Fall 2019
Learning From Data

Lecturer: Shao-Lun Huang & Yang Li
TA: Xinyi Tong & Zifeng Wang

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.

Intended Students

The course is geared towards students who are interested in understanding machine learning, and to carry out researches involving the applications of the machine learning problems. One of the objectives of the course is to understand the fundamental perspectives and develop solid connections between mathematical theory and learning systems.

Prerequisites

While this course does not assume any prerequisite, it will be helpful if the students can have some basic concepts in calculus, probability theory, and linear algebra. It is not required, but be prepared to work harder if you have not had it.

Problem Sets

There will be a total of 5 written and 4 programming problem sets, due roughly every 2-3 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. Copying will not be tolerated. If you collaborate, you must indicate all of your collaborators. Each problem set will be graded by TAs.

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 three. 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	20 %
Final Project	30 %
Problem sets (written & programming)	50 %

Course Syllubus

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

Week	Topic	Homework release
9/16-9/20	Introduction	
9/23-9/27	Supervised Learning I • Linear regression • Logistic regression	PA1
9/29 [2]	Supervised Learning II • Generalized linear model	WA1
9/30-10/4	Chinese National Day	
10/7-10/11	Supervised Learning III • Generative model: GDA • Generative model: naive Bayesian model	PA2
10/14-10/18	Supervised Learning IV • Support vector machines	WA2
10/21-10/25	Supervised Learning V • Deep neural networks	PA3

[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.

[2]Make-up class for Chinese National Holiday

10/28-11/01	Midterm	
11/4-11/8	Unsupervised Learning I • K-means clustering • Principal component analysis • Independent component analysis	WA3
11/11-11/15	Unsupervised Learning II • Canonical component analysis • Maximal HGR correlation	PA4
11/18-11/22	Unsupervised Learning III • Mixture Gaussian and EM algorithm • Auto-encoders	WA4, Final Project
11/25-11/29	Model selection I • Regularization • Empirical risk, VC dimension, model selection	
12/2-12/6	Model selection II • Hypothesis testing	
12/9-12/13	Reinforcement Learning • Markov decision process • Value iteration and policy iteration	WA5
12/16-12/20	Advanced Topic I • Transfer Learning	
12/23-12/27	Advanced Topic II • Semi-supervised learning	
12/30-1/3	Final Project Presentation I	
1/6-1/10	Final Project Presentation II	