Course Syllabus

Jump to Today ✎ Edit

CS 158: Machine Learning

Harvey Mudd College, Fall 2017

Who, Where, When

Instructor : Yi-Chieh (Jessica) Wu, yjw AT cs DOT hmc DOT edu, Olin 1277
When : MW 9:35-10:50am
Where : Shanahan 1480
Grutors : Teerapat Jenrungrot, Herrick Fang, Daniel King, Varsha Kishore, Marina Knittel

Course Description

Machine learning has been essential to the success of many recent technologies, including autonomous vehicles, search engines, genomics, automated medical diagnosis, image recognition, and social network analysis, among many others. This course introduces the fundamental concepts and algorithms that enable computers to learn from experience.

Learning Goals

Upon completing this course, students should

- [theory goal] be able to summarize the theoretical and computational foundations of core algorithms in the field of machine learning (at a level to do further research, e.g. in graduate school)
- [application goal] be able to apply core machine learning algorithms to modern problems and evaluate their performance (to become more sophisticated users of machine learning)

Prerequisites: CS70 (Data Structures and Program Development), HMC Math Core (Probability and Statistics, Multivariable Calculus, Linear Algebra). Previous experience with Python is strongly recommended. If you have not satisfied these prerequisites, you should speak with Prof Wu no later than the first day of class.

What This Course is Not

- a big data course -- We focus on the Math and CS foundations of ML and do not cover Databases, Map/Reduce, Hadoop, Spark, or Amazon Web Services.
- a "domain" (app-based) course -- We will use examples from a wide range of domains (e.g. NLP, Vision, CompBio) rather than developing specialty in a single field.
- a neural networks course -- Neural networks perform very well but are difficult to interpret, which is often a key factor when deciding among algorithms. Students interested in neural networks might try CS 152.

Course Format

This course meets twice a week for lecture; students are expected to attend and actively participate. You should be ready to work at the beginning of each class, participate fully in all class activities, and provide feedback when appropriate. I have no problem with you using computers or tablets to take notes or consult reference materials during class. Tempting though it may be, please do not check e-mail or visit websites that are not relevant to the course during class. It is a distraction for you, me, and (more importantly) for your fellow classmates. Additional opportunities for earning participation points include contributing to Piazza threads.

Weekly coursework includes (required) problem sets and (optional) mini-quizzes. Students must complete these assignments individually unless otherwise noted. Problem sets typically include both written and programming exercises. For the written exercises, you can either type your solutions or submit a scan or photo (converted to pdf) of your handwritten solutions (make sure your handwriting is legible please). For the programming exercises, please type your solutions.

Additionally, there will a project in the second half of the course, one midterm, and one final.

Course Management Systems

We will use Canvas and Piazza (https://piazza.com/class/j5sqqbibihy7k6). You should have received an email invitation to both sites.

Course Schedule

Course Schedule

Getting Help

Instructor office hours (Olin 1277) : Tue 4-5pm, Fri 1:30-2:30pm, by appointment
Grutoring hours (Olin 1261 CS Workroom) : TBD

As you will discover, I am a proponent of two-way communication, and I welcome feedback during the semester about the course. I am available to answer student questions, listen to concerns, and talk about any course-related topic. Come to office hours! This helps me get to know you. Additionally, there are many more exciting topics that we will not have time to cover in-class -- you are welcome to stop by and chat! Whenever you e-mail staff, be sure to include the phrase "[CS 158]" at the beginning of the subject line and to use a meaningful subject line. Your e-mail will catch my attention, and I will respond more quickly if you do this. However, *we ask that you use Piazza rather than e-mail to communicate with the course staff so that we can answer your questions as quickly and efficiently as possible.* Furthermore, other students invariably have the same questions as you do, and posting to the forum encourages discussion.

- Post questions publicly as much as possible (you can set your posts to anonymous to your classmates and staff).
- Answer your classmates' posts. Answering questions is a great way to learn, and it counts for participation credit.
- Do not forget to search the message board before posting! Your question may have been already asked and answered.
- Be careful not to post "spoilers". If in doubt, post privately to staff.

Of course, if you have a conceptual question about the material or a relatively "big" question about a homework problem, you are encouraged to come and talk to one of us in person.

Resources

Textbooks

There is not one comprehensive machine learning textbook that covers all material from this course. Our class is based heavily on Andrew Ng's CS 229 course at Stanford, so his **Machine Learning Notes** (http://cs229.stanford.edu/materials.html) provide the best preview / review of course material. These are optional, and you should also note that Andrew Ng is a "chalkboard lecturer", so these notes are best-used as supplements to class presentation of topics.

For those who would like a gentler introduction, I recommend readings from the following textbooks and resources:

- **A Course in Machine Learning** (http://ciml.info/) by Hal Daumé III. (Available online)
- **Learning from Data** (http://amlbook.com/) by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin, AMLBook, 2012.
- **Machine Learning: The Art and Science of Algorithms That Make Sense of Data** (https://www.cs.bris.ac.uk/~flach/mlbook/) by Peter Flach, Cambridge University Press, 2012.
- **Statistical Data Mining Tutorials** (http://www.autonlab.org/tutorials/) by Andrew Moore. (Available online) [Tutorial slides so covers a wide range of topics but lacks depth of textbooks.]

