Mathematics for Machine Learning

Summer 2024

Course Overview

What is this course?

This is a topics course meant to strengthen the mathematical fundamentals for students wishing to pursue further study in machine learning. The serious study of machine learning requires a student to be proficient in several prerequisite subjects: (i) linear algebra, (ii) multivariable calculus, and (iii) probability and statistics. This course assumes that the student has already taken courses in these subjects at the undergraduate level (it is *not* a replacement), but would like to be more comfortable with their mathematical maturity in any of these areas before approaching a formal course in machine learning at the level of, say, COMS 4771.

We will *not* give comprehensive treatment of each of these areas; instead, we will present the main results that are most relevant to the analysis and design of machine learning models. Alongside the theory, we will also motivate each topic with numerous applications and examples relevant to machine learning so they are more familiar when encountered in future study.

Topics will include (but are not limited to): abstract vector spaces, singular value decomposition, eigenvalues and eigenvectors (linear algebra), vector calculus, continuous optimization, convex optimization (calculus and optimization), review of basic probability, conjugacy, exponential families, and multivariate Gaussians (probability and statistics). Machine learning applications will include initial exposures to: principal component analysis, least squares regression, gradient descent, and density estimation.

Logistics and Details

Instructor: Samuel Deng (samdeng@cs.columbia.edu).

Dates and times: Monday and Wednesdays 10:10am - 1:20pm ET.

Location: *TBD.*

Instructor Office Hours: *TBD.*

- I would love to meet every student taking this course, so please come to office hours at least once, if only to introduce yourself!

Teaching Assisstants (TBD):

- Alice Alice (aa123@colmbia.edu)
- Bob Bob (bb123@columbia.edu)

TA Office Hours: *TBD.*

Prerequisites

This course is not meant to be a replacement for the undergraduate level courses that are already prerequisites to machine learning:

- Multivariable calculus at an undergraduate level (e.g., Math 1201, Math 1205).
- Linear algebra at an undergraduate level (e.g., Math 2010, COMS 3251).
- Probability theory (with calculus) at an undergraduate level (e.g., Math 2015, Math 1201).
- Discrete mathematics at an undergraduate level (e.g., COMS 3203).

Instead, this course assumes familiarity with the above prerequisites and our main goal will be to focus on building up to advanced topics, techniques, and applications from the above subjects that may have been glossed over in an introductory course that will be useful in further study of machine learning. That being said, we *do not* assume that you are an expert in any of the above prerequisites to take this course; we hope that the additional practice you receive in this course gets you a little closer to that.

This course will also integrate some basic numerical Python programming to allow students to get comfortable with numerical computing packages such as `numpy` and `pandas` before using them more intensively in a future machine learning course. Because of this, we *recommend* previous exposure to basic Python programming, but this is not strictly required if you are willing to learn the basics as the course progresses. Previous exposure to programming (possibly in a different language) should be sufficient for this.

If you are unsure if you meet the prerequisites above, please email the instructor.

Resources and Links

There will be no official textbook for this course, but we will roughly follow some of the topics and applications from the textbook *Mathematics for Machine Learning* by Marc Peter Deisenroth, A. Aldo Faisal, and Chen Soon Ong. All slides will be published before each lecture so students can follow along during lecture.

Though this class assumes that you've taken courses in each of the following areas, you may want to brush up. Here are a few undergraduate-level resources for the following subjects.

- **Linear Algebra.**
  - *Linear Algebra and Applications* by Gilbert Strang.
  - Daniel Hsu's course notes for Computational Linear Algebra.
  - 3Blue1Brown's "Essence of Linear Algebra" videos.
- **Probability Theory/Statistics.**
  - *A First Course in Probability* by Sheldon Ross.
  - *Introduction to Probability* by Blitzstein and Hwang.

    - *Probability and Statistics for Engineers and Scientists* by Wadpole et al.

- **Multi-variable Calculus.**
    - *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach* by Hubbard and Hubbard.
    - MIT OpenCourseware multivariable calculus course.

Assignments and Grading Policy

This course will be evaluated on the basis of six weekly problem sets and a final project. There are no exams.

