Mathematics for Machine Learning: A Bridge Course

Samuel Deng [1]

[1]Department of Computer Science, Columbia University

Overview & Learning Goals

What is this course? One-semester mathematics "bridge" course to graduate-level machine learning research or coursework for upper division undergraduates.

Motivation: Upper division machine learning coursework and research demands mathematical maturity in several core areas that students learn disjointly and without a view towards machine learning specific applications and techniques.

Goal: Strengthen students' mathematical foundations for rigorous study of machine learning and data science at the advanced undergraduate or graduate level. Along the way, we:

1. Bolster students' mathematical maturity and confidence by introducing key mathematical tools, techniques, and proof ideas that appear repeatedly in the study of machine learning.
2. Introduce students to a unified narrative that pulls together disparate ideas from linear algebra, calculus, and probability and statistics to motivate these subjects' core ideas.
3. Build comfort with abstract concepts through geometric/visual intuition and real applications in data science.

Main Issues

Here are some issues our course aims to address at our institution (and we suspect might crop up at other institutions as well):

1. **Variance in prerequisite courses.** Multivariable calculus, linear algebra, and probability and statistics taught in different departments, with different professors and syllabi.
2. **Insufficient focus on core machine learning concepts.** e.g., 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 and the implications of convexity.
3. **Lack of motivation for theory.** Prerequisite undergraduate math courses often present theory without sufficient motivation in why a student might need that theory, especially with an eye towards more applied domains. Our course links everything back to two main concepts in machine learning (least squares regression and gradient descent).
4. **Inclusivity of machine learning.** Some more fortunate students may have taken some combination of multivariable calculus, linear algebra, or probability (or even machine learning research) in high school. On the other hand, some students from more disadvantaged backgrounds may not have even seen calculus yet.

Course Website & Materials

Scan this QR code to see the course website! All materials are public and the site includes:

1. A full syllabus (for the six week summer version of this course).
2. Slides for every lesson.
3. Problem sets.
4. Interactive 3D visualizations for every lesson.

If you think that this course would be helpful at your institution, I'd love to chat — shot me an email at samdeng@cs.columbia.edu!

Course Narrative

This is a course with an overarching narrative. It revolves around two main ideas that underlie modern machine learning:

1. **Least squares regression.** The "what" — a representative problem that allows us to model and teach all the salient features of a canonical machine learning problem.
2. **Gradient descent.** The "how" — the workhorse algorithm in modern machine learning that allows us to introduce students how we think about learning algorithms.

Scan this QR code to see the course overview in slides (with a bunch of pretty interactive graphics!)

Blurb from the syllabus: *Every week, we'll develop and motivate these two ideas in lecture with the tools and concepts you learn from each part of the course. As the class goes on, you'll develop different perspectives on these two ideas from, first, what we learn in linear algebra, then calculus and optimization, and, finally, probability and statistics. The hope is that, by the end of the course, you'll have a deep understanding of both these ideas in ML while also having two concrete "applications" to motivate all the abstract mathematical tools and concepts you learn in the course.*

Feedback & Initial Experiences

We piloted this course in the summer of 2024 in a six-week format, and we received overwhelmingly positive evaluations and feedback, albeit from a small sample of 14 students.

4 - Course: Overall Quality

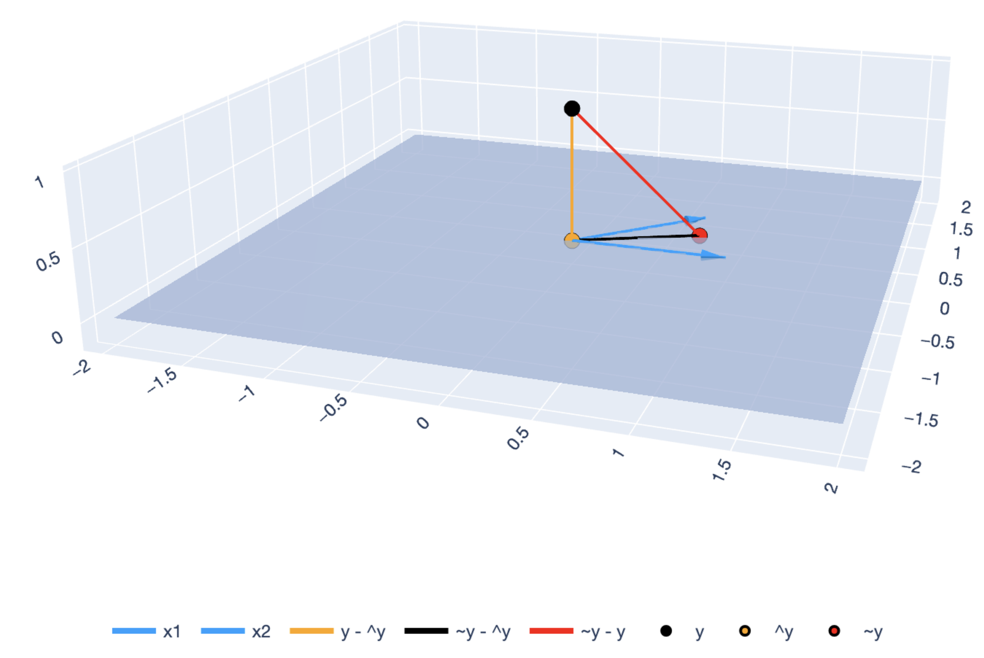
Response Option	Weight	Frequency	Percent
Poor	(1)	0	0.00%
Fair	(2)	0	0.00%
Good	(3)	0	0.00%
Very Good	(4)	1	11.11%
Excellent	(5)	8	88.89%

Response Rate	Mean	STD	Median
9/30 (30.00%)	4.89	0.33	5.00

Some more anonymous student feedback:

