New York University Tandon School of Engineering
Computer Science and Engineering

CS-GY 9223I (CS-UY 3943)
Algorithmic Machine Learning and Data Science, Fall 2019

Professor Christopher Musco
Wednesdays 3:20-5:50pm, Rogers Hall, RGSH 707.

PROFESSOR CONTACT

Email: cmusco@nyu.edu.

Office: 370 Jay St., #1105 (Floor 11).

Office Hours: Thursdays, 3-5pm, or by appointment.

Course Webpage: https://www.chrismusco.com/9223_2019/.

COURSE PREREQUISITES

I require a previous course in machine learning (for example, CS-UY 4563, CS-GY 6923, or ECE-GY 6143) and a previous course in algorithm design and analysis (for example, CS-UY 2413, CS-GY 6033, or CS-GY 6043). Experience with linear algebra and probability is also necessary. Email the course instructor if you have questions about your preparation for the course.

COURSE DESCRIPTION

This course gives a behind-the-scenes look into the algorithms and computational methods that make machine learning and data science work at large scale. How does a service like Shazam match a sound clip to a library of 100 million songs in under a second? How do scientists find patterns in terabytes of genetic data? How can we efficiently train neural networks with millions of parameters on millions of labeled images? We will address these questions and others by studying advanced algorithmic techniques like randomization, approximation, sketching, optimization, and spectral methods, and Fourier methods.

COURSE OBJECTIVES

1. Students will build experience with the most common algorithmic tools used in machine learning, including optimization, relaxation, spectral methods, Fourier methods, and Monte Carlo algorithms. They will also strengthen their understanding of the mathematical concepts behind these tools.

2. Through problem sets and in-class assignments, students will practice applying, combining, and modifying the methods learned in lectures to specific problems in machine learning and data analysis. The goal is to prepare students to use these tools in industrial or academic positions where they are developing computationally efficient machine learning methods.

3. Students will learn how theoretical questions (involving asymptotic complexity, communication complexity, convergence rate, approximation quality, etc.) can guide the design of new algorithmic methods in machine learning. We will practice framing and rigorously answering such questions.

4. The course will prepare students to contribute to research projects on computational methods for machine learning and data science.

5. Through assigned papers and discussions, students will improve their ability to read, summarize, and extract value from contemporary research papers in machine learning and data mining conferences like NeurIPS, ICML, KDD, ICLR and AAAI.

COURSE STRUCTURE

One meeting per week, involving both lecture and in class assignments. 4 **problem sets** involving analysis and application of methods learned in class, potentially with programming exercises to further explore lecture content (40% of grade). **In-class midterm exam** (20% of grade, scheduled for **10/23**). **Final exam** (30% of grade). **Class participation** is the remaining 10% of the grade.

Our meeting period is 2.5 hours long. We will typically begin and end with lecture, pausing in the middle of class for a break and also to work on (non-graded) in class assignments, which will resemble problem set questions. Solutions will be reviewed together in class.

READINGS

There is no textbook to purchase. Many of the topics covered are new, and not sufficiently addressed in existing textbooks. Course material will consist of my written lecture notes, as well as assorted online resources, including papers, notes from other courses, and publicly available surveys. Please refer to the course webpage (https://www.chrismusco.com/9223_2019/) before and after lecture to obtain links.

COURSE REQUIREMENTS

Lectures: Attend each lecture and participate in all in-class assignments, including in discussing solutions.

Problem sets: Complete all problem sets by the specific due date. The following tentative dates are subject to change:

- Problem Set 1, Out 9/11. Due 9/26.
- Problem Set 2, Out 10/2. Due 10/17.
- Problem Set 3, Out 11/6. Due 11/21.
- Problem Set 4, Out 11/25. Due 12/14.

While not required, I encourage students to prepare problem sets in LaTeX or Markdown (with math support.) Students will receive 10% extra credit on Problem Set 1 for preparing it in LaTeX or Markdown.

Problem set collaboration policy: *Discussion of high-level ideas for solving problems is allowed, but all solutions and any code must be written up independently.* This reflects the reality that research and algorithm design in machine learning are rarely done alone. However, any machine learning practitioner must be able to communicate and work through the details of a solution individually.

