CM146: Introduction to Machine Learning (Winter 2017)

Instructor: Kai-Wei Chang

Lecture: Tuesday/ Thursday

Discussions:

Course description

Machine Learning encompasses the study of algorithms that learn from data. It has been a key component in a number of problem domains including computer vision, natural language processing, computational biology and robotics. This class will introduce the fundamental concepts and algorithms in machine learning (supervised as well as unsupervised learning) as well as best practices in applying machine learning to practical problems. The class consists of lectures, problem sets that contain mathematical and programming exercises and two in-class exams.

Prerequisites

Undergraduate level training or coursework in algorithms, linear algebra, calculus and multivariate calculus, basic probability and statistics; an undergraduate level course in Artificial Intelligence may be helpful but is not required. A background in programming will also be necessary for the problem sets; specifically students are expected to be familiar with python and scikit-learn (a machine learning package for python) or learn it during the course.

Contact Info

Instructor: Kai-Wei Chang

Office Hours:

Email: kwchang at cs dot ucla dot edu

Teaching assistants

Textbooks

While there is not one textbook that covers all the material from this course, readings will come from the following texts:

- *A course in machine learning:* by Hal Daume III, which will be referred to as **CIML** (freely available online) is the primary reference. We will use version 0.9 of CIML.

- *Patter recognition and machine learning* by Christopher M. Bishop

For a more advanced treatment, the following are useful:
- *Machine Learning: A Probabilistic Perspective* by Kevin Murphy.
- *Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani and Jerome Friedman (freely available online)

Machine Learning requires a strong mathematical foundation. You may find the following resources useful to brush up your math background.
- Probability
  - Review notes from Stanford's machine learning course
- Linear algebra
  - Review notes from Stanford's machine learning course
- Optimization
  - Review notes from Stanford's machine learning course
  - Review notes from Stanford's machine learning course

Course format

- **Problem sets (40%)**: There will be periodic problem sets (aka homeworks). Questions on the problem sets will include math exercises, programming exercises and data analyses.
  - We will use gradescope to manage submission of problem sets.
  - Problem sets are due at 11:59 pm on the due date.
  - **Late submissions will not be accepted**
  - All solutions must be clearly written (or typed) ; unreadable answers will not be graded. We encourage using LaTeX to type out answers.
  - Solutions will be graded on both correctness and clarity. If you cannot solve a problem completely, you will get more partial credit by identifying the gaps in your argument than by attempting to cover them up.
- You are free to discuss the problems from the problem sets. However, you must write up your own solutions. You must also acknowledge all collaborators.
- **Mini quiz on math background (0%)**: This is a in class, closed-book and closed-notes mini quiz that will help you evaluate your background. This quiz does not count towards your final grade but is required for the course.
- **Exams (Mid-term: 25%, Final: 35%)**: There are two exams scheduled for ?? and ??. Exams are in class, closed-book and closed-notes and will cover material from the lectures and the problem sets. No alternate or make-up exams will be administered, except for disability/medical reasons documented and communicated to the instructor prior to the exam date. In particular, exam dates and times cannot be changed to accommodate scheduling conflicts with other classes.
- Problem set 0 and the math mini-quiz will be graded but not count towards your final grades. However, you need to submit both. We will not grade any other problem sets/exams unless you attempt problem set 0 and math mini quiz.

Software

We will extensively be using Python 2.7.x to implement ML algorithms and run experiments. You will require and need to familiarize yourself with the following packages:

- numpy: contains tools for numerical linear algebra, random number generation. For a numpy tutorial, see here .
- scipy
- scikit-learn : contains tools for machine learning and data science. For a tutorial, see here

Forums

Piazza

We will use Piazza for class discussions. Please go to this Piazza website to join the course forum (note: you must use a ucla.edu email account to join the forum). We strongly encourage students to post on this forum rather than emailing the course staff directly (this will be more efficient for both students and staff). Students should use Piazza to:

- Ask clarifying questions about the course material.
- Share useful resources with classmates (so long as they do not contain solutions).
- Look for project partners or other students to form study groups.
- Answer questions posted by other students to solidify your own understanding of the material.

The course Academic Integrity Policy must be followed on the message boards at all times. Do not post or request solutions to problem sets! Also, please be polite.

Gradescope

We will use gradescope to manage and grade problem sets and exams.

- Please see this guide

Policies

Academic Integrity Policy

Group studying and collaborating on problem sets are encouraged, as working together is a great way to understand new material. Students are free to discuss the homework problems with anyone under the following conditions:

- Students must write their own solutions and understand the solutions that they wrote down.
- Students must list the names of their collaborators (i.e., anyone with whom the assignment was discussed).

- Students may not use old solution sets from this class or any other class under any circumstances, unless the instructor grants special permission.

Students are encouraged to read the Dean of Students' guide to Academic Integrity.

Attendance and class participation

Although not a formal component of the course grade, attendance is essential for success in this course. If you are absent without a documented excuse, the instructor and TAs will not be able to go over missed lecture material with you. We emphatically welcome questions and your active participation in this course will enhance your learning experience and that of the other students.

Regrade requests

Regrade requests for homework and exams must be made through gradescope within one week after the graded homeworks have been released, regardless of your attendance on that day and regardless of any intervening holidays such as Memorial Day. We reserve the right to regrade all problems for a given regrade request.

Acknowledgments

The course website is based on material developed by Dan Roth, Sriram Sankararaman, Ameet Talwalkar and Fei Sha. Some of the administrative content on the course website is adapted from material from Jenn Wortman Vaughan, Rich Korf, and Alexander Sherstov.

Tentative Schedule (subject to change)

Date	Topics	Readings	Problem Sets
1/09	Introduction		Problem Set 0
1/11	Probability & Linear Algebra		
1/16	Math review mini quiz. Statistics		
1/18	Decision trees	CIML 1.3,1.5-1.10	

1/23	Nearest neighbors	CIML 2-2.3
1/25	Linear classification	CIML 3
1/30	Logistic regression	CIML 6.3
2/1	Linear regression	CIML 6-6.2, 6.4-6.6
2/6	Computational Learning Theory	
2/8	Kernels	CIML 9-9.2, 9.4-9.6
2/13	In-class mid-term	
2/15	Support Vector Machines	CIML 6.7
2/20	Ensemble methods	
2/22	Multclass classification	
2/27	Clustering	CIML 2.4, 13-13.1
3/1	Bayesian Learning	CIML 14-14.1
3/6	The Expectation Maximization algorithm	CIML 14.2
3/8	Hidden Markov Models (HMMs)	

3/13	HMMs continued
3/15	Neural networks