CS7470: Topics in Programming Languages
Foundations of Probabilistic Programming

Instructor: Steven Holtzen s.holtzen@northeastern.edu

Northeastern University

Fall 2023

https://www.khoury.northeastern.edu/home/sholtzen/CS7470Fall23/

Overview

1. Course logistics and introduction

2. What are probabilistic programming languages and why are they worth studying?

3. Overview of course content and format

What is this course?

- A graduate-level seminar on probabilistic programming languages (PPLs)

- Goals for this class (what you will know by the end)
  - In short: **bringing you up to speed on the modern research landscape of probabilistic programming**
  - What PPLs are and what they are useful for
  - How to design and implement your own PPL
  - How to read probabilistic programming papers
  - How to design and execute a research project using PPLs
  - …

- An experimental new course! We will learn as we go

Who is the audience?

- PhD. and advanced Master's students from programming languages, AI, or machine learning
  - Even if this is not you, you are still welcome! It's healthy to not be the "target audience" sometimes :-)

- Expected background:
  - Ability to program in OCaml (you can learn as part of the class, but it will be necessary; see exercise on course webpage)
  - Familiarity (or willingness to learn) some mathematical foundations from logic, computational complexity, and probability, programming languages

- Talk to Steven if you have any concerns/questions

Course Resources

- **Course webpage** (main source for content, "ground truth"):
  https://www.khoury.northeastern.edu/home/sholtzen/CS7470Fall23/

- **Course notes and reading**
  - See today's course notes!

- **Slack**: For course discussion, join using the link on the webpage
  - All course communication will happen on Slack

- **Office Hours**: by appointment

- **Email me** (please include "7470" in the subject)
  s.holtzen@northeastern.edu

Personal Resources

- I aim for the class to be an *inclusive environment* where **anyone** feels empowered to share their point of view
  - Do not be afraid of speaking up or volunteering

- Having trouble in the class? Feeling stressed?
  - Email me or schedule some office hours: we can work something out and I can help you find resources

Why probabilistic programming?

See the note at https://github.com/neuppl/CS7470-Notes/blob/main/lecture-1/lecture-1.pdf

Course Overview

Graded Material: Minor Projects

- Two minor projects, both due Oct 20
  - Implementing a discrete PPL
  - Implementing a continuous PPL

- Project info is available on course webpage

- Written in OCaml
  - https://course.ccs.neu.edu/cs4410sp22/hw_warmup1_assignment.html

All deadlines listed on Canvas

Graded Material: Term Project

- Self-directed research project due at the end of the term

- You choose the topic and define your own metrics of success

- Syllabus is on the website, please read

- Goal: deeply engage with the material, try something new, implement something from a paper, extend one of the two minor projects in an interesting way, …

Graded Material: Term Project

1. Proposal (due **Oct. 20**)
   - Tell me about what you want to do

2. Check-in (due **Nov. 16**)
   - Tell me how things are going
   - Are you on track to finish? Have you had to re-plan?

3. Final presentation in class **Dec. 4 and Dec. 8**
   - 20-30 minutes long. Summarize your project, key results, future work, etc.

4. Final report (due Dec. 8)

Course structure

CAUTION WORK IN PROGRESS

- Module I: Foundations Lectures
    "Everything you need for the first project"
    - Propositional logic
    - Discrete Probability
    - Weighted model counting & knowledge compilation
    - A discrete probabilistic programming language

"Reading" means read in advance. I will assume you have read it in advance

"Supplemental reading" means read if you are interested

Course structure

- Module 2: Continuous PPL Lectures

  "Everything you need for the second project"

  - Basics of continuous probability: probability densities, a tiny amount of measure theory

  - Continuous inference: importance sampling, Markov-chain Monte Carlo

Course structure

- Module 3: Topics
  - Differentiable programming
  - Static analysis and deductive verification
  - Potpourri (collection of papers
  - These are subject to change and discussion; if you want to read something, let me know!

- This module involves reading papers
  - I have listed some, but they are subject to change
  - Everyone will be expected to **pick one paper** to write a blog post about. You are required to create your blog post 24 hours before the session. See this for examples: https://www.cs.cornell.edu/courses/cs6120/2020fa/blog/
  - If you want to propose a paper to present that's not on the list, email Steven

Learning Resources

- See the "resources tab" on the webpage: specific textbook/tutorial

- There will be things you are not familiar with that I cover too quickly; it's a graduate-level course
  - some self-learning is expected; useful skill to learn

- Share things in Slack! Discuss with your peers, post questions as you're reading