- *"An excellent class where you get to truly understand and gain intuition for some key concepts needed for grasping ML algorithms. Seeing the connection between linear algebra, calculus, and statistics was very valuable. One of the best classes I have taken at Columbia."*
- *"An extremely well designed and executed course (considering the constraints of a summer semester course) that should be installed as a regular fixture of the CS course curriculum. A much needed bridge between undergraduate-level courses in mathematics with graduate-level ML coursework."*
- *"I wanted to send this email to thank you for this class. I really enjoyed it and thought that I learned a lot. While writing my final evaluation, I was actually amazed by how much more of the paper I understood. In the beginning it all truly looked like gibberish. But now, I could honestly follow what the authors were talking about and understand what computations were being made. I am more confident in my Linear Algebra, Calculus, and Probability and Statistics."*
- *"This is course is marketed to students preparing for COMS 4771, but I think its value far exceeds just that individual course. Make no mistake, there is a ton of content covered in this course and it is probably better suited for a 12-week session, but this course is a tremendous value in that it cuts through the filler of at least three other standalone courses and gets us straight to the most important, fundamental aspects of ML math."*

Part 1: Linear Algebra

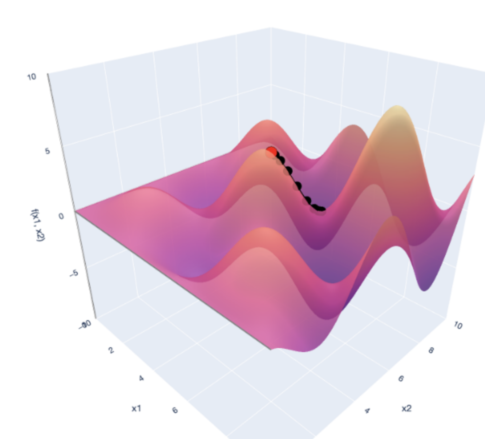


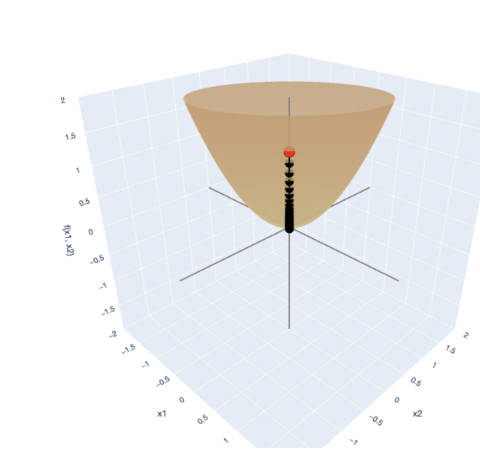
Course begins with tour of linear algebra with focus on building geometric intuition with least squares regression and projection.

- Module 1: Vectors, Matrices, and Least Squares
- Module 2: Bases, Subspaces, Orthogonality
- Module 3: Singular Value Decomposition
- Module 4: Eigenvalues and Eigenvectors

Culminates by introducing PSD matrices and that least squares regression, all along, was a quadratic form, leading to...

Part 2: Calculus & Optimization



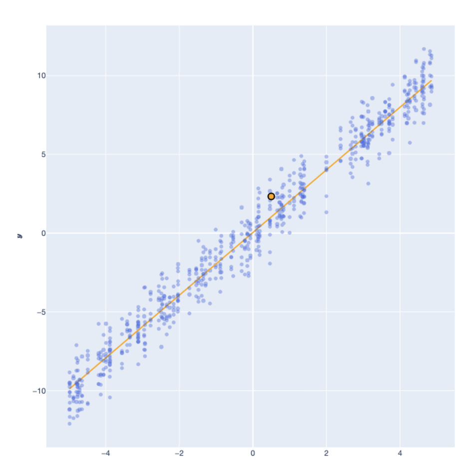


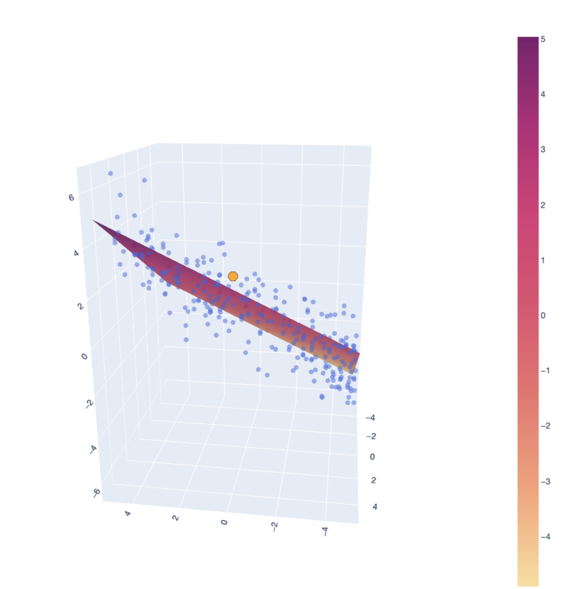
Multivariable calculus with a focus on optimizing the least squares quadratic form we introduced in Part 1.

- Module 5: Differentiation and Vector Calculus.
- Module 6: Taylor series and Linearization.
- Module 7: Optimization and the Lagrangian.
- Module 8 Convex Optimization.

Culminates in two more ways to solve least squares: one being the all-important first-order iterative solution of *gradient descent*.

Part 3: Probability & Statistics





Ground the epistemological assumptions of ML in the language of probability theory, with an eye towards analyzing statistical properties of least squares.

- Module 9: Probability Theory, Models, and Data.
- Module 10: Law of Large Numbers and Statistical Estimation.
- Module 11: Central Limit Theorem, Distributions, and MLE.
- Module 12: Multivariate Gaussian Distribution.