Math 466 Mathematics of Machine Learning, Spring 2025

Class Location:	Gross 318
Class Time:	MW 8:30 AM - 9:45 AM
Instructor:	Jiajia Yu
Office Hours:	(Tentative) MW 10:00 AM - 11:30 AM or by appointment
Email:	jiajia.yu@duke.edu Please include Math466 in subject line

Course Overview

This course introduces commonly used machine learning techniques with a focus on understanding them mathematically, utilizing tools from calculus, linear algebra, analysis, and probability.

Learning Objectives

By the end of this course, students are expected to:

- (General) know the common concepts, techniques and problems in machine learning; understand the concept of sparsity and know why and how to promote sparsity in machine learning; understand the difference between parametric and non-parametric model.
- (Supervised learning) know and understand the mechanisms of linear, ridge, LASSO regression, logistic regression and support vector machine; understand the concept of regularization and why it is important in machine learning;
- (Unsupervised learning) know and understand the mechanisms of PCA, MDS, isomap, Laplace eigenmap and diffusion map.
- (Deep learning) know the basics of fully connected neural network and convolutional neural network.
- (Optimization) know how to determine the well-posedness of an optimization problem and how to find the solution; understand why convexity is important in optimization; know why and how to compare convergence rate of different algorithms.
- (Statistical learning) know why and how to develop concentration inequality, VC dimensions.

Textbooks

There are no required textbooks. As necessary, links to papers and other reading materials will be provided. Some general references are:

- (General, supervised learning) [HTF09] The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Springer (2009). ISBN 9780387848570.
- (General, deep learning) [GBC16] Deep Learning, by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Cambridge: MIT Press (2016). ISBN 9780262035613.
- (Unsupervised learning) [BSS23+] Mathematics of Data Science, by A. S. Bandeira, A. Singer and T. Strohmer. Book in preparation, 2023 draft.
- (Optimization theory) [BSS06] Nonlinear programming: theory and algorithms, 3rd edition, by Mokhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. John Wiley & Sons, Inc (2006). ISBN 9780471486008.
- (Optimization algorithm) [NW06] Numerical optimization, by Jorge Nocedal and Stephen J. Wright. Springer New York (2006). ISBN 9780387303031
- (Statistical learning) [V18] High dimensional probability: An Introduction with Applications in Data Science, by Roman Vershynin. Cambridge University Press (2018). ISBN 9781108231596.

Prerequisites

- Math: multivariable calculus, linear algebra, and probability. We will discuss most of the necessary mathematical techniques as we proceed through the semester. But basic knowledge of these three courses is needed to understand the lecture. Please finish the self-assessment on the course canvas page before class.
- Coding: at least one programming language, python or MATLAB is preferred. There will be programming questions in homework and/or reports. The instructor and TA/grader are not responsible for helping debugging students' programming

Course Policy

- Attendance: Attendance in class is a vital part of the learning process. Regular class attendance is strongly encouraged. It is the student's responsibility to keep informed of any announcements, syllabus adjustments, or policy changes made during scheduled classes.

- Late policy: Students have three free "late" days they can use on homework and project throughout the semester, with at most one used for any one homework. A late day is defined as any whole or partial day after the submission deadline. These free late days are to be used for minor illnesses, balancing other coursework, conflicts associated with traveling, problems with your computer, etc. After using up the three free late days, late homework will be penalized 10% per day, and no credit will be awarded once solutions are posted (which can be as soon as the next class). Homework submitted on the due date but after the time specified will be penalized 5%. All work must be submitted by the assigned final exam time. Late submissions may result in a complete loss of credit for the corresponding portion.
- Collaboration policy: Group work and collaborative efforts for homework and projects are encouraged in this course. However, each handed-in problem set, and report must be independent work. You should name the students or other people with whom you had significant discussions about the problems, if any, on your hand-in solutions for homework. You should present a complete written solution/code to each problem, in your own words, without reference to the written solution of any other person. Any written sources, such as books and online sources other than the course textbook, that contribute significantly to your understanding of the problem should also be cited. Homework and report credits will not be given in case of violating the policies.
- AI tools policy: Students are encouraged to explore AI tools in homework and projects while maintaining academic integrity. The use must be properly documented and credited. For example, text or answers generated using ChatGPT-3 should include a citation such as: "Chat-GPT-3. (YYYY, Month DD of query). "Text of your query." Generated using OpenAI. https://chat.openai.com/" Material generated using other tools should follow a similar citation convention. Please be aware that AI does not always give the right answer or properly cite the references. Students using AI tools have the responsibility to make sure their final hand in homework and project report do not violate the academic integrity and include students' independent understanding and intellectual contributions.
- Academic Integrity: Students are responsible for informing themselves of Duke's policies regarding academic integrity. Students found in violation of the code are subject to penalties ranging from loss of credit for work involved to a grade of F in the course, and possible risk of suspension or probation. The academic integrity policy will be enforced in all areas of the course, including homework, exams, and projects. If you have any questions concerning this policy before submitting homework or project reports, please ask for clarification.

