CM146: Introduction to Machine Learning (Fall 2018)

Instructor: Kai-Wei Chang

Lecture: Mon, Wed 4:00 pm – 5:50 pm

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. Please also see other prerequisites set by school in the course information.

Contact Info

Instructor: Kai-Wei Chang

Office hours: Mon 3:00pm-3:50pm @ Eng VI 374

Email: kw+cs146@kwchang.net

Teaching assistants

- Md Rizwan Parvez rizwan@ucla.edu
- Sajad Darabi sajad10@ucla.edu
- Sudharsan Krishnaswamy sudharsansai@cs.ucla.edu
- Yichao Zhou yz@cs.ucla.edu

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. Please your math background.

● Probability

● Linear algebra

● Optimization

Course format

● Problem sets (38%): 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** (please plan your submission ahead. Due to the large class, we will not accept any excuse emergency.)

  ○ 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 (2%): A take-home mini quiz will help you evaluate your background. This quiz does not count towards your final grade but is required for the course. You have unlimited attempts for this mini quiz.

● Exams (Mid-term: 25%, Final: 35%): There are two in class, closed-book and closed-notes exams. They 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.

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.

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 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
10/1	Introduction		
10/3	Overview (I)		
10/8	Overview (II)		
10/10	Decision trees		
10/15	Nearest neighbors		
10/17	Linear classification		
10/22	Logistic regression & Neural Network		
10/24	Regression		
10/29	Computational Learning Theory		
10/31	Kernels		
11/5	In-class Midterm exam		
11/7	Support Vector Machines		
11/12	Veterans Day holiday (no class)		
11/14	Ensemble methods		
11/19	Multclass classification		
11/21	Clustering		
11/26	Bayesian Learning		
11/28	The Expectation Maximization algorithm		
12/3	Structured Learning: Hidden Markov Models (HMMs)		
12/5	Structured Learning & Deep Neural Network		
??/?	Final Exam		