CS 181 – Machine Learning

Instructors:	Finale Doshi-Velez
Lectures:	Monday and Wednesday, 9-10:30am
Location:	Northwest Building B103
Section:	Monday 4-5pm; 5-6:30pm (extended); Tuesday 4-5:30pm (extended); Wednesday 4-5pm, 5-6pm
URL:	`https://harvard-ml-courses.github.io/cs181-web/`

Basics

Course Description

This course provides a broad and rigorous introduction to machine learning, probabilistic reasoning and decision making in uncertain environments.

- **Who?** The course may be of interest to undergraduate students in computer science, applied mathematics, sciences and engineering, and lower-level graduate students looking to gain an introduction to the tools of machine learning and probabilistic reasoning with applications to data-intensive problems in the applied sciences, natural sciences and social sciences.
- **What?** This course will provide an overview of three major areas in machine learning: supervised learning, unsupervised learning, and reinforcement learning. These areas roughly correspond to: Given a set of inputs and outputs, how can we make predictions about future outputs? What are the major patterns in our data? How can we design agents to make decisions, rather than just predictions?
- **How?** Our approach to these questions will be conceptual, theoretical, and practical. 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. We will follow up by deriving the mathematical underpinnings for many common methods, as well as applying machine learning to challenges with real data.
- **Why?** Quoting one of our favorite superheroes: with great power (to run any kind of analysis) comes great responsibility (to do it properly)!

Prerequisites

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). We note that CS181 only requires portions of all of these courses, and many motivated students are successful in CS181 without all of these prerequisites. *Thus, we will not enforce any prerequisites, but it will by YOUR responsibility to learn any prerequisite material on your own. The course staff will not be responsible for teaching basic coding, matrix manipulation, etc.*

Course Logistics

Textbook The required textbook for the course is *Pattern Recognition and Machine Learning* by Bishop. We will post recommended readings on the web page for each lecture.

Course Website

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

Canvas The Canvas site will be used for submitted psets and for grading purposes.

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.
- **Private** messages can be used to ask sensitive, non-technical questions that are not appropriate for the entire class but are appropriate for the entire course staff. Depending on the content, you may also wish to catch Finale in person before or after class or catch Finale/TFs at their OHs. Email should be used sparingly.

Office Hours

Starting the week of 1/29, 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 and teammates!

Sections

Sections are optional problem-solving sessions, and another great place to find study partners! 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. Extended sections will cover the same material but with more time for students to work on solutions. *The staff may post additional practice questions or pointers to other practice resources. We do not guarantee solutions for these additional resources.*

Requirements and Grading

There are four practicals, making up 30% of the final grade. There are five individual homeworks, making up 30% of the final grade. There are two midterms, each counting for 20% of the final grade. The first midterm is on March 5 during class time, the second midterm will occur on the last day of class (April 25) and will focus on the material in the second half of the course.

Practicals

Practicals will be done in teams of three. The course staff can help you to find partners, or you can also seek them via Piazza. The goal of the practicals in the course is twofold: to help you master the technical material, and to show you how the techniques we are learning can be used to build powerful and cool applications. We call them "practicals" rather than "homeworks" to make the point that they are meant to be open-ended and encourage creativity. Each practical will usually be due two weeks after being handed out. Each practical will generally be centered around a particular methodology and task and involve programming. You will need to consider some conceptual issues, write a program to solve the task, and evaluate your program through experiments to compare the performance of different algorithms and methods.

Your main deliverable will be a short report. You'll be assessed on effort, the sophistication of your technical approach, the clarity of your explanations, the evidence that you present to support your evaluative claims, and the performance of your implementation. A high-performing approach with little explanation will receive little credit, while a careful set of experiments that illuminate why a particular direction turned out to be a dead end may receive close to full credit.

Collaboration Policy Outside your group, you may not share write-ups or code. You may discuss your high-level approach (we focused on feature engineering) and high-level results (we got 90% test accuracy). You must cite any sources used (online or otherwise.)

Homeworks

The homeworks help you practice the core concepts. These involve components that are theoretical and conceptual, and also require some programming. There will be five of these homeworks. Homework solutions must be submitted in LaTeX.

Self-Grading Policy Once the submission deadline has passed, students will receive the answer key for the assignment. Students will then be responsible for using this answer key to submit their own scores; this process is there to ensure that students check their understanding and to provide very rapid feedback. The staff will spot-check submitted scores for accuracy. *It is a honor code violation to look at the answer key if you haven't yet turned in your assignment (e.g. using late days). It is also an honor code violation to look at an answer key from a previous iteration of the class.*

Collaboration Policy You may not share write-ups or code. If you brainstorm with others, you must note their names 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. *Other than the preclusion of using an answer key from a previous year or another student's answer from this year or a previous year, 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.)*

Late Policy

Homeworks and practical writeups should be submitted electronically by 5:00pm on the due date, via the Canvas course website. This is a strict deadline, enforced by the Canvas site, so submit early enough that you don't accidentally discover that your clock is slow. Only one submission is required for each team of students for the practicals.

You have **five late days** that can be used for homework assignments and the write-ups on the practicals (note that the Kaggle competitions will close a day before the practicals are due, and that will not be extended). Start early.