Probability in High Dimensions

Lecturer: Shuchin Aeron
TA: TBD

Course Information

---

1 Basic Information

- **Instructor**: Shuchin Aeron, Halligan 144, shuchin@ece.tufts.edu
- **Main Textbook**: High-Dimensional Probability: An Introduction with Applications in Data Science, Roman Vershynin, Cambridge University Press. PDF available at: `https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf`
- **Teaching Assistants/Co-Instructor**: TBD
- **Office Hours**: TBD
- **Course Website**: The course will be actively managed via Piazza as well as using Gradescope. You will receive an email to enroll in the course on Piazza. All the exams and hw will be managed via gradescope and all the class discussions, class notes, will be managed via Piazza.

1.1 Other resources

1. Lecture Notes by Roman Van Handel, `https://web.math.princeton.edu/~rvan/APC550.pdf`.
2. Boucheron, Stéphane, Gábor Lugosi, and Pascal Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence (Oxford, 2013; online edn, Oxford Academic, 23 May 2013)
3. Lecture notes by Joel Tropp, `https://tropp.caltech.edu/notes/Tro21-Probability-High-LN.pdf`.
4. Related research papers, articles, and class handouts.

2 Overview

Typical problems in statistics and machine learning involve study of functions of a large number (hence high dimensional) of independent or weakly dependent random variables, e.g., PCA involves study of eigenvalues and eigenvectors of matrices with random entries, learning a classifier involves study of empirical risk that is average loss over many examples (labeled data). In this course we will study the behavior of these functions and in particular look at how they concentrate around their expected behavior – a phenomena broadly referred to as concentration of measure – with emphasis on non-asymptotic bounds.

Several interesting phenomena occur in high dimensions leading to striking and groundbreaking insights, which we will illustrate via several examples throughout the course. Core topics covered will include concentration of sums of independent random variables (sub-Gaussian, sub-exponential), concentration of norms of random vectors and matrices in high dimensions, with applications to community detection, dimensionality reduction, and clustering. The advanced and widely applicable topic of suprema of stochastic processes will be also be covered with several applications (theoretical bounds) to statistical machine learning and signal processing.

3 Tentative schedule and coverage from the main textbook

Note: The HW are issued at points when we are done covering the necessary material as indicated below. The due date for each HW will also depend on the length of the assignment and exam schedule.

1. Chapter 1: Motivation, basic preparation in probability.
   (a) Lectures 1-2.
   (b) Issue HW 1.
2. Chapter 2: Concentration of sums of independent (SubGaussian and SubExponential) random variables
   (a) Lectures 3-8 + State and prove a version of Johnson-Lindenstrauss (JL) Lemma.
   (b) Issue HW 2,3 on Chapter 2.

   Mid-term I: Take-home exam. Chapters 1 and 2.
3. Chapter 3: Random vectors in high dimensions [We will skip sections 3.5, 3.6., 3.7]
   (a) Lectures 9-12 + Pointers to some recent developments.
   (b) Issue HW 4.
4. Chapter 4: Random matrices
   (a) Lecture 13-15.
   (b) Issue HW 5.

   Mid-term II: Take-home exam. Chapters 1, 2, 3, 4.
5. Lectures 16-24 will cover advanced topics with roughly covering topics from these chapters. This may change depending on the interest of the class.
   (a) Chapter 5: Concentration without Independence (Optional)
   (b) Chapter 6: Symmetrization and Contraction Lemmas (Optional)
   (c) Chapter 7: Stochastic Processes, Suprema of Gaussian Processes
   (d) Chapter 8: Chaining and Generic Chaining, Applications to Machine Learning
   (e) Chapter 9: Matrix Deviation Inequalities, Applications to Signal Processing

Final take-home exam.

3.1 Homework and Exams Grade Distribution

1. **HW** - 50% **NOTE:** The HW are very important in this course. Please do spend a lot of time solving for them and also go over the solutions that are provided by the instructor and TA. **HW submission**: All HW are to be submitted and graded via gradescope.

   HW delays: Delays are not permitted unless there is a strong reason. Please contact the instructor and TA beforehand if the delay is to be anticipated and each case will be dealt on an individual basis.
2. **Take-Home Mid-Terms** - 15% each totaling to 30%
3. **Take-Home Final Exam** - 20%