Syllabus: Statistics 314 Advanced Statistical Theory

Instructor:	Yuting Wei, Sequoia 202; `ytwei AT stanford`
Lecture:	MW 1:30-2:50pm; 200-217
T.A.:	Qian Zhao; `qzhao1 AT stanford`
Office hours:	Yuting Wei: W 3-4pm (or by appointment), Location: Sequoia 202
	Qian Zhao: Th 8-10am: Sequoia Hall Rm 207 (Bowker)
Course website:	Sp19-STATS-314A-01 on *Canvas*

Synopsis: This is a Ph.D. topic class in theoretical statistics. We will cover a selection of topics in mathematical statistics and machine learning theory, with a focus on high-dimensional and non-parametric statistical models. One of the main goals is to provide students with some background to understand the ongoing statistical literature, which focus on dealing with data set with dimension, of same order or even larger than the sample size. We also hope to provide motivated students with some mathematical tools for working on statistical theory.

Tentative list of topics, some will be assigned as reading materials:

1. Concentration of measure
2. Uniform law of large numbers
3. Metric entropy and comparison inequalities
4. Sparse linear models and compressed sensing
5. Principal component analysis in high dimension
6. Matrix estimation and matrix concentration
7. Reproducing kernel Hilbert spaces
8. Nonparametric kernel estimation
9. Nonparametric least squares
10. Dimension reduction via random projection
11. Shallow neural networks

Main textbook: *High-dimensional statistics: A non-asymptotic viewpoint*, Cambridge University Press, 2019, by M. Wainwright.

Additional references:

- *High-Dimensional Probability, An Introduction with Applications in Data Science*, Cambridge University Press, 2018, by R. Vershynin.
- *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman & Hall, 2015, by T. Hastie, R. Tibshirani and M. Wainwright.
- *Introduction to Nonparametric Estimation*, Springer Series in Statistics, 2009, by A.B. Tsybakov.
- *Weak Convergence and Empirical Process: With Applications to Statistics*, Springer-Verlag, 1996, by Van der Vaart, A. and J. A. Wellner.
- *Empirical Processes in M-estimation*, Cambridge university press, 2000, by S. Van de Geer.

Course grading: max{40% homework + 60% Project, Project}

Homework: There will be two assignments. A hard copy of your homework must be turned in on Wednesday class. Please use Latex to typset your homework.

Course project: this project can either be a literature review or include original research:

- Literature review. We will provide a list of related papers that are not covered in the lectures, and the literature review should involve in-depth summaries and exposition of one of these papers.
- Original research. It can be either theoretical or experimental, with the approval from the instructor. If you choose this option, you can do it either individually or in groups of two.

Thress timestamps for the course project:

- **Proposal** (May 8th). Submit a short report stating the papers you plan to survey or the research problems that you plan to work on. Describe why they are important or interesting, and provide some appropriate references. If you elect to do original research, you are encouraged to connect this project with your current research (but is still related to our course content). Please do not propose an overly ambitious project. You will receive feedback from the instructor.
- **In-class presentation** (June 3/5th). Prepare an oral presentation with slides (the exact time will depend on the number of projects in the class). Focus on high-level ideas, and leave most technical details to your report.
- **A written report** (June 12th). You are expected to submit a final project report summarizing your findings / contributions.

Canvas folders: The folder architecture is fairly self-explanatory.
`Lectures` contains lecture notes / slides.
`Homework` contains homework assignments.
`Papers-for-review` contains a list of papers that is of interest. You have the freedom to select a paper of your own interest, as long as it is related to the topics of this course.