Introduction

AIE 6001 - Mathematics in AI

“This course equips you with the rigorous mathematical tools that underpin modern AI, from theoretical research to real-world applications.”

Jie Wang, 2025/09/01

Introduction to myself

- Assistant Professor at School of Artificial Intelligence (home department) and School of Data Science, The Chinese University of Hong Kong, Shenzhen
- Website: https://mypage.cuhk.edu.cn/academics/jwang/
- Got Bachelor in Pure Mathematics (2016-2020) at CUHKSZ
- Got Ph.D. in Industrial Engineering (2020-2025) at Georgia Tech
- Research focus: optimization and statistics for machine learning
- You can call me Prof. Wang or Dr. Wang

Contact and Office Hours

- Instructor: Jie Wang
- Email: jwang@cuhk.edu.cn
- Office Hour: 10am - 11am, every Tuesday, TA409a

- Teaching Assistant: Zhuoqi Tu
- Email: 223015150@link.cuhk.edu.cn
- Office Hour: 3pm - 4pm, every Wednesday, TA409a
- **Role:** The TA will host office hours, grade assignments, and provide support on problem sets and coding exercises.

"For urgent issues, please use email with a clear subject line. We aim to respond within 48 hours on weekdays."

Who is this course for?

This course is designed to serve the diverse goals of our graduate students:

- **Master's Students:** Preparing for high-impact careers in **AI/ML industry roles** (e.g., ML Engineer, Data Scientist, Quantitative Researcher).
- **Ph.D. Students:** Laying the essential groundwork for **cutting-edge AI research** and publication.

Why is a strong mathematical foundation critical?

- **For Industry:** Top tech and quantitative finance firms use rigorous technical interviews that test deep understanding of linear algebra, probability, and optimization—the very core of machine learning models.
- **For Research:** To innovate, you must first understand. This course ensures you can read, comprehend, and build upon the mathematical formalisms in papers from **NeurIPS, ICML, JMLR**, and other leading venues.

Why Mathematics for AI Practitioners

For Master's Students: Excelling in the Job Market

- **Technical Interviews:** A solid grasp of math is non-negotiable.
  - **Linear Algebra:** Matrix decompositions (SVD, Eigen), gradients, and their role in neural networks.
  - **Probability & Statistics:** Distributions, Bayesian inference, and model evaluation.
  - **Optimization:** Gradient descent, convexity, and convergence analysis.
- **Beyond Coding:** While programming skills are vital, **technical depth** separates candidates. Understanding the "why" behind algorithms leads to better model design, debugging, and innovation on the job.
- **Long-Term Growth:** A strong foundation enables you to quickly learn and adapt to new models and frameworks throughout your career.

"This course is your training ground for the technical challenges of the AI industry."

Why Mathematics for AI Practitioners

For Ph.D. Students: The Language of Research

- **Reading Research Papers:** The primary literature in AI is written in the language of mathematics. You must be fluent in:
  - Proofs of convergence and generalization.
  - Statistical bounds and complexity analysis.
  - Formal definitions of learning objectives and constraints.
- **Conducting Original Research:** To propose a new algorithm or theory, you need tools to:
  - Formally define the problem.
  - Analyze its properties (e.g., is it convex? What are its statistical guarantees?).
  - Compare it rigorously to existing methods.
- **Publishing:** Top conferences demand mathematical rigor. Your ability to present a clear, correct, and insightful mathematical argument is paramount.

"This course is designed to re-activate and deepen your mathematical intuition, bridging the gap between undergraduate math and modern AI research."

Why I teach this course?

My motivation stems from **industry experience** and **academic pursuit**:

- **Industry Insight:** Over the past two years, I have participated in numerous technical interviews with leading **high-tech companies (e.g., FAANG+) and quantitative hedge funds** for both internships and full-time roles. I understand the specific mathematical concepts and problem-solving skills they evaluate.
- **Research Experience:** I have dedicated the past nine years to AI-related research. My goal is to publish in the most prestigious venues, including **NeurIPS, ICML, JMLR, and ISIT**. This requires a daily engagement with advanced mathematics.
- This gives me a practical understanding of what students need.
- **My Commitment to You:** Throughout this course, I will not only teach the mathematics but also **share insights** on:
  - How to prepare effectively for technical interviews.
  - How to build and maintain a robust theoretical foundation for a successful research career.

Course Overview (1)

Linear algebra for AI:

- **Matrix operations:** Matrix multiplication, inverses, determinants, and their role in neural networks
- **Matrix decomposition:** QR factorization, eigenvalue decomposition, SVD decomposition, and their applications in dimensionality reduction
- **Vector Space and Norm:** Basis, orthogonality, and their applications in embeddings
- **Basics of Tensor:** Multidimensional arrays for deep learning

Course Overview (2)

Probability Theory for AI:

- **Probability distributions and statistical inference:** Common probability distributions and their usage in modeling data
- **Markov Chains:** Transition matrices, stationary distributions and their applications in page-rank
- **Stochastic processes:** Discrete and continuous-time Markov chains and their applications in diffusion models

Course Overview (3)

Optimization Theory:

- **Gradient methods:** Stochastic gradient methods, and their insights for training neural networks
- **Constrained optimization:** KKT Condition, duality, and their applications in SVM
- **Approximation Algorithm:** Convex relaxation for non-convex machine learning problems

Course Overview (4)

Information Theory:

- **Entropy & Mutual Information**: Measuring uncertainty and feature relevance
- **Comparing distributions:** KL-divergence, Wasserstein distance, and kernel-based distances to quantify the discrepancy of probability distributions
- **Cross-Entropy**: Loss functions for classification tasks.

Assessment Scheme

Component/Method	Weight
Assignments	40%
Midterm	25%
Final	30%
Participation	5%
Total	**100%**

- *For participation part, we will have weekly quiz every Thursday starting from the second week.*
- *The quiz will be released at the beginning of Thursday lecture time and you have 24 hours to finish it.*
- *The quiz typically has 1 or 2 multiple choice questions.*

Teaching Plan

Note: The schedule is subject to minor adjustments based on class progress

Week	Dates	Content / Topic / Activity
1	Sep. 1-7	General introduction, matrix operations, neural networks
2	Sep. 8-14	Matrix decompositions (QR factorization, eigenvalue decomposition, SVD
3	Sep. 15-21	Vector space, norm, basics of tensor, embedding
4	Sep. 22-28	Probability distribution and statistical inference
5	Sep. 29 - Oct. 12	Markov chains, transition matrices, stationary distributions
6	Oct. 13-19	Stochastic processes (discrete and continuous-time Markov chains),
7	Oct. 20-26	Stochastic gradient methods, training neural networks
8	Oct. 27 - Nov. 2	**Review and Midterm**
9	Nov. 3-9	KKT Condition, duality, SVM
10	Nov. 10-16	Convex relaxation for non-convex machine learning problems
11	Nov. 17-23	Introduction to information theory, entropy, mutual information
12	Nov. 24-30	KL-divergence, Wasserstein distance, and kernel-based distances
13	Dec. 1-7	Applications of information theory, review of class
14	Dec. 8	**Review of class**