Students must name any collaborators they discussed problems with on their assignments (list at the top of each problem separately). Do not write in parallel with other students: if problem ideas are discussed, solutions should be written at a later time, individually. I take this policy very seriously. Do not paraphrase, change variables, or in any way copy another students solution.

Midterm: Take the midterm exam, which will be held during the first half of class on **October 23rd**. Please let me know ASAP if you will not be able to attend that day so that we can make alternative arrangements.

Final: Take the final exam, which will be held during standard class hours (at 3:20pm) on **December 18th**. Undergraduates should let me know ASAP if this conflicts with any of your other scheduled exams.

COURSE SCHEDULE

The following schedule is tentative and subject to change.

THE POWER OF RANDOMNESS

1. 9/4, Concentration of random variables + applications to hashing and load balancing
2. 9/11, Sketching and streaming algorithms + models of computation for data processing
3. 9/18, The Johnson-Lindenstrauss lemma + applications to high dimensional data
4. 9/25, Nearest neighbor search + locality sensitive hashing

OPTIMIZATION

4. 10/2, Convexity in machine learning + vanilla, stochastic, and online gradient descent
5. 10/9, Conditioning, acceleration, coordinate descent, quasi-Newton methods
6. 10/16, Learning from experts + multiplicative weights
7. 10/23, **Midterm Exam** (first half of class), Beyond convexity: training neural networks and other non-convex models (second half of class)

SPECTRAL METHODS AND LINEAR ALGEBRA

9. 10/30, Singular value decomposition, Krylov methods + application to spectral clustering.
10. 11/6, Randomized linear algebra, sketching for linear regression, -nets arguments.
11. 11/13, Importance sampling, leverage scores, active learning.

FOURIER METHODS

12. 11/20, The Fourier transform, DFT, FFT + applications
13. 12/4, Compressed sensing, the restricted isometry property, basis pursuit
14. 12/11, Kernel methods in machine learning, random Fourier features
15. 12/18, **Final exam** (regular class hours)

ADDITIONAL INFORMATION

MOSES CENTER STATEMENT OF DISABILITY.

If you are student with a disability who is requesting accommodations, please contact New York Universitys Moses Center for Students with Disabilities (CSD) at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about the Moses Center can be found at www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 3rd floor.

NYU SCHOOL OF ENGINEERING POLICIES AND PROCEDURES ON ACADEMIC MISCONDUCT.

The complete Student Code of Conduct can be found here.

A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the Schools rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the Schools Policy on Academic Misconduct.

B. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

   1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another persons work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
   2. Fabrication: including but not limited to, falsifying experimental data and/or citations.

3. Plagiarism: intentionally or knowingly representing the words or ideas of another as ones own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
4. Unauthorized collaboration: working together on work meant to be done individually.
5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.
6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

NYU SCHOOL OF ENGINEERING POLICIES AND PROCEDURES ON EXCUSED ABSENCES.

The complete policy can be found here.

A. Introduction: An absence can be excused if you have missed no more than 10 days of school. If an illness or special circumstance has caused you to miss more than two weeks of school, please refer to the section labeled Medical Leave of Absence.

B. Students may request special accommodations for an absence to be excused in the following cases:

   1. Medical reasons
   2. Death in immediate family
   3. Personal qualified emergencies (documentation must be provided)
   4. Religious Expression or Practice

Deanna Rayment, deanna.rayment@nyu.edu, is the Coordinator of Student Advocacy, Compliance and Student Affairs and handles excused absences. She is located in 5 MTC, LC240C and can assist you should it become necessary.

NYU SCHOOL OF ENGINEERING ACADEMIC CALENDAR

The full calendar can be found here.

The last day of the final exam period is Friday, December 20, 2019. Final exam dates for undergraduate courses will not be determined until later in the semester. Final exams for graduate courses will be held on the last day of class during the week of December 16, 2019 – December 20, 2019. If you have two final exams at the same time, report the conflict to your professors as soon as possible. Do not make any travel plans until the exam schedule is finalized.

Also, please pay attention to notable dates such as Add/Drop, Withdrawal, etc. For confirmation of dates or further information, please contact Susana: sgarcia@nyu.edu