Math 547: Methods of Statistical Inference
Fall 2015

Instructor: Stas Minsker, KAP 406E
email: minsker@usc.edu
Classroom: KAP 150
Office hours: Tuesday 10am-noon

General Information

This course will provide an introduction to the mathematical foundations of *Statistical Learning Theory*. How is modern high-dimensional statistics and learning theory different 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."*

Prerequisites

Working knowledge (graduate or advanced undergraduate level) of Probability Theory, Real Analysis and Linear Algebra.

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.
- Introduction to the theory of Empirical Processes: symmetrization, comparison inequalities (Talagrand's contraction principle), concentration of measure, subgaussian processes, Dudley's entropy integral.
- Vapnik-Chervonenkis combinatorics and applications to Statistical Learning.
- Sparse recovery problems: $\|\cdot\|_1$-norm, high dimensional convex bodies and their sections, restricted isometries. Applications to compressed sensing and LASSO.

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

- Application of developed techniques to generalization error bounds. If time permits, we will discuss additional topics such as Matrix Concentration inequalities, Low Rank Matrix recovery and Active Learning.

Grading

Course grades will be based on

- (50%) homework assignments.
- (50%) final project/presentation. Each project can involve up to 3 people. Possible topics will be offered by the instructor, but students are welcome to make their own suggestions.
- 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. "Weak Convergence and Empirical Processes" by A. van der Vaart and Jon Wellner.
2. "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/.
3. "A Probabilistic Theory of Pattern Recognition" by L. Devroye, L. Györfi and G. Lugosi.
4. "Statistics for High-Dimensional Data: Methods, Theory and Applications" by P. Bühlmann and S. van de Geer.
5. "Estimation in High Dimensions: a Geometric Perspective" by R. Vershynin. Available at http://arxiv.org/pdf/1405.5103.pdf.
6. "Probability in Banach Spaces" by M. Ledoux and M. Talagrand.

Additional references 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 STU 301 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 ones own academic work from misuse by others as well as to avoid using anothers work as ones 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/.