Problem Sets

To give you practice and reinforce the concepts learned in class, there will be six weekly problem sets. The problem sets will usually have three to four theoretical problems with an additional coding exercise to reinforce the concepts and develop basic fluency in machine learning packages such as `numpy`, `pandas`, and `sklearn`. The problems will be proof-based, but we will aim to develop your mathematical maturity in reading and writing proofs throughout the course.

This course is intended to prepare students for further courses in machine learning, which all require LaTeX typesetting (at least at Columbia). LaTeX is also a useful skill to have for future courses or research, and LaTeX documents just look *clean.* Because of this, **we will require that your assignments be typed in LaTeX to give you practice**; the problem sets in COMS 4771 are hard enough without needing to wrestle LaTeX typesetting issues! Resources and a submission template will be provided to learn LaTeX to gently onboard students.

Final Project

The final project of this course will be to *attempt* to read a research paper in machine learning. Emphasis on *attempt*: there is no expectation that you will understand every single detail in the paper. However, you might be pleasantly surprised that you understand a bit more than you would've at the beginning of the course just by strengthening your mathematical foundations.

There are three parts to this project:

1. **Choose a paper.** *Within the first week.* We will provide a list of machine learning papers from recent conferences in machine learning that represent some topics from the cutting edge of current research. Your job is to just peruse the list, browse the abstracts, and choose a paper that seems interesting to you based on its title and abstract alone.

2. **Beginning of course evaluation.** *Within the second week.* You will attempt to read the whole paper. Research papers can be intimidating if you've never read one before (and even if you've read hundreds!) so we will provide some guidance on how to read a scientific paper in machine learning. Then, you will provide a critical evaluation of the paper to the best of your current ability based on a template we will supply. This will be graded on completion and effort.

3. **End of course evaluation.** *Final week.* At the end of the course, you will read the paper again. You will fill out a similar critical evaluation of the paper, per the same template.

This project will be graded on the clarity and quality of the evaluation, but we stress that we will not focus on how much you "get" the paper. The emphasis of this project is on your own growth — hopefully, you'll find that by the end of the course your chosen paper isn't quite as perplexing as it may have seemed in the beginning.

Course Content (Summer Schedule)

The course will be split into three main parts, for each "pillar" of mathematics that underlies machine learning. Throughout the course, we will aim to alternate between theory and application. We will first introduce the math behind a key ML concept; then, after you've learned the theory, we will present an application of that concept in machine learning. By doing so, we aim to reinforce each new piece of mathematics you learn with a concrete application that has relevance for those who wish to study machine learning further.

Although each third of the course will focus on a certain mathematical theme, they will not be disjoint – each week will build on the one before, so we encourage attending and carefully reviewing each lecture as the semester proceeds.

This is a course outline for a six week summer version. Each week will end with a specific ML application to wrap up that week's math content, which will take approximately half of the class time. Because of the accelerated schedule, less focus will be put on the coding aspect of the course as to not overwhelm students, particularly in the pilot version.

Linear Algebra.

- Week 1-1: Introduce the basic machine learning problem: regression. Review of basic linear algebra (vector spaces, matrices, linear independence and linear mappings).
- Week 1-2: Orthogonality, norms, and inner product spaces. *ML Application:* ordinary least squares regression.
- Week 2-1: Eigenvalues, eigenvectors, positive semidefinite matrices, and the spectral theorem.
- Week 2-2: Matrix decompositions and the singular value decomposition (SVD). *ML Application:* principal components analysis.

Calculus and optimization.

- Week 3-1: Vector calculus (derivatives, gradients, matrix calculus).
- Week 3-2: Linearization, Taylor series. *ML Application:* Gradient descent, basic neural networks, and backpropagation.
- Week 4-1: Basics of continuous optimization (constrained optimization, Lagrange multipliers).
- Week 4-2: Convex optimization and duality. *ML Application:* Linear programming and support vector machines.

Probability and statistics.

- Week 5-1: Probability fundamentals (random variables, expected value, variance, law of large numbers, central limit theorem).
- Week 5-2: Statistics fundamentals (basic distributions, sufficient statistics). *ML Application:* Maximum likelihood estimation.

