Mathematics for Machine Learning

Samuel Deng

Summer 2024

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.