MATH 547: MATHEMATICAL FOUNDATIONS OF *STATISTICAL LEARNING THEORY*

INSTRUCTOR: STANISLAV MINSKER, KAP 406E
EMAIL: MINSKER@USC.EDU
TEACHING ASSISTANT: TBA
CLASSROOM: KAP 147, TIME: MWF 10 – 10:50AM
OFFICE HOURS: MONDAY 3-4PM, WEDNESDAY 4-5PM, OR BY APPOINTMENT.

General Information:

This course provides an introduction to the mathematical foundations of *Statistical Learning Theory*. 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."*

One of the goals of the course is to prepare participants for independent research, and to help them navigate and understand publications related to the course material and the area of statistical learning in general.

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).

(Approximate) list of covered topics:

- Binary classification: plug-in method, curse of dimensionality, and empirical risk minimization.
- Linear separators, kernel trick, and Reproducing Kernel Hilbert spaces.
- 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, Dudley's entropy integral (based on references (1) and (4)).
- Vapnik-Chervonenkis combinatorics and applications to Statistical Learning (based on references (1) and (4)).

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

- 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 completion".
- Application of developed techniques to generalization error bounds. If time permits, we will discuss additional topics such as Matrix Concentration inequalities.

Grading:

Course grades will be based on

- (55%) Homework assignments. As some problems are challenging, collaboration is encouraged, however, you **must** write down your own solutions.
- (5%) Scribing (2 lectures each): the scribe notes should be the summary of the class material, written so that someone who did not attend the class will understand the material. It is good practice for communicating your future results with others; the template will be provided.
- (10%) Attendance, checked at random times during the semester (excluding religious holidays).
- (30%) Final project/presentation. Each project can involve up to 3 people (for groups consisting of 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 either be a short presentation made in class, or a written report 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-20153/calendar/.

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.

Students Requiring Special Accommodation:

Any student requesting academic accommodations based on special needs is required to register with DSP each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to the instructor as early in the semester as possible. DSP 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.

Academic Integrity:

USC seeks to maintain an optimal learning environment. 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 these principles. The Student Guidebook contains the Student Conduct Code in Section 11.00, while the recommended sanctions are located in Appendix A: http://www.usc.edu/dept/publications/SCAMPUS/gov/. 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: http://www.usc.edu/student-affairs/SJACS/.