

CS165B

Introduction to Machine Learning

Winter 2020

Lecture:

MW 11:00am - 12:15pm - PSYCH 1902

Discussion:

M 4:00-4:50 pm - GIRV 1112

M 5:00-5:50 pm - GIRV 1112

M 6:00-6:50 pm - PHELP 1444

Final exam:

Thursday, March 19, 12:00pm - 3:00pm

Instructor:

Yuan-Fang Wang

Office Hours: MW 12:30pm - 1:30pm, HFH, Rm. 3113

email: yfwang@cs.ucsb.edu

TA:

Sikun Lin

Office Hours: (in CSIL) Monday 1:30-3:30 pm, Thursday 11am -1pm

email: sikun@ucsb.edu

Chinmay Garg

Office Hours: (in CSIL) Tuesday 2-4 pm, Friday 10 am-12 pm

email: chinmay@ucsb.edu

Required textbooks

Machine Learning: The Art of Science of Algorithms that Make Sense of Data, Peter Flachest
(reading assignments and practice problems will be drawn from here)

The following is a partial list of graduate-level books that serve as good references. Note that you can easily find PDF version or web version of these reference books. Hence, there is no need to shell out big bucks to buy hard copies.

- T. Hastie, R. Tibshirani, and J. H. Friedman, [The Elements of Statistical Learning, 2nd edition](#), Springer, 2009
- C. Bishop, [Pattern Recognition and Machine Learning](#), Springer, 2016
- K. Murphy, [Machine Learning: A Probabilistic Approach](#), MIT Press
- I. Goodfellow, T. Bengio and A. Courville, [Deep Learning](#), MIT Press
- D. Mackay, [Information Theory, Inference, and Learning Algorithms](#), Cambridge University Press, 2003
- R. Duda, P. Hart, D. Stork, [Pattern Classification, 2nd Ed.](#), John Wiley & Sons, 2001.
- Sutton and Barto, [An Introduction to Reinforcement Learning](#) MIT Press, 1998.

Labs:

- CSIL Lab, Harold Frank Hall 1132
- CSTL Lab, Phelp 3525

Grading:

- Assignments = 60%
- Midterm = 15%
- Final = 25%

Topics to be Covered:

1. Introduction to machine learning
2. Regression
3. Error analysis: bias vs. variance for regression , precision vs. recall for classification
4. Classification
5. Logistic regression
6. Introduction to deep learning
7. Back propagation
8. Support vector machine and kernel methods
9. Decision tree Learning
10. Bagging, boosting and ensemble analysis
11. Semi-supervised learning
12. Unsupervised learning and dimension reduction

As we will lose two Monday classes due to holiday, we might not be able to cover all these topics. Topics may be added and/or deleted depending on the interests and preparation of the participants.

Class objectives, policies and announcements:

- The course covers the core principals and key techniques of machine learning (ML), which is the study of algorithms that learn from data and experience. Topics including regression, classification, decision trees, Bayesian learning, and many others. Both theory and practice will be emphasized. However, this is an introductory course - you won't learn everything about machine learning here, but you'll be prepared to go deeper in the subject.
- Note that CS165A is a complementary course on Artificial Intelligence, although it is not a prerequisite to CS165B. Artificial Intelligence is a broader area and encompasses topics beyond machine learning such as logic, problem-solving and search, expert systems, intelligent agents, knowledge representation, planning, and natural language processing. You can take one or both, concurrently or sequentially.
- We will be using Piazza Forum for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TAs, reader, and instructor. Rather than emailing questions to the teaching staff, you should post your questions online. Important forum etiquette:
 - Be professional and courteous - this is not a place to vent your frustration!
 - Describe your problems clearly but succinctly - you are wasting your time composing long questions in the Twitter era,
 - Remember the rule of gives-and-takes: you want people to help you, you should try to help others if possible, however,
 - Don't post suggestions that you are not sure of or are intentionally misleading, and
 - Most importantly, never ever post your codes onto the discussion website (small snippets are ok).
- You can code and debug your assignments using your home PC. However, it is imperative you run your programs on machines in CSIL at least once before you turn in your source codes electronically for grading. This step is essential to make sure that no incompatibility (different header and library versions) exists that might prevent your programs from being run in CSIL.
- For all programming assignments, you must keep a copy of your codes in CSIL. You must not edit or change them in any way after you turn them in electronically. The purpose of this backup copy is to help resolving disputes regarding assignment grading (e.g., the turnin process corrupts the codes). You need to have a backup copy with a time stamp showing the codes' last modification date. A backup copy on your own computer is not acceptable.

- You should save important email correspondences with your readers/TAs/instructor (e.g., those grant you a late turnin without penalty, those change your grades, etc.) and retain all graded written assignments. This evidence is a must if you later want to contest grades as recording error.
- Homework due dates will be clearly marked on the handouts. In general, you will have about one to two weeks to finish a programming assignment. Plan your schedule wisely as late homework turnin, without a documented proof of an emergency, will be severely penalized. 20% will be taken off for any fraction of a day late, up to two days. No late assignment turnins will be accepted after the second day past the due day.
- There is no group assignment in this class. You can discuss problems with your classmates and consult your TAs and instructor if you have any questions, say, via Piazza. However, you must do all the assignments on your own; i.e., all answers and codes you turn in must be your own work. Any act of cheating, plagiarism or collaboration on assignments will be met with severe penalty.
- If you have questions concerning your program/exam grades, you should bring them to your readers and TAs first. If the questions cannot be resolved to your satisfaction, you should then consult the instructor. Any dispute over grades must be resolved in a timely manner (within two weeks after the grades are announced).
- Class attendance is highly recommended. You are responsible for everything that goes on in class. "I wasn't there" and "I didn't know" are not valid excuses for missing something important! Caveats: class slides are NOT for self study. They only provide major "talking points" with details filled in the lecture. For self study, you will need to have a book (any of the textbook and references should be good).
- Weekly one-hour discussion sections, led by the TAs, will focus on (1) presenting supplementary material (such as practical examples) to the lectures and (2) answering questions (about lecture topics, homework problems, programming issues, etc.). It is expected that you attend these sessions - you are responsible for what goes on during the discussion sessions also. For consistency, the discussion sessions in the same week will be lead by the same TA and discuss the same topics. You can go to any of these sessions no matter which one you register. Caveats: The assignments will comprise many topics and questions not covered in the lecture (especially assignments #1 and #2), you will be at a significant disadvantage if you miss the discussion sessions where the TAs explain the assignments.