Term : 2

Year : 2024

PHIL7010: Formal Methods for AI, Ethics and Society

Professor: Boris Babic
Email: babic@hku.hk
Office: Run Run Shaw Tower, 10.03
Office hours: Mon 3-4pm
TAs:
Yiwen Li, evenli@hku.hk
Office hours (by Zoom): Wed 1-2pm
Iris Yeung, irishmy@hku.hk
Office hours (by Zoom): Wed 2-3pm

Time: Tuesday 6:30-8:20pm
Location: LE7

Course Description

This course will be a general introduction to the fundamental data science tools and techniques used in modern machine learning and artificial intelligence systems. The course is a continuation of PHIL 7001. We will begin with supervised classification, particularly neural networks (starting from the single layer perceptron), and move on additional learning models, including clustering, nearest neighbours analysis, principal components analysis, and natural language processing.

The goal will be to develop each skill from the ground up, focusing on (a) conceptual understanding and (b) students' construction of these models in R.

There will be several case studies throughout the class. The case studies present issues that commonly arise when machine learning and AI are introduced into corporate finance, criminal justice, and health care. During the case studies, you will work in small groups to solve several problems, which we will later debrief in class.

Materials

I will post all required readings to the course webpage. No textbook is required but we will draw from the sources below. (These textbooks are freely available online). ISL is introductory, and ESL is advanced.

ISL: Witten et al, An Introduction to Statistical Learning with Applications in R.

ESL: Hastie et al, Elements of Statistical Learning.

Assignments

Grades will be based on the following:

Midterm: 30% Final: 30% Homework (15% each for 2 assignments): 30% Participation and case studies: 10%

The midterm and final exams will be in-class. You will have two hours to complete each test. The test will be closed book, but you can bring one "cheat sheet" (instructions will be provided in class). You can use a basic (non-graphing) calculator, but no smart devices of any kind are permitted.

Participation and case study grades will be calculated on the basis of class attendance, group participation, case study completion, and in-class contributions. This is a large class, and we will keep in mind that in class contribution may be limited. However, there will still be opportunity to engage.

For example: I may call on students to answer a question, or address a reading, and when called on, students will be expected to provide a meaningful/insightful answer. This does not mean I will reward those who "talk the most" or punish those who "talk the least". You are simply required to attend, be attentive, be present, and contribute meaningfully.

Please come to class on time, we will start promptly and avoid interruptions.

Homework assignments are designed to test your understanding of the material as preparation for the exam. There will be two homework assignments.

Prerequisites

The prerequisite for this course is PHIL 7001.

Part I: Methods

Week 1: Introduction and Neural Networks

Optional reading

ISL Chapter 10, 10.1(pgs. 403-406), 10.2(pgs. 407-411), 10.6(pgs. 432-433), 10.7(pgs. 434-439).

Week 2: Neural Networks with Applications in R

Optional reading

ISL Chapter 10, 10.9(pgs.443-448).

Week 3: Introduction to Unsupervised Learning and Nearest Neighbours Analysis

Optional reading

Chapter 2, 2.2.3(pgs.39-42), Chapter 3, 3.5(pgs.105-110), Chapter 4, 4.7(pgs.181-185), Chapter 12, 12.1(pgs.495-496).

Week 4: Clustering

Optional reading

Chapter 12, 12.4(pgs.514-519, pgs.528-530), 12.5(pgs.536-538).

Week 5: Understanding Latency and Principal Components Analysis

Optional reading

Chapter 12, 12.2(pgs.496-508), 12.5(pgs.530-533).

Week 6: Natural Language Processing

Week 7: Class suspension for Reading week

Week 8: Midterm (March 12)

Part II: Applications

Week 9: Case Study A

Week 10: Case Study B

Week 11: Case Study C

Week 12: Case Study D

Week 13: Final Exam (April 16)

Note: There is no class on Tuesday Feb 13 (CNY) and on Tuesday March 5 (Reading week)

Submitting Assignments and Late Policy

Submittable work must be submitted online by 11:59pm on the day it is due, unless otherwise instructed.

If you anticipate needing more time on an assignment, you should ask me in advance. Otherwise, late assignments will be penalized by one-fourth of a letter grade for each day they are late.

Students with Disabilities

If you think you may need accommodation for a disability, please contact me or campus accessibility services as soon as possible.

Plagiarism

Written work submitted for a grade in this course must be your own. You are responsible for making sure that none of your work is plagiarized. Cite the sources you rely on, and err on the side of caution where necessary.