For a more advanced treatment of machine learning topics, I recommend one of the following books:

- **Machine Learning** (http://www.cs.cmu.edu/~tom/mlbook.html) by Tom Mitchell, McGraw Hill, 1997. (On reserve in library)
  Pros: One of the "bibles" of ML.
  Cons: Lacks depth (compared to other "bibles"). Somewhat out-of-date (1997), with new edition forthcoming.
- **Pattern Recognition and Machine Learning** (http://research.microsoft.com/en-us/um/people/cmbishop/prml/) by Christopher M. Bishop, Springer, 2006. (On reserve in library)
  Pros: One of the "bibles" of ML, presented from a Bayesian perspective. My personal favorite.
  Cons: Requires much higher degree of mathematical sophistication. Poor at providing intuition (unless you gain intuition from math).
- **The Elements of Statistical Learning** (http://statweb.stanford.edu/~tibs/ElemStatLearn/), 2nd edition by Trevor Hastie, Robert Tibshirani and Jerome Friedman, Springer-Verlag, 2008. (Available online)
  Pros: One of the "bibles" of ML, presented from a statistical perspective. Free and online.
  Cons: Terse explanations. Least readable of the "bibles".
- **Information Theory, Inference, and Learning Algorithms** (http://www.inference.phy.cam.ac.uk/mackay/itila/) by David MacKay, Cambridge University Press, 2003. (Available online)
- **Machine Learning: A Probabilistic Perspective** (http://www.cs.ubc.ca/~murphyk/MLbook/) by Kevin P. Murphy, MIT Press, 2012.
- **Pattern Classification** (http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471056693.html) by Richard Duda, Peter Hart and David Stork, Wiley & Sons, 2001.
- **Bayesian Reasoning and Machine Learning** (http://www.cs.ucl.ac.uk/staff/d.barber/brml/) by David Barber, Cambridge University Press, 2012. (Available online)

Software

We will be using **Python** (https://www.python.org/) 2.7.x throughout the course to implement various ML algorithms and run experiments. You will require the following packages:

- **numpy** (http://www.numpy.org/) (**tutorial** (http://wiki.scipy.org/Tentative_NumPy_Tutorial))
- **scipy** (http://www.scipy.org/)
- **scikit-learn** (http://scikit-learn.org/) (**tutorial** (http://scikit-learn.org/stable/tutorial/index.html))

Other popular software:

- **Weka** (http://www.cs.waikato.ac.nz/ml/weka/downloading.html) -- We will not use Weka for this course, as it is too "black-box". If you are interested in exploring this software, be sure to use the "Stable Book 3rd Edition" version. Weka is built using Java, so you can download it into your home directory and run it directly there.

Math References

Machine learning requires a strong mathematical foundation. You may find the following resources useful to brush up on your math background:

- Probability
  - Murphy or Bishop Chapter 2 (see also Bishop Appendix B)
  - **Review notes** (http://cs229.stanford.edu/section/cs229-prob.pdf) from Stanford's machine learning class
  - Sam Roweis' **probability review** (http://cs.nyu.edu/~dsontag/courses/ml12/notes/probx.pdf)
- Linear algebra
  - Bishop Appendix C
  - **Online class from MIT** (http://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/)

    - **Review notes** (http://cs229.stanford.edu/section/cs229-linalg.pdf) from Stanford's machine learning class
    - Sam Roweis' **linear algebra review** (http://cs.nyu.edu/~dsontag/courses/ml12/notes/linear_algebra.pdf)
- Calculus
    - Bishop Appendix D and E (Lagrange multipliers)
    - **Notes** (http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-867-machine-learning-fall-2006/readings/lagrange.pdf) from MIT on Lagrange multipliers
    - Dan Klein's **Lagrange Multipliers without Permanent Scarring** (http://www.cs.berkeley.edu/~klein/papers/lagrange-multipliers.pdf)
- Optimization
    - **Convex Optimization** (http://www.stanford.edu/~boyd/cvxbook/) by Stephen Boyd and Lieven Vandenberghe (can be downloaded as PDF)

Grading

There are 500 "course points" distributed among the course components:

Problem Sets	250
Project	100
Mini-quizzes	0 (optional)
Midterm	50
Final	75
Participation	25
Total	**500**

That is, each "course point" is worth 0.2 points towards your overall course grade. Additionally, there will be opportunities for extra credit throughout the course.

Course Grades

To pass the course, you must submit a (course) project and take the midterm and final. Assuming these are true, we will use the following scale to convert course grades to letter grades:

A	[93,100]
A-	[90,93)
B+	[87,90)
B	[83,87)
B-	[80,83)
C+	[77,80)
C	[73,77)
C-	[70,73)
D range	[60,70)
F	[0, 60)

Questions about Grading

You may have questions or concerns about grading on occasion. When such issues arise, please send a private note to staff ("Instructors" on Piazza) indicating the specific problem and the nature of your question or concern. Please send any concerns regarding exam grading directly to Prof Wu.

