University of Colorado at Boulder **Fall 2020**
APPM 4/5515, High-Dimensional Probability for Data Science

Syllabus and Policies

Course Information

1. Credit Hours: 3
2. Semester: Fall 2020
3. Meeting times and location: Monday, Wednesday, and Friday 1:50 PM – 02:40 PM, ZOOM: https://cuboulder.zoom.us/j/98459013215
4. Course website: https://canvas.colorado.edu/courses/63514
5. Relevant course material, including this syllabus, and course related resources will be made available at the course website.

Instructor Information

1. Professor François G. Meyer
2. Office: ECOT 321
3. Email: fmeyer@colorado.edu
4. Office Hours on ZOOM: https://cuboulder.zoom.us/j/94460263690

Course Description

The course provides a senior undergraduate / first-year graduate level course for students interested in probabilistic methods for data science. The emphasis is on big data, wherein the size or dimensionality of the data requires new mathematical methods, not just bigger computers.

General Policies

APPM4/5515 is a course that you take because you are really interested in the subject. No one is forced to take this course. We therefore consider all of you to be students in training and will try to provide assignments and feedback that help you gain all the skills and knowledge you will need to.

You are expected to come to all the lectures, to arrive prior to the starting time, to remain for the entire class, and to follow basic classroom etiquette. In general, we will follow CU default policies; I will assume that you have read them and agree to abide by them. You are expected to know everything that is announced on the class website available on canvas, or discussed during lectures.

Prerequisite

APPM 3570 (Applied Probability); APPM 3310 (Matrix Methods).

Course Outline

1. Review of probability and random variables
2. Concentration of sums of independent random variables
   (a) Hoeffding's inequality
   (b) Chernoff 's inequality
   (c) Bennett's inequality
   (d) Bernstein's inequality
3. Geometry in high dimensions
   (a) Norms and angles
   (b) Volumes and surfaces
4. Concentration of measure without independence
   (a) Concentration on the sphere
   (b) Grothendieck's inequality
   (c) Concentration of Lipschitz functions on the sphere
   (d) Concentration on other metric measure spaces
5. Application: random matrices deviation inequality
   (a) Matrix calculus
   (b) Matrix Bernstein's inequality
   (c) Matrix deviation inequality
6. Application: random matrices: universality of spectral statistics
   (a) Wigner Matrices and Semicircular Law
   (b) Sample Covariance Matrices and the Marcenko-Pastur Law
   (c) Spectrum of random graphs
7. Application: Johnson-Lindenstrauss lemma
   (a) The Johnson-Lindenstrauss lemma
   (b) Random projections

Students Learning Outcomes

Students acquire the fundamental theoretical skills necessary to develop the future data-analytics algorithms. The theoretical training, based on recent groundbreaking developments in mathematics, computer science, and engineering, is complemented with hands-on programming experience.

Textbook

High-Dimensional Probability: An Introduction with Applications in Data Science, Roman Vershynin, Cambridge University Press 2018.

Problem sets

Problem sets will be due on a weekly basis. Type your solution using Word or LaTeX, submit a PDF document on canvas. HANDWRITTEN SOLUTIONS WILL NOT BE GRADED.

You must show all work, correct solutions without adequate justification will receive no credit. All work turned in must be your own.

Computing

Some assignments will require that you be proficient in at least one programming language (e.g., R, Python, MATLAB, etc.) You must have access to a computer on which you can install software.

Help and Resources

There are multiple resources that are available to you.

You can use office hours to ask questions about the homework, and the class material. Students who cannot come to official office hours sessions should schedule an appointment.

In addition, the beginning of class is reserved for questions about the class material, and general questions about homework. Questions are always welcome in class.

I organize in-class review sessions for each midterm. I write problems that will help you study for the test, and will help you asses your understanding of the class material. You should be able to solve these problems with closed book and notes. I write the solutions to these problems. Please solve these problems and attend these review sessions.

Plagiarism and Cheating

The homework are designed to encourage you to work on the class material. You are welcome to discuss the material with other students and the TA in order to derive your own solution. However, copying from each other is not legitimate collaboration but fundamental cheating, you will automatically get zero on that assignment. If you realize that you rely too much on your classmates for solving problem sets, you need to contact the instructor and the TA to develop other strategies.

If we find that your answer in any assignment is copied from online sources or a textbook, you will automatically get zero on that assignment. Your final grade will be computed using the following weights,

- Homework assignments $W_H = 0.10$
  To allow for unexpected difficulties, your lowest homework score will be dropped.
- Midterms: $W_M = 0.6 = 3 \times 0.2$
- Final exam: $W_F = 0.3$

For letter grade equivalents, we will be using the following algorithm.

