Fall 2021
Learning From Data

Instructor: Yang Li
TA: Weida Wang & Guoqing Zhang

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

Basic concepts in calculus, probability theory, and linear algebra.

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 **three 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. Additional late homeworks will not be graded. (Talk to the instructor if you have special circumstances.)

Course Syllubus

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

Date	Topic	Homework release
9/17	Introduction	WA0 (don't need to hand in)
9/19	Review session (optional)	
9/24	Supervised Learning I • Linear regression • Logistic regression	PA1
10/1	Chinese National Day	
10/8	Supervised Learning II • Generalized linear model	WA1
10/15	Supervised Learning III • Generative model: GDA • Generative model: naive Bayesian model	PA2

[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.

10/22	Supervised Learning IV • Support vector machines	WA2
10/29	Supervised Learning V • Deep neural networks	PA3
11/5	Midterm	
11/12	Unsupervised Learning I • K-means clustering • Principal component analysis	WA3
11/19	Unsupervised Learning II • Independent component analysis • Canonical component analysis	PA4 , Final Project
11/26	Unsupervised Learning III • Maximal HGR correlation • Spectral clustering	WA4
12/3	Reinforcement Learning • MDP, value and policy iterations • Deep Q-learning	
12/10	Practical Tips in Machine Learning • Bias and variance trade off • Model selection and feature selection	WA5
12/17	Machine Learning Theory • Regularization • Empirical risk, VC dimension	
12/24	Advanced Topic • TBA	
1/7	Final Project Presentations	