- Week 6-1: Statistics fundamentals continued (conjugacy, exponential family). Gaussian and multivariate Gaussian distribution.
- Week 6-2: Multivariate Gaussian. *ML Application:* Gaussian mixture models and density estimation.

Collaboration Policy

Learning is best done in collaboration with peers. To this end, you will be allowed to collaborate with other students on the problem sets in groups **up to three students (including yourself).** All collaborators must write the names and UNIs of their group at the top of each problem set. All collaborators must also *type up everything in their own words.* You are free to discuss, whiteboard, and brainstorm with your collaborators. However, when it comes to sitting down and solving the actual problem, you must do it yourself, away from your collaborators.

For the final project, only individual work is allowed. The final project is meant to track your own individual growth in mathematical maturity (which may have improved from collaborating with other students!), so you should aim to do all parts of it yourself.

Course Philosophy and Feedback

The goal of this course is to reinforce and deepen important mathematical fundamentals, gain better intuition of these mathematical tools, and develop confidence in mathematical maturity. All of these require work that may sometimes seem daunting, but I believe that *any* student is capable of growing in the course, so long as they continually grapple with the concepts and do the work. This may, at times, be difficult, but struggle is a totally normal part of the process. I was in your shoes, at one point (and still am!), and I can assure you that many of these concepts seem really difficult until they inevitably, after plugging away for a while, become natural. I hope you, the student, come away with this feeling as well.

By the nature of this course, students will come from widely different levels of background, and it is my job to make sure that no student is left behind or glossed over because of this. To this end, if there's anything I can do to help you learn better, do not hesitate to contact me directly or leave anonymous feedback.

This is also a new course, so any feedback would be much appreciated to iterate and improve it! If you have any feedback at all about the course or my teaching of the course, please submit through this anonymous feedback form.

Rationale

This is a new course, and, if you'd like to learn more about why this course exists, this section includes some of the instructor's rationale for designing it.

The goal of this course is to give students the preparation and confidence to pursue machine learning and machine learning research at Columbia by addressing a root issue: lack of sufficient mathematical preparation. A prevalent issue in upper-division machine learning courses in the department, felt by both instructors and students, is that students coming into such courses lack sufficient foundations in multivariable calculus, linear algebra, or probability (even if students have formally acquired these prerequisites). This is particularly felt in 4771, our "core" machine learning course, where there is a steep jump in expected mathematical maturity. This course seeks to remedy this issue by providing a second exposure to these prerequisites, focusing on each of these mathematical "pillars" with a view towards developing fluency in the topics that matter specifically for machine learning.

There are several main issues that this course will address:

- **Variance in prerequisite courses.** The computer science department allows students to take multivariable calculus, linear algebra, and probability and statistics in different departments, with different professors and syllabi. This introduces a large amount of variance in what students might learn in these courses, and knowledge gaps in *any* of these three areas can lead to downstream confusion. This course aims to have a consistent syllabus that patches up such potential holes.

- **Insufficient focus in core ML topics.** Due to being "offloaded" to different departments, the prerequisite courses may not focus on speecific topics and results that have particular importance in machine learning. For example, a linear algebra course may gloss over topics such as matrix factorizations and SVD, which are central to many machine learning models and algorithms. A multivariable calculus course may focus on line and path integrals while a student interested in machine learning may need more time learning about optimization. This course aims to focus on the core topics from each prerequisite in greater detail than their coverage in prerequisite undergraduate courses.

- **Lack of motivation in the form of applications.** Prerequisite undergraduate courses often present theory without sufficient motivation in why a student might need that theory. This course aims to provide numerous examples from machine learning for each concept to reinforce the theory learned each week. This has the added benefit that a student taking future ML courses will already be familiar with some of the topics they encounter, allowing a deeper understanding on second exposure.

- **Mathematical maturity.** Students in subsequent ML courses often lack mathematical maturity, particularly in reading and writing proofs (and even in typesetting LaTeX). This course aims to develop mathematical maturity by shifting focus from the rote exercises students may have seen in their prerequisite courses to reading and writing proofs. The proofs in this course will gradually increase in difficulty, providing students opportunity to grow as mathematicians. Plus, LaTeX will be mandatory — 4771, for instance, is hard enough without wrangling with LaTeX code.

