

STAT 416: HIGH DIMENSIONAL STATISTICS II

Syllabus, Winter 2017

High dimensional statistics II is graduate level course on the mathematical and statistical properties of estimation and inference procedures on high dimensional data, where the number of variables is large relative to the number of observations. Topics include primal/dual methods based on convex optimization, confidence intervals and inference, variable selection for parametric and nonparametric models, concentration of measure, classification, and algorithms that are scalable to high dimensional and large data.

Schedule

LECTURES Mondays and Wednesdayss 10:30-11:50 am Jones 226

Contact Information

Instructors:

Rina Foygel Barber Jones 214 rina@uchicago.edu
Office hours by appt.

John Lafferty Jones 121 lafferty@uchicago.edu
Office hours by appt.

Course Assistant:

Wooseok Ha haywse@gmail.com
Office hours: Mon. 6:00-7:30 Jones 308

Prerequisites

Prerequisites are Stat 30100 or Stat 30400 or 31015, or permission of the instructors.

Course Structure and Grading

The course will have a standard lecture format, and will have four written assignments. Assignments will have a mix of theory and simulation. Collaboration on homework assignments with fellow students is encouraged. However, such collaboration should be clearly acknowledged, by listing the names of the students with whom you have had discussions concerning your solution.

Course Calendar

The course calendar and other materials will be posted on the course Piazza website, <https://piazza.com/uchicago/winter2017/stat416/home>

The schedule of topics, exams, and assignments follows.

Week	Date	Topic	Assignments
1	Jan 2 Jan 4 _{JR}	— overview	
2	Jan 9 _J Jan 11 _J	primal-dual methods	assn 1 out
3	Jan 16 Jan 18 _J	— nonparametric variable selection	
4	Jan 23 _R Jan 25 _R	confidence intervals & inference	assn 1 due; assn 2 out
5	Jan 30 _R Feb 1 _R	model free inference	
6	Feb 6 _J Feb 8 _J	classification and low noise	assn 2 due, assn 3 out
7	Feb 13 _J Feb 15 _J	graphical models	
8	Feb 20 _R Feb 22 _R	concentration of measure	assn 3 due, assn 4 out
9	Feb 27 _R Mar 1 _R	risk minimization	
10	Mar 6 _J Mar 8 _J	algorithms	assn 4 due

Textbook

The course will not follow a textbook. The following books may be relevant to some of the course material:

- P. Bühlmann and S. van de Geer, “Statistics for High-Dimensional Data: Methods, Theory and Applications,” Springer Series in Statistics, 2011.
- L. Györfi, M. Kohler, A. Krzyżak, H. Walk, “A Distribution-Free Theory of Nonparametric Regression,” Springer, 2002, pdf available at <http://web.stanford.edu/class/ee378a/books/book1.pdf>.
- T. Hastie, R. Tibshirani, M. Wainwright, “Statistical Learning with Sparsity: the Lasso and generalizations,” CRC press 2015, pdf available at https://web.stanford.edu/~hastie/StatLearnSparsity_files/SLS_corrected_1.4.16.pdf.