Spring 2024
Learning From Data

Instructor: Yang Li
TA: Yanru Wu, Boshi Tang and Jiahao Lai

Description

This introductory course gives an overview of many concepts, techniques, and algorithms in machine learning, beginning with topics such as logistic regression and SVM and ending up with more recent topics such as deep neural networks and reinforcement learning. The course will give the student the basic ideas and intuition behind modern machine learning methods as well as a bit more formal understanding of how, why, and when they work. The underlying theme in the course is statistical inference as it provides the foundation for most of the methods covered.

The course includes 2.5 hours of **lectures** (by the Instructor) each Friday from 9:50 am to 12:15 pm, and 40 mins of **optional recitation session** (by the TAs) before each class from 9:00 am to 9:40 am. Homework includes both written exercises and programming exercises. There will be a written midterm exam and a final course project includes a proposal, final presentation, and report.

Intended Students

The course is geared towards students who are interested in understanding machine learning algorithms, and to prepare them for research in this field. One of the objectives of the course is to understand the fundamental perspectives and develop solid connections between mathematical theory and learning systems.

Prerequisites

Basic concepts in calculus, probability theory, and linear algebra. Basic knowledge of Python programming.

Problem Sets

There will be a total of 5 written and 4 programming problem sets, due roughly every two weeks. The content of the problem sets will vary from theoretical questions to more applied problems. You are encouraged to collaborate with other students while solving the problems but you will have to turn in your own solutions. **Homework plagiarism will not be tolerated. If you collaborate with others in any part of the homework, you must acknowledge them in your submission. If you get homework help from an online resource (e.g. github), you must also give credit to the source.**

Final Project

The final project for the course will involve using applied techniques on learning related applications or theoretical explorations of machine learning. The instructor will provide a list of suggested datasets for students to chose from, but students are encouraged to find their own dataset or topic, subject to the approval of the instructor.

The final project will be done in groups of two. Each group will submit a written report and optionally present in class[1].

Grading

Your overall grade will be determined roughly as follows:

ACTIVITIES	PERCENTAGES
Midterm	15 %
Final Project	25 %
Problem sets (written & programming)	60 %

Each person will have **two free chances** to turn in a late homework assignment (except for the final project). Each late homework must be turned in within three days after its original deadline. Beyond the three-day grace period, late homework will not be graded. (Talk to the instructor if you have special circumstances.)

Textbook

The primary course materials are the lecture slides. You can also read the following reference materials:

- (Recommended) Machine Learning Lecture Notes by Andrew Ng: `https://github.com/mxc19912008/Andrew-Ng-Machine-Learning-Notes`
- Pattern Recognition and Machine Learning, 2nd Edition, by Christopher Bishop

Course Syllubus

Note: PA stands for "programming assignment"; WA stands for "written assignment".

[1] We may not have enough time for all groups to present during the final week. The final presentation method will be up to discussion later.

Date	Topic	Homework release
2/29	Review Session (optional)	WA0 (don't need to hand in)
3/1	Introduction	
3/8	Supervised Learning I • Linear regression • Logistic regression	WA1
3/15	Supervised Learning II • Generalized linear model	PA1
3/22	Supervised Learning III • Support vector machines	WA2
3/29	Supervised Learning IV • Generative model: GDA • Generative model: naive Bayesian model	PA2
4/5	Qing Ming	
4/12	Supervised Learning V • Deep neural networks	WA3
4/19	Midterm	
4/26	Unsupervised Learning I • K-means clustering • Principal component analysis	WA4, Final Project
5/3	May Day	
5/10	Unsupervised Learning II • Independent component analysis • Canonical component analysis	PA3
5/17	Unsupervised Learning III • Maximal HGR correlation • Spectral Clustering	WA5

5/24	Reinforcement Learning • MDP, value and policy iterations • Deep Q-learning	PA4
5/31	Practical Tips in Machine Learning • Bias and variance trade off • Model selection and feature selection	
6/7	Machine Learning Theory • Regularization • Empirical risk, VC dimension	
6/14	Advanced Topic • TBA	
6/21	Final Project Presentations	