CS 57800 – Statistical Machine Learning

Time: MW 5:30pm-6:45pm

Location: Peirce Hall 277

Instructor: Elias Bareinboim

Webpage: http://www.cs.purdue.edu/~eb/

Summary

This course is motivated by the need for providing different levels of "explanation" to the vast amount of data collected in different fields of human inquiry, including engineering, medicine, and the empirical sciences. Departing from structural causal models, we will study the concepts, principles, and algorithms necessary to solve modern, large-scale machine learning problems. We will study the tradeoffs involved in answering multiple types of queries, including associational, causal, and counterfactual. For instance, we will discuss under what conditions non-experimental data can be used to infer causal knowledge.

Prerequisites

In order to be successful in this course, you should have a basic knowledge of:

- Discrete Math (proof techniques, search algorithms, and graph theory)
- Calculus (find min/max of functions)
- Linear Algebra (vectors and matrices)
- Statistics (basic probability, modeling, experimental design)
- Some programming experience (with special understanding of complexity analysis)

People

Name	Email	Office Hours	Location
Elias Bareinboim	eb@purdue.edu	Wed 11:30-12:30pm	LWSN 2142L
Daniel Kumor (TA)	dkumor@purdue.edu	Fri 12-1pm	LWSN B116

If you have questions about lectures or material, please ask them on piazza, rather than emailing.

Material (tentative)

The following is a rough outline of the material we will cover. Depending on how fast things go, we might switch some things around during the semester.

Week	Subject	Material	Textbook
1	Introduction	Motivation, 3-layer scientific hierarchy, review probability.	
2	Probabilistic Graphical Models	Graphoids, i-maps, d-separation, Bayesian Networks, Naive Bayes, Markov Random Fields, Hidden Markov Models, Relation Models, variable elimination, belief propagation, sampling, parameter learning.	Probabilistic Graphical Models: Principles and Techniques. Daphne Koller and Nir Friedman ISBN: 8601401113034
3			
4			
5			
6	Causal Models	Causal Bayesian Networks, Structural Learning, do-operator, confounding bias, Simpson's paradox, Back-door criterion, do-calculus, sampling selection bias, transportability, structural causal models, counterfactual reasoning.	Causality: Models, Reasoning and Inference. Judea Pearl ISBN: 9780521895606
7			
8			
9	Midterm		
10	Supervised Learning	Decision Trees, Regression, SVM, VC dimension, Perceptron, Deep Learning, PAC learning.	Pattern Recognition and Machine Learning, Christopher Bishop. ISBN: 9780387310732 Reinforcement Learning: An Introduction, Richard S. Sutton. ISBN: 9780262193986
11			
12	Unsupervised Learning	Clustering, k-means, hierarchical, expectation maximization.	
13			
14	Reinforcement Learning	Bandits, Markov Decision Process, Value Function, Policy search.	
15			
16	Final Exam		

Grading

- Midterm exam: 30%
- Final exam: 40%
- Homeworks: 30%
    - About 5 homework assignments
- Attendance and participation: bonus

Please review the Purdue honor code. While working on assignments in small teams is okay, your homework solutions must be your own.

Course Policies

You are expected to attend lectures and participate in class. While taking notes on laptops and *small* snacks are allowed, please make sure you are quiet and respectful of those around you, including those behind you who might be distracted by your snacks/devices.

You are expected to come prepared to class, and to participate in class discussion, *especially* when something is not clear. If you are too shy to ask in class, please post on piazza or attend office hours.