STAT-S 782 -Topics: Statistical Learning Theory

Instructor: Daniel McDonald
Office: BH 669
Phone: 812-855-7828
email: dajmcdon@indiana.edu

Office Hours: TBA

Course Web Page: https://github.com/stats-782fa2017

Slack: https://stat-s782fa2017.slack.com/

Lectures: TR 9:30–10:45, BH 247

Text (recommended): Boucheron et al. (2013); Boyd and Vandenberghe (2004); Tsybakov (2009)

Prerequisite: A thorough understanding of statistics and probability at the graduate level (equivalent to STAT-S721–722) or permission of instructor.

Course Objective

Statistical learning theory is a burgeoning research field at the intersection of probability, statistics, computer science, and optimization that studies the performance of computer algorithms for making predictions on the basis of training data.

The following topics will be covered: basics of statistical decision theory; concentration inequalities; supervised and unsupervised learning; empirical risk minimization; complexity-regularized estimation; generalization bounds for learning algorithms; VC dimension and Rademacher complexities; minimax lower bounds; online learning and optimization.

Along with the general theory, we will discuss a number of applications of statistical learning theory to signal processing, information theory, and adaptive control.

Lectures

Class time will consist of a combination of lecture, discussion, questions and answers, and problem solving. You are strongly encouraged to attend lectures on a regular basis.

Textbook

The recommended textbooks each address a portion of the material we will cover in class. We will begin with convexity and material from (Boyd and Vandenberghe, 2004). This will be used to examine concentration inequalities and uniform convergence as in (Boucheron et al., 2013). Finally, we will develop complementary lower bounds as discussed in (Tsybakov, 2009). Additional material will come from the following sources as well as other places: Devroye et al. (2013); van de Geer (2000); van der Vaart and Wellner (1996); Vapnik (1998).

Grading

10% : Homework
15% : Scribe
25% : Paper presentations
50% : Project

Homework

There will be 3–4 homework assignments during the semester to practice (essentially 1 for each module). Late homework is unacceptable unless special arrangements are made.

Homeworks and their solutions will be available on the course website. Please submit legible, stapled paper copies of your homework. I will not accept electronic copies unless you have cleared it with me in advance.

I encourage you to discuss assignments with other students. The best way to work with others on homework is to do as much as you can on your own before discussing the problems. While I encourage you to work together, the written solutions to homework problems must be your own and not copied from someone else. You must give credit to all collaborators on your assignments by listing the names of those collaborators.

Scribe duties

Every lecture will have a designated scribe. The duties of the scribe are to take careful notes during class and then typset the notes in LaTeX. I will provide a template for the notes with instructions. The typset notes will then be posted on the course website (following my edits) so that everyone may access them. The notes file will be due to me within one week of the lecture. The goal of the scribe is not only to take thorough notes but also to supplement them as needed with appropriate references or examples that they deem useful.

Paper presentations

I have chosen a number of papers from recent ML conferences which involve techniques we will cover in class. Each student will choose one to present during class for 20 minutes. We will discuss 1 paper per week and I will present the remainder. The goal here is to familiarize yourself with the sorts of topics and techniques which lead to a good conference paper. The goal is *not* to understand every step of every proof.

Project

The following is a description of the minimal project necessary to do well in this course. As graduate students, your goal is to publish papers and pad your CV, so I encourage you to think of this project as an opportunity to create something toward that goal while also satisfying course requirements. Here are the rules:

1. You may work by yourself or in teams of two.
2. The goals are (i) to use methods you have learned in class or, if you wish, to develop a new method and (ii) present a theoretical analysis of the methods.
3. You will provide: (i) a proposal, (ii) a progress report and (iii) and final report.
4. The reports should be well-written.

Proposal. A one page proposal is due Thursday, September 13. It should contain the following information: (1) project title, (2) team members, (3) description of the data (if used), (4) precise description of the question you are trying to answer, (5) preliminary plan for analysis, (6) reading list. (Papers you will need to read).

Progress Report. Due Thursday, November 15. Three pages. Include: (i) a high quality introduction, (ii) what have you done so far, (iii) what remains to be done and (iv) a clear description of the division of work among teammates, if applicable.

Project Spotlight. On Thursday, December 6, you will have 10 minutes to present your project to the class.

Final Report: Due Thursday, December 13 at noon. The paper should be in ICML format. Maximum 8 pages. (You can have an appendix with extra material if needed. If working in groups of two, please include a clear description of the contribution of each person in the appendix.) You should submit a pdf file electronically. It should have the following format:

1. Introduction. A quick summary of the problem, methods and results.
2. Problem description. Detailed description of the problem. What question are you trying to address? What have people done before?
3. Methods. Description of methods used and algorithms.
4. Theory. This section should contain a cogent discussion of the theoretical properties of the method. It should also discuss under what assumptions the methods should work and under what conditions they will fail. You should do your best to develop new theoretical results.
5. Simulation studies and/or data example. Results of applying the method to simulated and/or real data sets.
6. Conclusions. What is the answer to the question? What did you learn about the methods? Mention any future directions of interest.

Notes: You can also choose to do a purely theoretical project. In this case, you should choose an area of interest, read several key papers, and provide a clear, unified summary of the theoretical results in these papers. The formatting and deadline are to suggest that, perhaps with some extra effort, you can submit this paper to an ML conference (perhaps ICML) or further develop it for journal submission. This is a research opportunity, and I will help you throughout the process to take advantage, including modifications for eventual submission.

Rough schedule of topics

1. Introduction (1 week)
2. Convexity and optimization (4 weeks)

   Convex sets and functions; Duality; Gradient descent; Coordinate descent; Projected gradient;
3. Concentration inequalities (3 weeks)

   Laws of large numbers; Sub-Gaussian random variables; Hoeffding; McDiarmid;
4. Uniform convergence (3 weeks)

   Rademacher averages; VC-dimension; Covering numbers; Chaining;
5. Minimax lower bounds (4 weeks)

References

Boucheron, S., Lugosi, G., and Massart, P. (2013), *Concentration Inequalities*, Oxford University Press, Oxford, UK.

Boyd, S., and Vandenberghe, L. (2004), *Convex optimization*, Cambridge Univ Press, Cambridge, UK.

Devroye, L., Györfi, L., and Lugosi, G. (2013), *A probabilistic theory of pattern recognition*, Springer, New York.

Tsybakov, A. (2009), *Introduction to Nonparametric Estimation*, Springer, New York.

van de Geer, S. (2000), *Empirical Processes in M-estimation*, Cambridge series in statistical and probabilistic mathematics, Cambridge University Press, Cambridge, UK.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Vapnik, V. (1998), *Statistical learning theory*, John Wiley & Sons, Inc., New York.