- **Scaffolding for those interestd in "core" machine learning.** Many students find that the jump in difficulty from the undergraduate prerequisites to, say, 4771 is steep. This course aims to provide a bridge that eases this transition. In addition, the inclusion of this course as a possible elective provides a clear roadmap for students wishing to pursue machine learning research in earnest. The math department has a "strongly suggested" track for those wishing to study mathematics at the PhD level: Honors Math A, Honors Math B, Modern Analysis I and II, Modern Algebra I and II. For math undergraduates, all these courses in analysis and algebra are not strictly required but are strongly suggested for those pursuing graduate study. This course might fill a similar role for those pursuing graduate studies in machine learning. A possible track for such a student would be: take the prerequisite courses during freshman and sophomore year, consolidate those prerequisites with this course during sophomore or junior year, and then take 4771.

- **Confidence to pursue machine learning research further.** Pursuing research in the computer science department at Columbia can often feel nebulous, particularly in machine learning. Undergraduates are often intimidated and don't even know where to begin looking for opportunities. To this end, the cumulative project in this course is a guided exercise in reading and dissecting a research paper in machine learning of the student's choice, with focus on how a student's understanding grew through the semester (see "Final Project"). This assignment's goals are twofold: (i) get the student's feet wet in reading cutting-edge research and (ii) instilling confidence that this is a learnable skill.
- **Inclusivity of machine learning at Columbia.** It is often the case that students from disadvantaged backgrounds are already behind on prerequisites as first years. Some more fortunate students may have taken some combination of multivariable calculus, linear algebra, or probability in high school. Some may have even dabbled with machine learning research before coming to Columbia. On the other hand, some students from more disadvantaged backgrounds may not have even seen calculus yet. One aim of this proposal is to level this playing field and add a clearer bridge for students to at least have the fundamentals and confidence necessary to pursue research, particularly for students who haven't had as much exposure to mathematics before coming to Columbia.

The author's personal experiences. I took COMS 4771 in Fall 2018 as an undergraduate at Columbia. This was when the computer science requirements still included tracks (mine was Intelligent Systems, the track for AI and ML), and students were required to take either Linear Algebra (exclusive) *or* Probability and Statistics. I chose to take Linear Algebra in the math department and never took Probability and Statistics. Still, I signed right up for COMS 4771 thinking, "How bad could it be?"

Upon receiving the weed-out Homework 0, I realized that it could be *pretty bad.* My linear algebra foundations were weak, particularly in eigenvalues and eigenvectors. Topics like SVD, positive definite matrices, and inner product spaces were simply not covered or covered very, very briefly in my linear algebra course. My Calculus III course, which I took long ago first semester my freshman year, focused a bit too little on differential vector calculus, optimization, and matrix derivatives for actual fluency. Worst of all, I had absolutely *no* background in probability and statistics. I remember initially seeing the symbol $\mathbb{E}[\cdot]$ and wondering why it was everywhere on the problem set. I didn't know what an expectation was!

I somehow learned enough probability online from MIT OpenCourseware that week to pass Homework 0, but, for the entire rest of the semester, it was an uphill struggle teaching myself probability from scratch while keeping up with the course. Needless to say, this was *not* an ideal way to learn machine learning (or, for that matter, what an expectation is). Of course, this was probably for lack of preparation on my part; perhaps I could've avoided this by being more proactive in planning my courses. Homework 0 was definitely meant to weed out students like me, and I somehow got in the course with a little luck and a generous curve.

I've learned through the years, however, that my experience wasn't particularly unique — many of my peers felt similarly unprepared taking COMS 4771 or other upper division machine learning courses. Some avoided such courses altogether because word-of-mouth from peers that these courses were tough and mathematically demanding. And, while that may be true, I believe that the mathematical prerequisites, maturity, and confidence needed in these courses can be learned through a more focused second exposure. That belief is what led to this course, and I hope that this course satisfies a need for "bridge" between undergraduate prerequisites and more advanced study, while encouraging more students to pursue machine learning at Columbia.