Up to date syllabus: ORIE6217/CS6384 Syllabus -- Spring 2023 - Google Docs

CS6384/ORIE6217: Applied Bayesian Data Analysis for Research

Instructor: Prof. Nikhil Garg (he/him), ngarg@cornell.edu

Credits: 3 hours

Lecture Timings: M/W 11:25 - 12:40

Course communication

Course communication will primarily be over Slack and email. Important announcements will be sent over email, but material may be shared over Slack. Except for personal requests/questions, I expect all student questions to be over Slack.

Please join the #bayesian_data_analysis_sp2023 channel in the Slack workspace here: https://join.slack.com/t/gargngroup/shared_invite/zt-1o8qsgkkk-_nsYakyT37SVVTH_Q~RiZA

Course credit note: The course has been approved to count for the IS Computational Methods Core PhD requirement, and is cross-listed as an ORIE/CS PhD course.

Course website: https://cs6384-orie6217.github.io/Spring2023/

Zoom link:
https://cornell.zoom.us/j/97962933007?pwd=V3paOHVqelRpdVdrMmRGbXEvQnNLdz09

Course Description

Bayesian modeling and data analysis is a powerful tool for computational research. It consists of writing a probability model and then fitting it with observed data, while handling uncertainty. The model can be flexible, encompassing hierarchy, spatio-temporal dynamics, graphs, and high-dimensionality.

This course is a graduate, hands-on introduction to Bayesian analysis in Stan and/or Pyro. The focus will be on writing and fitting models in practice for computational research, including the applied Bayesian statistics workflow: model building, checking, and evaluation. The course will also discuss research papers that use such methods.

After the course, students should be able to:

- Start with a research question and estimand of interest, and (1) construct a data generating process for the setting, and (2) construct a Bayesian model reflecting that process.
- Determine whether (and in what cases) the estimand of interest is identifiable.
- Write down the model in a Bayesian programming language such as Stan and/or Pyro.
- Fit and evaluate model fit using data.

The class will be a mixture of lectures by the professor, paper discussions, student presentations (of their own work and that of others), guest lectures, and hands-on coding sessions (Stan and/or Pyro).

Please follow these norms in all communications with the instructor and other students.

Course topics + rough schedule

1. Introduction and overview of Bayesian inference (~1 weeks)
2. Developing and debugging basic models in Stan/Pyro (~2 weeks)
    a. Single parameter models, multivariate models, Bayesian regression, Multinomial choice models, Hierarchical models
    b. Basics of sampling algorithms
3. Bayesian workflow (~1-2 weeks)
    a. Model checking (posterior predictive checks, diagnosing identifiability issues, etc)
    b. Evaluating, comparing models
4. Research paper case studies (~4-5 weeks)
    a. Each lecture we will discuss 1 paper. There will be student presenters for the paper (via slides), and then you will replicate in code the model + fitting from the paper.
    b. Some guest lectures
5. Miscellaneous topics (~1-2 weeks)
    a. Advanced models – gaussian processes/dynamic models, HMMs and Spatial models, conditional random fields, nonlinear and nonparametric models
    b. Sampling algorithms (Gibbs, Metropolis Hastings, HMC, Variational Inference)
    c. Computation concerns (speeding up models, GPU usage, etc)
6. Student project presentations (1-2 weeks)

Assignments + Grading

1. Homework assignments (1 assignment) [10%]
    a. Developing, fitting, and debugging models in Stan/Pyro
    b. Peer grading: Give feedback to each other (in pairs/triplets) on the homework.
2. Research paper presentation + discussion lead (in teams of 4) [30%]
    a. Present the paper via slides + submit a report
        i. Discuss model in detail, how it connects to the research question
    b. Replicate the paper in Stan or Pyro
        i. Simulate data according to model proposed in the paper
        ii. Get their code working on sample data
        iii. Walk through the code in class

    c. Presentation discussion + peer review:
        i. Read the paper another group has chosen to present.
        ii. Send me paper summary + discussion questions before class
        iii. Give feedback to the presenting team on presentation quality/clarity
3. Course project (in teams of 2) [40%]
    a. Define a research question of interest, formulate a model for a data generating process, write the model in Stan/Pyro, fit it using either synthetic or real data, and write a project report.
4. Participation + Attendance [20%]
    a. I will take attendance on random days, and this will make up part of the "attendance/participation" grade. Class participation in discussion will also contribute to this grade.