Late Policy

Problem Sets

Problem sets that are submitted after the deadline will receive a 25% penalty for each day (24-hour period) that the assignment is late. Submissions received more than two days late will not be accepted.

Each student is allowed four "slip" days. It is up to you to track how many late days you have used and mark your count at the top of each late submission. (For example, if you turn in one assignment 3 hours late, you would write at the top "Turning in 3 hours late, using 1 of 4 late days.") Each day can only be used in its entirety (so, for example, you cannot use "half" a day to turn in an assignment twelve hours late), and you may use a maximum of two slip days per assignment. We count written and programming exercises as separate assignments. (For example, if you turn in both components one day late, you would have to use two slip days to submit without penalty.) If working in teams, slip days are deducted from each members' allotment. Late submissions (regardless of whether slip days are used) are not eligible for extra credit.

Projects

There are several project components: Proposal Conference (in-person), Proposal Presentation (in-class), Proposal Writeup (written), Status Update

(written), Final Presentation (in-class), Attendance / Participation (in-class), and Work Log (written). The written components are subject to the same rules and share the same slip day allotment as problem sets. In-person and in-class components will not be accepted late (meaning you cannot use slip days on them either).

Extensions for Illness

If you get sick or confront some other emergency and cannot turn in an assignment on time, we understand! To be fair and consistent, here is the policy: Go to Baxter, get a note from a doctor or Dean Bassman (or a dean at your home college if you are not a Mudder), and send it to Prof Wu as a private note. We will work together to find an appropriate arrangement.

Collaboration Policy

You are encouraged to discuss the material and work together to understand it. It is fine (and you are encouraged) to discuss the topics covered in the homeworks, to discuss approaches to problems, and to sketch out *general* solutions. If an assignment is meant to be completed individually, that means it should not be a collaborative effort or copied from a common source (e.g. whiteboard).

We will often allow you to work with a partner for programming assignments. If an assignment indicates that you can work in a pair, then you can choose your own work style (from full pair programming to each tackling a separate component). But regardless of how you divide the work, **every student is responsible for all work submitted by his or her group**.

For all assignments, you must indicate on each submission the names of people with whom you worked with or consulted (with the exception of course staff and course textbook or notes). The use of Internet resources (e.g., online Python tutorials) to aid in course work is acceptable as long as it does not substituted for an understanding of the course material. Plagiarism and direct use of external materials (e.g., books, online resources, or solution sets from previous offerings of this or other courses) to find hints or answers to any assigned work is strictly prohibited.

Exams, of course, must be your own individual work.

Assignments, exams, and any other course materials should not be committed to dorm repositories or otherwise redistributed for future or external students.

If you have any questions as to what types of collaborations are allowed and which are dishonest, please ask **before** you make a mistake.

Honor Code

Students enrolled in this course are expected to adhere to the Harvey Mudd College **Honor Code** (http://www.hmc.edu/studentlife1/ashmc1/constitution.html#VII) . As per the Student Handbook, students suspected of academic dishonesty will be reported to the Judiciary Board.

The Writing Center

The Writing Center provides a welcoming space for writers to get feedback on their composition projects, whether written, spoken or visual pieces. Writing Center Consultants are prepared to assist students in any discipline with any stage of the writing process, from developing an idea to polishing a final draft. Even the most accomplished writers benefit from seeking feedback at the writing center. The center is open Sunday through Thursday evenings from 7-11 and Saturday and Sunday afternoons from 3-5. It is located in Shanahan 1470, just up the walkway from the cafe. You may schedule an appointment through their **website, or you may simply drop in during normal hours.** (www.hmc.edu/writingcenter)

YOU WILL LIKELY FIND YOUR WRITING CENTER VISIT MORE VALUABLE IF YOU GO EARLIER THAN THE NIGHT BEFORE YOUR FINAL DRAFT IS DUE.

Accommodations

It is the policy of The Claremont Colleges to accommodate students with temporary or permanent disabilities. Any student with a documented disability who requires reasonable accommodations should contact Deborah Kahn, Coordinator for Student Disability Resources at (909) 607-3148 or **dkahn@hmc.edu** (mailto:dkahn@hmc.edu) , as soon as possible. Students from the other Claremont Colleges should contact their home college's disability officer.

- CMC: Julia Easley (**julia.easley@cmc.edu** (mailto:julia.easley@cmc.edu))
- CGU: Chris Bass (**chris.bass@cgu.edu** (mailto:chris.bass@cgu.edu))
- Pitzer: Gabriella Tempestoso (**gabriella_tempestoso@pitzer.edu** (mailto:gabriella_tempestoso@pitzer.edu))
- Pomona: Jan Collins-Eaglin (**jan.collins-eaglin@pomona.edu** (mailto:jan.collins-eaglin@pomona.edu)
- Scripps: Leslie Schnyder (**lschnyde@scrippscollege.edu** (mailto:lschnyde@scrippscollege.edu))