

COMS 4771 Fall 2023 Syllabus

1. Description

COMS 4771 is a graduate-level introduction to machine learning. The course covers basic statistical principles of supervised machine learning, as well as some common algorithmic paradigms.

2. Course information

- **Lecture times:** Tue/Thu 1:10pm–2:25pm (Section 1), 2:40pm–3:55pm (Section 2)
- **Lecture venue:** 451 CSB
- **Instructor:** Daniel Hsu

See course website <https://www.cs.columbia.edu/~djhsu/coms4771-f23/> for up-to-date information and announcements.

3. Learning goals

- Apply mathematical and statistical principles to understand and reason about machine learning problems and algorithms.
- Apply algorithmic techniques to construct machine learning algorithms.
- Apply machine learning algorithms in some simple application domains (e.g., text classification).

4. Prerequisites

There are several prerequisites for this course.

- You must be fluent in multivariate calculus, linear algebra, and basic probability, all at the undergraduate level.
- You must be comfortable with writing code to process and analyze data in Python.
- You must be familiar with basic algorithmic design and analysis.
- You must have general [mathematical maturity](#).

[A more detailed list of topics is available here.](#)

[Review notes for some of the prerequisites are available here.](#)

Some online resources for course prerequisites are as follows.

- Multivariable calculus: [MIT Open Courseware](#)
- Linear algebra: [Hefferon's Linear Algebra](#), [COMS 3251 lecture notes](#)
- Probability: [Grinstead and Snell's Introduction to Probability \(local copy with hyperref\)](#)
- Programming in Python with NumPy: [Python Tutorial](#), [NumPy: the absolute basics for beginners](#)

5. Topics

The tentative list of topics is as follows.

- Nearest neighbors
- Classification using generative models
- Statistical models for prediction
- Decision tree learning
- Linear regression
- Linear classification
- Feature maps and kernel methods
- Inductive bias and regularization
- Dimension reduction
- Optimization by gradient methods
- Optimization problems and duality
- Multi-class linear prediction
- Calibration and bias
- Generalization theory
- Neural networks

[A brief overview of the course topics is available here.](#)

6. Requirements

Students are expected to attend lectures, complete required reading assignments, complete homework assignments, and take in-class exams.

7. Readings

Lectures will be mostly self-contained; required reading assignments will be posted alongside the lecture schedule. Pointers to optional reading from (some of) the following texts will also be given.

- [A Course in Machine Learning](#) (CML) by Daumé
- [Pattern Classification](#) (PC) by Duda, Hart, and Stork
- [Patterns, Predictions, and Actions](#) (PPA) by Hardt and Recht
- [Mathematics for Machine Learning](#) (MML) by Deisenroth, Faisal, and Ong
- [The Elements of Statistical Learning](#) (ESL) by Hastie, Tibshirani, and Friedman

All of these texts are available online, possibly through [Columbia University Libraries](#).

8. Assessment

The overall course grade is comprised of the following.

- homework assignments (40%)
- in-class exams (60%) on **Thursday, October 26** and **Thursday, December 7**

There are no make-up assignments/exams available.

- Do not enroll in the course if you do not expect to be able to take the exams on the scheduled dates during lecture.
- Extensions on homework assignments are only granted for medical or family emergencies. Late homework assignments are not accepted for any other reason.
 - Requests for such extensions must be submitted in writing to the instructor **by your advising dean or your academic advisor** after you have explained your situation and provided any necessary documentation to them.
- If you miss an exam due to a medical or family emergency, you may be granted an “incomplete” for the course, which you can complete by taking a comparable exam in a future offering of this course, or you may “withdraw” from the course.

Overall course grades will be curved.

9. Columbia video network

CVN students may be subject to other policies related to the video network format; please contact CVN administration for details.

10. Disability services

If you require accommodations or support services from [Disability Services](#), please make necessary arrangements in accordance with their policies within the first two weeks of the semester.

11. Academic rules of conduct

You are expected to adhere to the [Academic Honesty policy of the Computer Science Department](#), as well as the following course-specific policies.

11.1. Policies on plagiarism

Any work you submit must be written completely in your own words.

11.2. Policies on collaboration

Homework assignments must be completed **individually** or in **groups of two or three**. All students must abide by the following rules regarding collaboration.

- If you work in a group, all members of the group must contribute to the solution of each problem.
- You may not look at or take another student's/group's homework write-up/solutions/code (whether partial or complete).
- You may not show or give your homework write-up/solutions/code (whether partial or complete) to another student/group.
- Discussions between different homework groups must not go as far as one group telling the other group how to solve a problem.

Exams must be completed individually. **Collaboration or discussion between students on exams is NOT PERMITTED.**

11.3. Policies on the use of outside references and resources on homework assignments

Outside reference materials and resources (i.e., texts and sources beyond the assigned reading materials for the course) may be used on homework under the following rules.

- **Any reference or resource used must be acknowledged and cited in the write-up.**
- **Explicitly searching or querying the internet/large language models/“AIs”/etc. for answers or hints on homework assignments is not permitted.**
- If you *inadvertently* come across a solution to (or substantial hint about) a problem, you must:
 - acknowledge this source and document the circumstance in your homework write-up;
 - produce a solution without looking at the source; and
 - as always, write your solution in your own words.
- If you have already seen one of the homework problems before (e.g., in a different course), please re-solve the problem without referring to any previous solutions.
 - In your write-up, please also indicate that you had seen the problem before. (You won't lose any credit for this; it would just be helpful for us to know about this fact.)

Outside references and sources **CANNOT** be used on exams.

You are welcome to use resources found in the library, on the internet, embedded in large language models, etc., to help you learn about the course topics. But please note that these resources may contain (often very subtle) inaccuracies, and the course staff may not be able to help you discern whether a particular resource is correct or not. Also, these resources may not be used on exams.

11.4. Consequences of policy violations

Violation of any portion of these policies will result in a penalty to be assessed at the instructor's discretion (e.g., a zero grade for the assignment in question, a failing letter grade for the course), even for a first offense.

12. Getting help

You are encouraged to use office hours and message board to discuss and ask questions about course material and reading assignments, and to ask for high-level clarification on and possible approaches to homework problems. If you need to ask a detailed question specific to your solution, please do so on the message board and mark the post as “private” so only the instructors can see it.

Questions, of course, are also welcome during lecture. If something is not clear to you during lecture, there is a chance it may also not be clear to other students. So please raise your hand to ask for clarification during lecture. Some questions may need to be handled “off-line”; we'll do our best to handle these questions in office hours or on message board.