

CS 181 – Machine Learning

Instructor: Finale Doshi-Velez
Lectures: Monday and Wednesday, 9-10:30am
Location: Northwest Building B103
Section: 1:30-2:30pm Monday, 4:30-5:30pm Monday, 10:30-11:30am Tuesday,
3-4pm Tuesday, 4:30-5:30pm Wednesday
URL: <https://harvard-ml-courses.github.io/cs181-web/>

1 Course Description

This course provides a broad and rigorous introduction to machine learning, probabilistic reasoning and decision making in uncertain environments. We will discuss the motivations behind common machine learning algorithms, and the properties that determine whether or not they will work well for a particular task. You will derive the mathematical underpinnings for many common methods, as well as apply machine learning to challenges with real data. In doing so, our goal is that you gain a strong conceptual understanding of machine learning methods that can serve as your map and lodestar as you pursue future theoretical and practical directions.

Where CS181 fits with other ML/AI options. As noted above, our goal is to combine mathematical derivation and coding assignments to lead to a strong and rigorous conceptual grounding in the basics of machine learning (e.g. being able to reason about how different methods should behave in different circumstances). Students interested primarily in theory may prefer Stat195 and other learning theory offerings. Students interested primarily in practice may prefer CS109, CS10, and other data science offerings. Students interested primarily in going deep into a particular machine learning or AI topic, rather than a survey, may prefer CS28x and other graduate seminars.

Prerequisites The material is aimed at an advanced undergraduate level. Students should be comfortable with writing non-trivial programs (e.g., CS 51 or equivalent). All staff-provided code will be in Python, and the staff will not support questions in any other language. Students should also have a background in basic probability theory (e.g. STAT 110 or equivalent), and some level of mathematical sophistication, including calculus and linear algebra (e.g., Math 21a and 21b or equivalent).

That said, we note that CS181 only requires portions of all of these courses. Every year, many motivated students are successful in CS181 without all of these prerequisites. I want to continue to welcome those hard-working students who are willing and able to independently fill in any gaps that they might have in their knowledge. Part I of *Math*

for *Machine Learning*, available at <https://mml-book.com> may be a useful resource for mathematical background (specifically Sections 2.1-2.6; 3.1-3.5; 4.1-4.2; 5.1-5.6; 6.3).

Regardless of your background, it is YOUR responsibility to learn any prerequisite material on your own. The course staff will not be responsible for teaching basic coding, matrix manipulation, etc.

2 Course Logistics

2.1 Lecture, Section, Office Hours

Team The CS181 team consists of a course instructor (responsible for the content and grade assignments), a course manager (responsible for the logistics of the course—posting things, managing exceptions, etc.), and a large staff of TFs lead by a head-TF (responsible for section, most of the office hours, and grading).

Lectures Lectures will be used to introduce new content as well as explore the content through conceptual questions. They will occur on the board and will not be recorded. During lecture, I may remind the class about upcoming deadlines, clarify points in the homework, and respond to questions about upcoming assignments and midterms. Not all of these interactions may make it to the class announcements. Thus, if you miss a lecture, we strongly recommend asking friends about anything that was mentioned.

Sections Sections will employ a flipped classroom format, in which students will work on questions that will be good preparation for both homeworks and midterms. The teaching staff will introduce the questions, assist students in solving them, and wrap up with the solutions. These solutions will be posted. *The staff may post additional practice questions or pointers to other practice resources. We do not guarantee solutions for these additional resources.*

While sections attendance is optional, attendance will be taken and strong participation is one way we may choose to decide letter grades for students who are near a boundary. Section is also a great place to find study partners!

Office Hours The office hour times will be posted on the website. Please make use of office hours! In addition to getting questions answered by the staff, office hours are also a great place to find study partners.

2.2 Materials and Resources

Textbook There is no official textbook for the course. We have a version 0.0 of course notes available at <https://github.com/harvard-ml-courses/cs181-textbook>. We emphasize that these notes are due to the awesome effort of a past CS181 student who

decided to create a course textbook as a senior thesis. *Expect errors*. When you spot errors, be a good citizen and put in a pull request.

Course Website The course web site will be used for posting section notes and links to assignments, and includes pointers to other resources we'll use, including Piazza and GradeScope.

GradeScope GradeScope will be used for submitting assignments, posting grades.

