

Introduction to Machine Learning

DSC 240 Winter 2025, UC San Diego

Instructor: Prof. Yu-Xiang Wang

Lectures: Tuesday and Thursday 9:30 AM - 10:50 AM at CENTR 105

Course website: <https://cseweb.ucsd.edu/~yuxiangw/classes/DSC240-2025Winter/>

Discussion Forum: https://piazza.com/ucsd/winter2025/dsc240_wi25_a00/home
(We use Piazza for Q&A, discussion, and most announcement!)

Gradescope: <https://www.gradescope.com/courses/941346>
(We will be using the gradescope for submitting code and project reports.)

Canvas: <https://canvas.ucsd.edu/courses/61389> The Canvas site is active, but it is only reserved for communications that are restricted to people who are officially registered.

1 Overview

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data," in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task (Wikipedia).

We will cover the following topics:

This is an introductory course in machine learning offered as part of a computer science and data science curriculum, so we are not primarily concerned with the approaches of philosophy, cognitive science, etc., although we will come back to them from time to time. We will cover the following technical topics:

- Linear Algebra needed for machine learning
- Machine Learning Basic Concept
- Classification
- Regression
- Ensemble Methods
- Optimization
- Modeling principles
- Neural Networks, Deep Learning

2 What you will learn?

By the end of the course, you will understand what ML is all about and what it has contributed, and may contribute, to computing. You will have a working knowledge of the basic tools of machine learning, which are applicable to a wide range of computing problems. You will be able to solve problems using classification and clustering. You will have experience building models that, to some degree, can learn from experience.

This is not primarily a programming course - that is, the main goal is to learn the concepts, not to learn a language or particular programming techniques. However, coding examples of the concepts is the best way to demonstrate (and facilitate) your knowledge of them. Good programming practices (proper file structure, comments, etc.) are expected.

Quite different from a typical course you might have taken, you will learn how to apply what you have learned before in mathematics, statistics and coding skills to solve problems. You will get some practices in simple mathematical derivations and translating the mathematical equations into codes in Python / Numpy.

3 Prerequisites

The formal prerequisite is DSC 210, 212. However, as this is a graduate level course, there is no hard pre-requisite, but students entering the class are expected of the following.

- Students are expected to be able to follow rigorous mathematical arguments and perform derivation and proofs using notations, definitions and theorems from calculus, linear algebra, probability and statistics.
- Students are expected to have working knowledge of basic algorithms and data structures, and could write simple code in Python / Numpy.
- If you have taken a course in graduate-level mathematical statistics, statistical machine learning, linear / convex optimization, or are familiar with topics such that concentration inequalities, then you are very well-prepared to enjoy the course.

You will be asked to complete HW0 (self-assessment) in the first week.

4 Textbooks

We will not follow the pedagogy of any text book. The course will be self-contained. However, it will help you with your learning if you complete the readings assignments on the course website from the following references books.

- “Pattern Recognition and Machine Learning” by Chrs Bishop
- “Dive Into Deep Learning”, by Zhang, Lipton, Li and Smola.

Both books are open access.

5 Assignments and Grades

The grades will be based on the following breakdowns.

- 30% Coding projects
- 30 % Homework assignments

- 20% Midterm
- 20% Final Exam
- -5% if you don't complete HW0 on time.

Homework will include 4 assignments and 3 coding assignments. Unless otherwise instructed, all assignments must be submitted on the due date. No late submission is allowed — but only your best three homeworks will be included in your final evaluations.

You are encouraged to discuss homework assignments with classmates at a general level. However, you shall not share answers/code or collaborate on solutions unless otherwise directed to do so. All work turned in must be completely your own, including programming assignments. (See the Policy on Academic Integrity, below.)

There will be no makeup exams except for extraordinary circumstances.

6 Exams

The midterm and final will test your understanding of the materials from the lectures and discussion sections. They will be easier and shorter than the homeworks.

7 Logistics

Lectures: The instruction will be in person.

Office hours: Office hours are announced on the course website and on Piazza.

Attendance policy: The attendance to lectures and weekly discussion sections are *required*.

Late submission policy: No late days allowed. The best 3 out of 4 homeworks are counted. Since this is a grad level class, if you miss more than one homework, you won't get 0. Instead, your missing homework score will be *replaced* by the lower of your two exam scores.

8 Policy on Academic Integrity

Please read this section carefully.

The university, the department, and this instructor all take the issue of academic integrity **very** seriously. A university requires an atmosphere of mutual trust and respect. While collaboration is an integral part of many scholarly activities, it is not always appropriate in a course, and it is never appropriate unless due credit is given to all participants in the collaboration.

Here are some examples:

- Allowed: Discussion of lecture and textbook materials
- Allowed: Discussion of how to approach assignments, what techniques to consider, what textbook or lecture material is relevant
- Allowed: Collaboration on homeworks. You need to declare your collaborators, and describe what you get from the person who helped you. Each student still needs to complete their own homework and write their own solutions.
- Allowed: Refer to online resources (including ChatGPT and other language models). Please declare sources.

- Not allowed: Turning in someone else's work as your own, even with that person's permission.
- Not allowed: Allowing someone else to turn in your work as his or her own.
- Not allowed: Turning in work without proper acknowledgment of the sources of the content (including ideas) contained within the work.

For some views on academic integrity at UCSD see the Academic Integrity (AI) Office.

Summary: Academic integrity is absolutely required - dishonesty (cheating, plagiarism, etc.) benefits no one and hurts everyone. Violations of these honor codes on academic integrity will be reported to the Office of Student Conduct. If you find yourself in such a situation, please contact the instructor. If you are not sure whether or not something is appropriate, please ask the instructor.

9 Code of conduct

The University of California, San Diego has a general code of conduct for all students published here: <https://sage.ucsd.edu/>

The University's commitment to Diversity, Equity and Inclusion is published here: <https://ucsd.edu/about/principles.html>.

10 Students with Disabilities

If you are a student with a disability and would like to discuss special academic accommodations, please contact the instructor. In addition, students with temporary or permanent disabilities are referred to the Office of Students with Disabilities (OSD) at UCSD. OSD will arrange for special services when appropriate (e.g., facilitation of access, note takers, readers, sign language interpreters). Please note that it is the student's responsibility to communicate his or her special needs to the instructor, along with a letter of verification from OSD.

11 Copyright of course materials

My lectures and course materials, including presentations slides, written notes, homework assignments and similar materials, are protected by U.S. copyright law and by University policy. I am the exclusive owner of the copyright in those materials I create. You may take notes and make copies of course materials for your own use. You may also share those materials with another student who is enrolled in or auditing this course. You may not reproduce, distribute or display (post/upload) lecture notes or recordings or course materials in any other way — whether or not a fee is charged — without my express prior written consent. You also may not allow others to do so.

If you do so, you may be subject to student conduct proceedings under the UCSD Student Code of Conduct.