We first compute a grade, X, on a scale from 0 to 4 as follows,

$$X = 0.04 \times (W_H X_H + W_M X_M + W_F X_F)$$

where X_H, X_M, and X_F are your grades (on a 0 to 100 scale) for the homework, midterms, and final, respectively.
Finally, your letter grade is given by the following intervals,

- $3.77 \leq X \leq 4.0$ is A
- $3.60 \leq X < 3.77$ is A-
- $3.48 \leq X < 3.60$ is B+
- $3.36 \leq X < 3.48$ is B
- $3.20 \leq X < 3.36$ is B-
- $3.08 \leq X < 3.20$ is C+
- $2.96 \leq X < 3.08$ is C
- $2.80 \leq X < 2.96$ is C-
- $2.68 \leq X < 2.80$ is D+
- $2.56 \leq X < 2.68$ is D
- $2.44 \leq X < 2.56$ is D-
- $0 \leq X \leq 2.44$ is F

Incomplete grades will not be issued without a valid documented reason. A petition is required, delineating how the incomplete grade will be resolved. Retaking the course is not an acceptable way to resolve an incomplete.

Late Work Policy

Everyone in the class should have the same amount of time to deal with the assignments. To ensure that your work is graded in comparison with the other students' assignments, we need to be able to grade the work of all the students at the same time.

Work that is handed in after the deadline, but before the answers are posted will get partial credit. Obviously, any work that is handed in after answers are posted will not receive any credit, except under rare and well-documented circumstances (see below). Late assignments without a formal excuse (see next paragraph) will have 10% credit deducted for each calendar day that it is late. For instance, a homework that is due on Friday, and is turned in on Monday is two days late and will count only for 80 % of the maximum grade. Homework that are five days late will not be graded.

If you know in advance that one of the deadlines conflicts with an obligation (e.g., religious holiday, travel to an interview, etc.) it is your responsibility to contact me or the TA and arrange an alternative due date for

your work. If you do this in advance, you will not be penalized.

If you get sick or have an accident that prevents you medically from finishing your work, you may also be given extra time, but only under these circumstances: you must get a note from a doctor or registered health professional.

University Policies

Classroom Behavior

Both students and faculty are responsible for maintaining an appropriate learning environment in all instructional settings, whether in person, remote or online. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. For more information, see the policies on classroom behavior and the Student Code of Conduct.

Requirements for COVID-19

As a matter of public health and safety due to the pandemic, all members of the CU Boulder community and all visitors to campus must follow university, department and building requirements, and public health orders in place to reduce the risk of spreading infectious disease. Required safety measures at CU Boulder relevant to the classroom setting include:

- maintain 6-foot distancing when possible,
- wear a face covering in public indoor spaces and outdoors while on campus consistent with state and county health orders,
- clean local work area,
- practice hand hygiene,
- follow public health orders, and
- if sick and you live off campus, do not come onto campus (unless instructed by a CU Healthcare professional), or if you live on-campus, please alert CU Boulder Medical Services.

Students who fail to adhere to these requirements will be asked to leave class, and students who do not leave class when asked or who refuse to comply with these requirements will be referred to Student Conduct and Conflict Resolution. For more information, see the policies on COVID-19 Health and Safety and classroom behavior and the Student Code of Conduct. If you require accommodation because a disability prevents you from fulfilling these safety measures, please see the "Accommodation for Disabilities" statement on this syllabus.

Before returning to campus, all students must complete the COVID-19 Student Health and Expectations Course. Before coming on to campus each day, all students are required to complete a Daily Health Form. In this class, you may be reminded of the responsibility to complete the Daily Health Form and given time during class to complete it.

Students who have tested positive for COVID-19, have symptoms of COVID-19, or have had close contact with someone who has tested positive for or had symptoms of COVID-19 must stay home and complete the Health Questionnaire and Illness Reporting Form remotely.

Accommodation for Disabilities

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website. Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition or injury, see Temporary Medical Conditions under the Students tab on the Disability Services website.

Preferred Student Names and Pronouns

CU Boulder recognizes that students' legal information doesn't always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructors' class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name.

Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code honor@colorado.edu; 303-492-5550). Students found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found at the Honor Code Office website.

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

The University of Colorado Boulder (CU Boulder) is committed to fostering an inclusive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct (harassment, exploitation, and assault), intimate partner violence (dating or domestic violence), stalking, or protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, anonymous reporting, and the campus resources can be found on the OIEC website.

Please know that faculty and instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, dating and domestic violence, stalking, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance.

In this class, you may turn in your homework early if you cannot come to class when the homework is due. If you know in advance that one of the deadlines or exam conflicts with a religious observance, it is your responsibility to contact me or the TA and arrange an alternative due date for your work, or reschedule and exam. If you will be missing an exam, please inform us at least one week prior to the exam so that we can schedule a makeup exam.

See the campus policy regarding religious observances for full details.