L32 Pol Sci 583: Computational Social Science

Instructor: Christopher Lucas
Assistant in Instruction: Ryan Johnson

Fall, 2020

Lucas Office: Online	Johnson Office: Online
Lucas Office Hours: W 4-6pm	Johnson Office Hours: F 2-4pm

Course Description

We cover a variety of supervised and unsupervised learning methods, introducing students to algorithms, probabilistic models, and techniques for approximating solutions when there is no tractable closed-form.

In addition to problem sets, this course involves a substantial research component. Students will conduct research that implements or extends a model covered during the semester.

This course is designed for Ph.D. students. It is not recommended for undergraduate or masters students, except in exceptional circumstances with the permission of the instructor.

Materials

- *Bishop, Christopher M.*. Pattern Recognition and Machine Learning.
- *Murphy, Kevin*. Machine Learning: A Probabilistic Perspective.
- *Hastie, Trevor, Robert Tibshirani, and Jerome Friedman*. The Elements of Statistical Learning: Data Mining, Inference, and Prediction.

Grading

- Problem Sets (40% total; 4 psets, 10% each)
- Course Project (40% total; components listed below as a share of the grade on the course project)
  - Preliminary abstract (5%)
  - Data and descriptive results (5%)
  - Intermediate results (10%)
  - Presentation (20%)

  - First draft of paper (20%)
  - Final draft of paper (40%)
- Participation (20%)
  - Discussion (10%)
  - Paper Feedback (10%)

Course Policies

During Class

You are welcome to take notes electronically. However, refrain from using computers for anything but course-related activities. The use of cell phones is prohibited, as they are of no use in this course. You're welcome to eat and drink, but be respectul of your classmates and avoid disruptions.

Attendance Policy

Attendance is a prerequisite for satisfactorily completing this class.

Late Assignments and Missed Exams

No credit will be given for late assignments. Exceptions will only be made in extraordinary, well-documented circumstances. In the event that you miss a pset due to less-the-extraordinary circumstances, you will receive a 0.

Academic Integrity and Honesty

Problem sets: You are permitted to discuss problem sets with each other. However, *all code and words must be your own*.

Accommodations for Disabilities

Student with disabilities are encouraged to make a private appointment with me during the first two weeks of the course so appropriate accomodations can be made.

Religious Observances

Students with religious observances that conflict with participation are encouraged to make a private appointment with me during the first two weeks of the course so appropriate accomodations can be made.

Course Outline

Our priority is to ensure that all students are learning, and so our pace through the material is subject to revision.

Key Dates

- October 7: Preliminary abstract
- October 28: Data and descriptive results
- November 11: Intermediate results
- November 25: Presentation
- December 2: First draft of paper
- December 9: Peer comments
- December 16: Final paper

Core Units

The following units represent the core of the class. The section following this one overviews elective topics, covered if time permits and according to student research interests.

Introduction

- What can we do with statistical learning?
- Where does statistical learning fit in your research?
- Course logistics

Supervised Learning

- What is supervised learning?
- Logistic regression
- SVM
- Linear discriminant analysis
- K-nearest neighbors

Dimension Reduction

- Principal component analysis
- t-distributed stochastic neighbor embedding

Mixture Models

- Clustering introduced
- Finite mixture models
- The expectation-maximization algorithm

Optional Units

If we complete the units in the previous section at a sufficient pace, we will cover a subset of the following units.

Topic Models

- Latent dirichlet allocation
- Correlated topic model
- Structural topic model
- STM applications

Neural Networks

- Convolutional neural networks
- Recurrent neural networks

Hidden Markov Models

- Hidden markov models introduced
- The model of audio and speech structure

Variational Inference

- Variational inference
- Variational Bayes

Bayesian Bootstrap

- Bayesian bootstap in general
- Application to the model of audio and speech structure

Gaussian Processes

- Guest lecture