc/`
  - Counseling and Mental Health Center (CMHC): `https://healthyhorns.utexas.edu/cmhc/`
- Behavior Concerns Advice Line (BCAL): `https://bcal.utexas.edu/`

We are here to help you learn. If you have questions about this or your course, please let us know!