USC Viterbi School of Engineering

EE 599,
High-dimensional probability and Statistics for the Data Sciences

Units: 03
Mon-Wed-10:30-11:50 AM:

Location: VKC 256

Instructors: Mahdi Soltanolkotabi

Mahdi Soltanolkotabi
Office: EEB422
Office Hours: 5:30-7 PM Mon-Wed
Contact Info: soltanol@usc.edu

IT Help: TBA
Hours of Service:
Contact Info:

Course Description
This course focuses on probability and statistics in high dimensions with a view toward applications in the data sciences. Fundamental topics include: concentration of measure, concentration or random vectors and matrices in high dimensions, community deterction, covariance estimation and clustering, randomized dimensionality reduction, stochastic processes and statistical learning. Applications to machine learning, statistics and signal processing will be presented.

Learning Objectives
Students will learn some of the key probabilistic methods and results that should form an essential toolbox for a mathematical data scientist. The students will be exposed to numerous applications of probability theory in a multitude of applied and theoretical desciplines.

Prerequisite(s):

EE 599 Enrollees: EE441 (Linear Algebra), EE 503 (Probability), and Mathematical maturity

Co-Requisite (s):
Concurrent Enrollment: course(s) that must be taken simultaneously: N/A
Recommended Preparation: N/A

Course Notes
Grading Type: letter grade
The course is Web-Enhanced **(Blackboard)**.
Copies of lecture slides and other class information will be posted on Blackboard.

Technological Proficiency and Hardware/Software Required
None.

Required Readings and Supplementary Materials
Required textbook:

High Dimensional Probability for Mathematicians and Data Scientists, Roman Vershynin

Additional textbook:
Concentration Inequalities: A Nonasymptotic Theory of Independence. Stephane Boucheron, Gabor Lugosi, Pascal Massart

Description and Assessment of Assignments
Students will be assigned a homework every other week. Homework will consist of solving textbook problems and will sometimes include a "research-oriented" problem to stimulate and probe students' creativity. Homeworks are to be submitted in class on the due date. Late homeworks will not be accepted unless prior approval for late submission has been obtained.

Grading Breakdown

Assignment	Points	% of Grade
participation		10%
homework		90%
TOTAL		1

Assignment Submission Policy

Homework to be submitted in class two weeks after assignment. Late homeworks will not be accepted unless prior approval for late submission has been obtained. Best 5 of 6 homeworks will count towards final grade.

Additional Policies

Attendance of the lectures is expected.

Course Schedule: A Weekly Breakdown

	Topics/Daily Activities	Readings and Homework	Deliverable/ Due Dates
Week 1 Dates	Preliminaries on random variables: Basic quantities associated with random variables, Law of large numers and the central limit theorem	Textbook Chapter 1,	
Week 2 Dates	Concentration of sums of random variables: sub-Gaussian distributions, Hoeffding, Chernoff and Khinchine's inequalities,	Textbook Chapter 2,	Homework #1 assigned
Week 3 Dates	applications to random graphs, sub-exponential distributions and Bernstein's inequality	Textbook Chapter 2,	
Week 4 Dates	Random vectors: Norm concentration, covariance matrices and isotropic distributions, sub-Gaussian distributions in higher dimensions	Textbook Chapter 3,	Homework #1 due, Homework#2 assigned.
Week 5 Dates	Applications: Grothendieck's inequality and semidefinite programming, Maximum cut for graphs, Kernel trick, and tightening of Grothendieck's inequality	Textbook Chapter 3,	
Week 6 Dates	Sub-Gaussian random matrices: Nets, covering numbers and packing numbers, upper bounds on sub-Gaussian matrices	Textbook Chapter 4, Sections 4.1-4.3	Homework #2 due, Homework#3 assigned.
Week 7 Dates	applications in community detection in networks, two-sided bounds on sub-Gaussian matrices, Applications in covariance estimation and	Textbook Chapter 4, Sections 4.3-4.5	

	clustering		
Week 8 **Dates**	Concentration without independence: Concentration of Lipschitz functions on the sphere, concentration on other metric measure spaces, applications in dimensionality reduction	Textbook Chapter 5, Sections 5.1-5.3	Homework #3 due, No homework assigned because of midterm.
Week 9 Dates	Matrix concentration inequalities, Matrix Hoeffding and Bernstein, community detection in sparse networks, covariance estimation for general distributions	Textbook Chapter 5, Sections 5.4-5.7 An introduction to matrix concentration inequalities by J. A. Tropp	
Week 10 Dates	Random Processes I: Gaussian processes, Slepian's inequality, sharp bound on Gaussian matrices	Textbook Chapter 7, Sections 7.1-7.4	Homework#4 assigned
Week 11 Dates	Random Processes II: Sudakov's minoration, empirical method for net construction, Guassian width, random projections of sets	Textbook Chapter 7, Sections 7.5-7.8	.
Week 12 Dates	Chaining I: Dudley's inequality, empirical processes, uniform laws of large numbers, Applications in statistical learning	Textbook Chapter 8, Sections 8.1-8.3	Homework #4 due, Homework#5 assigned.
Week 13 Dates	Chaining II: Generic chaning, majorizing measure and comparison theorem VC- dimension	Textbook Chapter 8, Sections 8.4-8.7	
Week 14 Dates	Sparse recovery and generalizations	Text book Chapter 10	Homework #5 due, Homework#6 assigned.
Week 15 Dates	Low-rank recovery and generalizations	Text book Chapter 10	Last homework # 6 due.
Week 16 Dates	Applications to high-dimensional statistics and Learning		

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards* https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct/.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu/ or to the *Department of Public Safety* http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* *http://emergency.usc.edu/* will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.