Prerequisites/Corequisites and Preparation summary

This is a PhD course. There are no explicit requirements, but mathematical/statistical/coding maturity is recommended, especially prior probability background (undergraduate level background in probability/statistics/machine learning); a desire to learn the course material is necessary. First year PhD students in ORIE/CS/IS are likely to have the technical background necessary for the course.

Textbook(s) and/or Other Required Materials

None required to purchase. All readings will be distributed throughout the semester or are readily found online. We will use the following textbooks:

1. Bayesian Data Analysis
2. A First Course in Bayesian Statistical Methods
3. Statistical Rethinking

These textbooks are accompanied by online video lectures. We may require students to watch the video lectures and then spend our class-time on coding models and discussion.

Covid and attendance policy

Attendance and participation is expected at every lecture. Lectures will be hybrid, as the course is split across two campuses; if you're attending virtually, I expect "video on" for the vast majority of the class. Class participation (attendance, active discussion of papers, presenting papers) is an essential part of the course. If you cannot attend a lecture due to extenuating circumstances (e.g., illness, religious observances, etc.) please email me *ahead of time* so we can discuss make-up work to replace in-class discussion. This will usually involve watching the lecture video and writing a 500 word reflection on the contents.

About the instructor

Nikhil is an assistant professor of Operations Research and Information Engineering at Cornell Tech, whose research is at the intersection of computer science, economics, and operations -- on the application of algorithms, data science, and mechanism design to the study of democracy, markets, and societal systems at large. Things he's worked on include crowdsourcing systems, surge pricing, rating systems, how to vote on budgets, gerrymandering, stereotypes in word embeddings, and political polarization on Twitter. Outside of academia, Nikhil has been a data scientist at Uber, collaborated with Upwork and other freelancing marketplaces, and most recently led campaign data science at PredictWise during the 2020 US election cycle. He has used Bayesian modeling techniques in both research and industry.

Academic Integrity

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity. Any work submitted by a student in this course for academic credit will be the student's own work. The policy can be found on the university's website here: https://theuniversityfaculty.cornell.edu/academic-integrity/. Students are encouraged to discuss papers and course content with each other.

Students with Disabilities

Your access to this course is important. Please give me your Student Disability Services (SDS) accommodation letter early in the semester so that we have adequate time to arrange your approved academic accommodations. If you need an immediate accommodation for equal access, please speak with me after class or send an email message to me and/or SDS at sds_cu@cornell.edu. If the need arises for additional accommodations during the semester, please contact SDS. You may also feel free to speak with Student Services at Cornell Tech who will connect you with the university SDS office.

Religious Observances

Cornell University is committed to supporting students who wish to practice their religious beliefs. Students are advised to discuss religious absences with their instructors well in advance of the religious holiday so that arrangements for making up work can be resolved before the absence.

Cornell Tech Cares

The Cornell Tech community is a diverse and vibrant group of students, faculty, and staff. We take our responsibility to look out for one another seriously. As members of this community, your openness and proactive communication will allow us all to better care for students and respond to their needs, whether they be interpersonal or academic. Please help us continue to build and strengthen our community by reaching out if you are having an issue or are concerned about a fellow student. Contact studentwellness@tech.cornell.edu with concerns and we will

make sure to care for one another. In the event of an emergency, please call 911 and Cornell Tech Safety & Security at 646-971-3611 (This number is also located on the back of your Cornell ID), when safe to do so.