36-788: Topics in High Dimensional Statistics I
Fall 2015

Instructor:

Alessandro Rinaldo
Email: arinaldo@.cmu.edu
Phone: x8-7362
Office: Baker Hall 229I
Office Hour: by appointment.

Course Objectives:

This course is the first of a sequence of two minis intended to introduce Ph.d. students in Statistics and Machine Learning to some of the mathematical tools used in the analysis of high-dimensional statistical models. In big-data problems, it is often the case that the statistical complexity of a single datum is increasing in the sample size. In these instances, the classical statistical modeling approach, which is predicated on the assumption that the number of model parameters remains fixed as the sample size grows unbounded, is no longer applicable. In contrast, high-dimensional statistical models allow for the number of model parameters to change with the sample size, so that, as more data are gathered, models of increasing complexity and predictive power can be used. In order to formalize statistical tasks and evaluate the theoretical performance of statistical procedures in high-dimensional settings, it is necessary to use a different suite of techniques and theories than in traditional statistics. The learning objectives of this course are two-fold.

1. The first goal is present various concentration inequalities techniques to derive finite sample upper bounds on the performance of high-dimensional algorithms.

2. The second goal is to exemplify the use of such techniques on various problems borrowed from the current literature, including high-dimensional regression and compressed sensing, matrix completion, high-dimensional graphical modeling, community detection, network analysis, ect ect.

Lectures:

Tuesday and Thursday, 1:30pm - 2:50pm, PHA18A.

Class Website:

`http://www.stat.cmu.edu/~arinaldo/36788/`

Prerequisites:

This course is intended for second-year Ph.D. students in Statistics. 36-705 (Intermediate Statistics) or equivalent course, or the instructor's permission. 36-755 (Advanced Statistics) is recommended.

Class material:

Reading material, references and notes will be posted on the class website.

Course Grading:

Your assessment and grades will be determined as follows:

- Homework assignments.
- Scribe duties.
- In class presentations (if applicable).
- Attendance and class participation.

Any failure to turn in any assignment, to fulfill the scribe duties and to miss a significant number of lectures without informing me of your absence or without a reasonable excuse will result in a lower grade.

Scribe duties:
Each student will take turn in transcribing the notes of every lecture in electronic format using the latex template available at http://www.stat.cmu.edu/~arinaldo/36788/schedule.html. The scribe has to attend class, take good and accurate notes, check for mistakes and inconsistencies, write them up in latex, adding references and expanding the material if appropriate and after consulting with me The resulting pdf and latex files have to be submitted for my approval *within one week.* The pdf files containing the lecture notes will be posted on the class website.

Homeworks:
Homework assignments will not be graded. I will only check if an honest effort has been made to solve the assigned problems.

Attendance:
It is important that you attend class, as the selection and organization of the topics will be different from the textbook. If you know you will be absent for few consecutive lectures, please let me know.

Involvement:
Please come and see me any time you are confused or stuck and don't be shy in class: the more questions you ask and the more feedback I receive from you, the better I will be able to tailor the lectures to your specific needs.

References and reading material:
There is no textbook for this class. Lecture material will be assigned or provided at each lecture. Below is a list of references on concentration inequalities.

- Concentration Inequalities: A Nonasymptotic Theory of Independencei, by S. Boucheron, G. Lugosi and P. Massart, Oxford University Press, 2013. Concentration Inequalities and Model Selection, by P. Massart, Springer Lecture Notes in Mathematics, vol 1605, 2007.
- The Concentration of Measure Phenomenon, by M. Ledoux, 2005, AMS. Lecture notes on concentration of measure Inequalities, by G. Lugosi, A Probabilistic Theory of Pattern Recognition, by L. Devroye, L. Gyfi and G. Lugosi, Springer, 1996.
- A New Look at Independence – Special Invited Paper, by M. Talagrand, the Annals of Applied Probability, 24(1),1–34, 1996.
- Concentration of Measure for the Analysis of Randomized Algorithms, by D.P. Dubhashi and A, Panconesi, Cambridge University Press, 2012.

- Concentration by C. McDiarmid. In M. Habib. C. McDriamid, J.Ramirez-Alfonsin and B. Reed, editors, Probabilistic Methods for Algorithmic Discrete Mathematics, 195–248, Springer, 1998.
- Measure Concentration, Lecture Notes for Math 710, by Alexander Barvinok, 2005.
- Probability Theory and Combinatorial Optimization, by M. Steele, CBMS Regionel Conference Series in Applied Mathematics, vol. 69, 1997.