Piazza The piazza site for the course will be used for three purposes:

- **Content** questions are technical questions posted to other students. (Please keep in mind collaboration policies when asking about code or solutions.) The course staff will *not* be responsible for immediate responses but will answer when possible; technical questions to TFs should be brought to office hours (or to section when appropriate).
- **Clarification** questions are posts about logistical details (Is there really class on XYZ holiday or is that a mistake?) or questions about homework phrasing or typos (Should question 1a of the homework be asking for the integral of x , not y ?). We will make every effort to respond to these questions as quickly as possible. Tag clarification questions as “clarification.”
- **Private Message** There may be times when you wish to send a private messages. These may include procedural things such as requests for additional time on midterms, additional late days, regrades; you may also have other concerns that you wish to share. We ask you to send those as *private* messages on Piazza with the *appropriate tag*: Regrade, Extension, Special Midterm, and Other. We will be using these tags to make sure that the right people get your request. We will *only* consider procedural requests via this form (this is as much for you as it is for us: we want to have a record of all requests to hold us accountable).

Please note that Piazza is not a formally secure, private, or confidential form of communication, and what you send may be seen by the entire course staff. If you have a sensitive concern for which such a medium is not appropriate, then please catch Finale in person before or after class or catch Finale/TFs at their OHs. (Note, after a conversation, we may still ask you to create a generic request via Piazza so we have record of a promised extension, etc.) Email should be used sparingly if at all.

3 Requirements and Grading

The main grading components of the course will be the six homeworks (10% each) and two midterms (20% each, on March 11 and April 29). Participation (in section, office hours, lecture) may be used to bump up (but not down) grades that end up near a letter-

grade boundary.

Grading Errors If you believe there has been a grading error, . However, please note that a) we will regrade the entire assignment, which may result in your total grade going up or down and b) we will only allow one regrade request per problem.

Special Circumstances More broadly, we understand that sometimes life throws a set of circumstances that impact your performance in the course. Should this happen to you, please let me or the course manager know so we can help determine a plan to navigate a tough situation.

3.1 Homeworks

The homeworks help you practice the core concepts. These involve components that are theoretical and conceptual, and also require some programming. Homework solutions must be submitted in LaTeX and will be returned with grades and solutions. Due to the volume of the grading, it may not always be possible for the staff to provide detailed feedback. It is your responsibility to look at the solutions, identify gaps, and come to office hours to fill those gaps.

Collaboration Policy You may work with others, but your write-up must be entirely written by yourself in your own words. You may help each other debug code, but again, the code must be written by you. Include the names of anyone you worked with in your write-up. We encourage you to spend time thinking about and understanding the homeworks on your own before collaborating with others to practice for the midterms. *It is an honor code violation to copy parts of another person's assignment or jointly type up an assignment.*

You can make use of textbooks and online sources to help in answering questions but you must cite your sources (and you should be ready to explain your answer to a member of the teaching staff.) *It is also an honor code violation to look up solutions from the internet or other sources (e.g. friends from previous years).*

Late Days Policy Homeworks should be submitted electronically by midnight on the due date, via the GradeScope course website. This is a strict deadline, enforced by the site, so submit early enough that you don't accidentally discover that your clock is slow.

You have **five late days** that can be used for homework assignments. Up to **two late days** can be used on any assignment. Start early and plan ahead! *It is an honor code violation to look at the answer key if you haven't yet turned in your assignment.*

The staff will consider giving 50% credit to assignments turned in past their late days at their discretion. *It is almost always in your interest to turn in partial or late homework rather than not turning in any homework at all.*

3.2 Midterms

Midterms are a chance to demonstrate your learning individually. Section problems, homeworks, and concept questions are all great starting points for study. You will be allowed to bring in one sheet of 8.5 by 11 paper, front and back, as notes and the staff may, at their discretion, provide additional formulas. *Your notes are the only resource you are allowed to use during the midterm. Using any other resources, as well as sharing information about the midterm to students who are taking it at a different time, is an honor code violation.*

4 Philosophy

The goal of this course is to instill a strong technical background for you to responsibly apply machine learning in the world. Thus, in addition to the derivations and the practicals, each class will include a story about realworld applications of machine learning and one full lecture and part of an assignment will be devoted to the ethical implications.

To be blunt: Given the the increasing use of machine learning systems, the users and developers of these systems must hold themselves to high professional and ethical standards. One can cause real harm by pursuing a good cause via poor engineering choices. Quoting one of our favorite superheroes: with great power (to run any kind of analysis) comes great responsibility (to do it properly)!

Relatedly, we expect all participants in this course—instructors, staff, students—to be committed to a open, professional, and inclusive environment. Just like the maths, these qualities take cultivation and effort. I will start with the premise that we're all decent people trying our best and expect you to do the same. I welcome constructive feedback to improving the course environment.