Data Mining

Modern Machine Learning and Probabilistic AI

Course Number: 16:960:588
Credits: 3
Semester: Fall 2026
Instructor: Prof. Yifan Hu
Rutgers University

Course Description

This course provides an introduction to modern machine learning and data mining, with an emphasis on the probabilistic foundations and techniques underlying contemporary artificial intelligence.

We will study modern approaches to probabilistic modeling, uncertainty quantification, deep learning, generative modeling, and representation learning. Throughout the course, we will emphasize connections between mathematical and statistical principles and current developments in generative AI, multimodal learning, and foundation models.

A second objective of the course is to train students to use modern AI tools, especially large language models (LLMs), **effectively, critically, and responsibly** in technical problem solving. The course therefore evaluates both traditional mathematical understanding and the ability to work productively with AI-assisted workflows.

Learning Objectives

By the end of the course, students should be able to:

- understand the probabilistic foundations of modern machine-learning methods;
- formulate and analyze models involving uncertainty and probabilistic inference;
- understand the main ideas behind Gaussian processes, Bayesian optimization, variational inference, and Bayesian deep learning;
- understand modern generative and representation-learning methods, including VAEs, diffusion models, self-supervised learning, and contrastive learning;
- understand how representation learning connects to multimodal learning and foundation models;
- use large language models as tools for mathematical reasoning, coding, literature exploration, and technical problem solving;
- critically evaluate AI-generated arguments and solutions rather than treating model outputs as authoritative; and
- conduct an open-ended technical project by identifying important difficulties, developing appropriate approaches, and iteratively investigating the problem.

Assessment

The **base course grade is 100%**, consisting of:

- **Midterm Examination: 40%**
- **AI-Assisted Course Project: 60%**

Students may additionally earn **up to 20% bonus credit** through challenging homework questions.

Midterm Examination — 40%

Date: October 19, 2026

The midterm consists of two components:

Part I: Closed-Book Examination — 2 hours

Students solve mathematical and conceptual problems without access to books, notes, LLMs, or other external resources. This component evaluates individual understanding of the core concepts and techniques developed in the course.

Part II: AI-Assisted Problem Solving — 1 hour

Students may use LLMs to:

- revisit problems they were unable to complete during Part I;
- diagnose and correct errors in their reasoning; and
- solve additional, more challenging problems designed for AI-assisted work.

The objective is not merely to obtain answers from an AI system. Students are expected to formulate useful queries, evaluate AI-generated reasoning, identify mistakes, and produce correct final solutions.

AI-Assisted Course Project — 60%

The project is the primary assessment component of the course.

Students will work individually or in groups of **at most four students** on a substantial technical problem. Modern AI tools should be used actively as part of the project workflow.

Possible projects include:

- investigating an open question suggested by the research literature;
- reproducing, testing, or extending a recent machine-learning method;
- developing a new modeling or algorithmic approach to a technical problem;
- conducting a systematic empirical investigation; or
- constructing an end-to-end pipeline connecting data, modeling, prediction, and decision-making.

The project is intentionally open-ended. **A project will not be evaluated simply according to whether the original problem is ultimately solved.** Instead, evaluation will emphasize:

- understanding and formulation of the problem;

- identification of the main technical or empirical difficulties;
- quality of reasoning and methodology;
- appropriate and critical use of AI tools;
- concrete technical progress and milestones;
- ability to evaluate unsuccessful approaches and revise the strategy;
- clarity of communication; and
- depth of understanding demonstrated in the final work.

The final written product will be a **technical blog post** explaining the problem, the investigation, the main findings, difficulties encountered, and lessons learned.

Projects will also be presented during the final course meeting through **project presentations and a poster session**.

Bonus Homework — Up to 20%

Throughout the semester, challenging problems will be provided as optional bonus homework.

These problems may be substantially harder or more open-ended than standard exercises. Students are encouraged to explore them using any appropriate resources, including mathematical derivations, computation, literature searches, discussions, and AI tools.

The purpose of these problems is to encourage exploration rather than routine completion.

Policy on Large Language Models and AI Tools

The use of LLMs is an integral part of this course.

Students are expected to learn how to use modern AI systems effectively while remaining responsible for the mathematical and factual correctness of their work.

LLMs may be used for course projects, bonus problems, programming, exploration, and other course activities unless explicitly stated otherwise.

There is one important exception:

LLMs and other external resources are prohibited during the two-hour closed-book portion of the midterm examination.

They are explicitly permitted during the one-hour AI-assisted portion of the midterm.

Whenever AI tools materially contribute to submitted work, their use should be **clearly disclosed**. Students remain responsible for verifying AI-generated claims, calculations, references, code, and arguments.

The central principle of the course is:

> AI can assist the reasoning process, but responsibility for the final reasoning remains with the student.

Tentative Schedule

Date	Topic
Tue., Sep. 8	**Introduction to Modern Machine Learning and Probabilistic AI**; Bayesian Linear Regression
Mon., Sep. 14	**Gaussian Processes**
Mon., Sep. 21	**Bayesian Optimization**
Mon., Sep. 28	**Variational Inference**
Mon., Oct. 5	**Bayesian Deep Learning**
Mon., Oct. 12	**Active Learning and Uncertainty Quantification**
Mon., Oct. 19	**Midterm Examination:** 2-hour closed-book exam + 1-hour AI-assisted problem solving
Mon., Oct. 26	**LLMs for Problem Solving:** Hands-on Tutorial
Mon., Nov. 2	**Generative Models:** Latent-Variable Models, VAEs, and Diffusion Models
Mon., Nov. 9	**Representation Learning I:** Embeddings, Autoencoders, and Learning Useful Representations
Mon., Nov. 16	**Representation Learning II:** Self-Supervised Learning and Predictive Representations
Mon., Nov. 23	**Contrastive Learning:** Objectives, Negative Sampling, and Modern Methods
Mon., Nov. 30	**Multimodal Representation Learning and Foundation Models**
Mon., Dec. 7	**Project Presentations and Poster Session**

Calendar note: Monday, September 7 is Labor Day. Rutgers designates **Tuesday, September 8 as a Monday class day**, so the course meets on Tuesday that week.

The schedule is tentative and may be adjusted according to the pace of the course and student interests.

Course Philosophy

Modern machine learning increasingly involves two complementary forms of expertise.

The first is the ability to understand models, assumptions, mathematical principles, and algorithms independently. The second is the ability to use increasingly capable AI systems to accelerate exploration and problem solving without surrendering judgment to those systems.

Accordingly, this course deliberately trains and evaluates **both capabilities**. Students will first develop the foundations needed to reason independently and will then learn how those foundations can be combined with modern AI tools to address substantially more challenging problems.

The goal is not simply to learn a collection of machine-learning methods, but to develop the ability to **understand, investigate, and solve new problems in an AI-assisted research environment**.