Grading Policy

Letter grades will be computed from the semester average. Maximum lower bound cutoffs for A, B, C and D grades are 93%, 83%, 73% and 60% respectively. These bounds may be moved lower at the instructor's discretion.

- Homework (40%): Homework will be assigned by topic. Depending on the content of the lecture, the homework assignment may have a programming part, in which case both codes and a written summary of experimental results (preferably in latex/markdown or any typed-up format) are required. Homework is graded both for accuracy, clarity, and completeness. For some questions, e.g., those that are broken down into branches, partial credit may be granted depending on the quality and completeness of the handed-in solution. Efforts to give meaningful partial solutions are always encouraged, but the grader has the right to determine the points to be credited for the handed-in solutions.

- Midterm and final (15%+15%): We will have two written exams. The mid-term exam will be an in-class 75-minute closed-book exam and no use of computational aids is allowed. There is NO make-up exam except for university-approved excuses. The final exam will either be a closed-book, in-person exam during final exam week or an open-book, take-home exam to be submitted during final exam week.

- Course project (30%): The course project can be either (i) a review report on a selected topic, or (ii) a technical report on a small research project on machine learning or data analysis. You can select a topic related to the content of the class.

  - For the review report, you will summarize the main results in the field of study, organize and cite the related literature, and present the content in a clear, correct, and organized way.
  - For the technical report, you will design the content of your project, and present your results (theoretical, programming, or both) in the report.

  For either type of report, you will also prepare one slide for a lightning presentation in class. The evaluation will be based on topic selection (2%), proposal (4%), update (8%) the in-class presentation (8%) and the final written report (8%). For more details, please refer to the project guidance file. We will also discuss and provide more guidance in the class.

Topic and Schedule (Tentative)

- Introduction (1 lecture)

- Supervised learning (4-5 lectures)

  - Regression: linear regression, Ridge regression and LASSO.
  - Classification: logistic regression, SVM and kernel SVM, KNN. (Naive Bayes, Random forest)
  - Neural network for regression and classification.
  - Sparsity: l_1 norm and nuclear norm. (Sparse signal coding, low-rank matrix completion, matrix factorization)
  - Parametric vs Non-parametric models.

- Optimization (4-5 lectures)
  - Convex optimization: local and global minimizer, existence and uniqueness of minimizer, optimality conditions.
  - Gradient descent, subgradient and proximal gradient: convergence analysis, convergence rate.
  - Stochastic gradient descent and its variants.
- Unsupervised learning (6 lectures)
  - Dimension reduction and manifold learning: PCA, classical MDS, Isomap; metric and non-metric MDS; Laplace eigenmap, diffusion map
  - Clustering: k-means, spectral clustering.
  - Generative models: VAE, GAN and normalizing flows.
  - Graph. (PageRank)
  - Kernel method. (RKHS)
- Statistical learning (4 lectures)
  - Concentration inequality
  - VC dimension
- Select topics (if time permits)
  - Gaussian process
  - Optimal transport
  - Diffusion models

For the latest schedule, please check the canvas front page.

Week	Monday	Wednesday	Notes
1, 01/08	No class	Introduction	
2, 01/13, 01/15	Regression models	Classification models	
3, 01/20, 01/22	No class	Sparsity in machine learning	01/22 Drop/Add ends
4, 01/27, 01/29	Neural network for regression and classification	Neural network; Parametric vs non-parametric models	**01/30 Project topic selection due**
5, 02/03, 02/05	Optimization and convex analysis basics	Optimality conditions	**02/06 HW1 due**
6, 02/10, 02/12	Gradient descent and its convergence	Subgradient and proximal gradient	
7, 02/17, 02/19	Stochastic gradient and its variants	TBD (unsupervised learning overview and generative models)	**02/20 HW2 due**; 02/21 mid-grade due
8, 02/24, 02/26	SVD and PCA	MDS, classical MDS and Isomap	**02/27 Project proposal due**
9, 03/03, 03/05	Graph, PageRank; Mid-term review	Mid-term exam, cover HW1,2 and PCA	
10, 03/10, 03/12	No class	No class	
11, 03/17, 03/19	k-means clustering and spectral clustering	Laplace eigenmap and diffusion map; kernel method	**03/20 Project update due**
12, 03/24, 03/26	Concentration inequality	Concentration inequality	**03/27 HW3 due**; 03/26 last day to withdraw
13, 03/31, 04/02	VC dimension	VC dimension	
14, 04/07, 04/09	Select topic	Select topic	**04/10 HW4 due**
15, 04/14, 04/16	Select topic	Select topic	
16, 04/21, 04/23	Presentation	Presentation	
04/24-04/27			**04/24 Project report due**
04/28-05/03			Final exam