MATH 547: MATHEMATICAL FOUNDATIONS OF *STATISTICAL LEARNING THEORY*
FALL 2022

INSTRUCTOR: STANISLAV MINSKER, KAP 406E
EMAIL: MINSKER@USC.EDU
TEACHING ASSISTANT: TBA

CLASSROOM: CPA 210, TIME: MWF 1 – 1:50PM
OFFICE HOURS: MONDAY 9:30-11AM, OR BY APPOINTMENT.

General Information:

This course provides an introduction to the mathematical foundations of *Statistical Learning Theory* as well as select topics in high-dimensional statistics. How do modern high-dimensional statistics and learning theory differ from the classical statistical techniques? As Leo Breiman wrote in 2001 [1], *"There is an old saying 'If all a man has is a hammer, then every problem looks like a nail.' The trouble for statisticians is that recently some of the problems have stopped looking like nails."*

Statistical learning framework often does not assume that the data we observe strictly follows the underlying model (e.g., Gaussian distribution). Instead, (quoting L. Breiman), *"The approach is that nature produces data in a black box whose insides are complex, mysterious, and, at least, partly unknowable. What is observed is a set of x's that go in and a subsequent set of y's that come out. The problem is to find an algorithm $f(x)$ such that for future x in a test set, $f(x)$ will be a good predictor of y."* Throughout the course, we will try to understand when the approach of statistical learning is advantageous, and what are its limitations. We will introduce the underlying basic principles and mathematical tools, and apply them to the design and analysis of classical algorithms.

One of the goals of the course is to prepare participants for independent research, and to help them navigate and understand the scientific literature devoted to the subject of statistical learning.

Please note: the course material will be communicated via the (black/white)board, and the lectures will not be recorded. However, the notes summarizing the material will be provided.

Prerequisites:

Working knowledge (graduate or advanced undergraduate level) of Probability Theory (Math 407/505a, 507a recommended), Real Analysis (Math 425a/b, Math 525a recommended) and Linear Algebra (Math 471). Understanding of the basic principles of Mathematical Statistics (Math 408 or Math 541a).

(Approximate) list of covered topics:

- Introduction to binary classification: plug-in method, curse of dimensionality, and empirical risk minimization.
- Linear separators, kernel trick, and Reproducing Kernel Hilbert spaces.

[1]Leo Breiman (2001). "Statistical Modeling: The Two Cultures", *Statistical Science*, Vol. 16, No. 3.

- Voting algorithms (AdaBoost), Support Vector machines: derivation from the basic principles (see reference (8)).
- Introduction to the theory of Empirical Processes: symmetrization, comparison inequalities (Talagrand's contraction principle), concentration of measure, sub-Gaussian processes, generic chaining and Dudley's entropy integral (based on references (1) and (4)).
- Vapnik-Chervonenkis combinatorics and applications to Statistical Learning (based on references (1) and (4)); generalization error bounds.
- Sparse recovery problems: $\| \cdot \|_1$-norm, high dimensional convex bodies and their sections, restricted isometries. Applications to compressed sensing and LASSO (based on references (3) and (6)).
- Matrix recovery problems: matrix concentration inequalities and applications to matrix completion. If time permits, we will discuss additional topics such as Matrix Concentration inequalities.

Grading:

- (60%) Homework assignments. The problems are intended to be challenging, therefore collaboration and discussions are allowed; however, you **must** write down your own solutions. You <u>are not allowed</u> to search for assistance/solutions online but are encouraged to ask the instructor for hints if you get stuck (I will not deduct points for this).
- (40%) Final project/presentation. Most projects are individual but I will allow up to 3 people (for groups consisting of 2 or 3 people, please get my permission and describe the role of each person involved in the project). Possible topics will be offered by the instructor, but students are welcome to make their own suggestions. You will be expected to read paper(s) on a particular topic, and/or implement algorithms involved. Result of the project will be communicated via a short presentation to the instructor, and a report (written in LaTeX) that summarizes results of the project.
- Late submissions of homework assignments and projects will not be accepted.
- Please see the registration calendar for additional information, including the last day to drop the course: https://classes.usc.edu/term-20223/calendar/.

Grades will be computed on the following scale: $[90\%, 100\%] = A$, $[87\%, 90\%) = A-$, $[85\%, 87\%) = B+$, $[81\%, 85\%) = B$, $[79\%, 81\%) = B-$, $[77\%, 79\%) = C+$, $[71\%, 77\%) = C$, $[0\%, 72\%) = C-$.

Reminder: C is a minimum passing for graduate credit (except in courses designated by a school or department to have a higher minimum standard for passing), C minus is a failing grade for graduate credit.

Scribing:

Everyone is expected to participate in creating the lecture notes. These notes will be typeset in LaTeXand will include the summary of the material covered during the class, written in a way that someone who did not attend the class will understand the main ideas. It is good practice for communicating your future research; a template and a draft will be provided.

Books and useful references:

Course material does not follow a single book, hence there is no mandatory textbook requirement. Useful references include

(1) "Mathematical Foundations of Infinite-Dimensional Statistical Models" by E. Giné and R. Nickl (if you want to get one book, get this one as it is well-written, and contains many important recent results).

(2) "High-Dimensional Probability: An Introduction with Applications in Data Science" by R. Vershynin, available for free at
http://www-personal.umich.edu/ romanv/papers/HDP-book/HDP-book.pdf
(3) "Estimation in High Dimensions: a Geometric Perspective" by R. Vershynin. Available at http://arxiv.org/pdf/1405.5103.pdf.
(4) "Weak Convergence and Empirical Processes" by A. van der Vaart and Jon Wellner.
(5) "Probability in Banach Spaces" by M. Ledoux and M. Talagrand.
(6) "Statistics for High-Dimensional Data: Methods, Theory and Applications" by P. Bühlmann and S. van de Geer.
(7) "The Elements of Statistical Learning" by T. Hastie. R. Tibshirani and Jerome Friedman. Authors generously made it available online:
http://statweb.stanford.edu/ tibs/ElemStatLearn/.
(8) "Understanding Machine Learning: From Theory to Algorithms" by Shai Shalev-Shwartz and Shai Ben-David, available online at
http://www.cs.huji.ac.il/ shais/UnderstandingMachineLearning/
This book is very well-written and easy to read.

Additional references, including papers, will be provided whenever necessary.

Important dates:

- September 9: last day to drop a class without a mark of "W," except for Monday-only classes, and receive a refund for Session 001; last day to change enrollment option to Pass/No Pass or audit for Session 001;
- October 7: last day to drop a course without a mark of "W" on the transcript. Mark of "W" will still appear on student record and STARS report and tuition charges still apply; last day to change between letter grade or Pass/No pass in a letter graded course for Session 001;
- Fall recess: October 13-14, Thanksgiving break: November 23-27.

Students Requiring Special Accommodation:

I would be happy to discuss any special accommodations at the beginning of the course. Any student requesting academic accommodations based on special needs is required to register with OSAS. A letter of verification for approved accommodations can be obtained from OSAS as well. Please be sure the letter is delivered to me as early in the semester as possible. OSAS is located in GFS 120 and is open 8:30 a.m. till 5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776 and the email is DSPFrontDesk@usc.edu.

Academic Integrity:

General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that **individual work will be submitted unless otherwise allowed by an instructor**, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by the principles. In particular, plagiarism - presenting someone else's ideas and work as your own, either verbatim or recast in your own words - is a serious academic offense with equally serious consequences.

The Student Guidebook contains the Student Conduct Code (SCampus, Part B: https://policy.usc.edu/scampus/). Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: https://sjacs.usc.edu/students/academic-integrity/.