STATS271/371: Applied Bayesian Statistics
Instructor: Scott Linderman
Term: Spring 2021
Stanford University

Course Description:
This course is a modern treatment of applied Bayesian statistics with an emphasis on building probabilistic models, algorithms for approximate Bayesian inference, and methods for checking, criticizing, and revising models. Some of the models we will study include classic Bayesian mixture and regression models, hierarchical models, factor models, topic models, Gaussian processes, Dirichlet process mixture models, and deep generative models. Alongside these models we will study algorithms for approximate Bayesian inference including Markov Chain Monte Carlo and variational inference algorithms. Finally, we will discuss methods for checking, criticizing, and revising models in an iterative manner, completing a virtuous cycle of applied Bayesian statistics.

Prerequisites:
Students should be comfortable with probability and statistics as well as multivariate calculus and linear algebra. Specifically, we assume you are familiar with the material in Chapters 1-3 of Bayesian Data Analysis, 3rd ed. (link below); e.g., through STATS270/370. This course will emphasize implementing models and algorithms, so coding proficiency (in Python, R, or Julia) is required.

Logistics:
Time: Monday and Wednesday, 1:00-2:20pm.
Level: advanced undergrad and up
Grading basis: credit or letter grade
Office hours: Yu: Tues 5:30-7:30pm; Scott: Wed 2:30-3:30pm; Ishmael Thurs: 9-11am.
Final evaluation: Project

Resources:
Our primary references will be Bayesian Data Analysis, 3rd ed. by Gelman et al and selected journal articles. The following references may be useful as well:

- A First Course in Bayesian Statistical Methods. P. Hoff
- Bayesian Reasoning and Machine Learning. D. Barber
- Probabilistic Machine Learning. K. Murphy.

Tentative Schedule:
The course is organized around “Box’s Loop,” a conceptual framework for approaching applied data analysis. It’s an iterative process of modeling, inference, criticism, and refinement. We’ll take multiple “laps” around this loop throughout the quarter, each time introducing new probabilistic models, inference algorithms, and model checking procedures. This structure entails a bit of hopscotch through BDA3, but it’s worth it!

Mon, Mar 29	Introduction: Box’s Loop	"Build, compute, critique, repeat: Data analysis with latent variable models" (Blei, 2014)
Wed, Mar 31	Lap 1: Bayesian Linear Regression, Exact Posterior Inference, and Model Selection	BDA3 Ch 14
Mon, Apr 5		BDA3 Ch 7
Wed, Apr 7	Lap 2: Generalized Linear Models and Laplace Approximation, Posterior Predictive Checks	BDA3 Ch 16
Mon, Apr 12		BDA3 Ch 6
Wed, Apr 14	Lap 3: Robust models, Hamiltonian Monte Carlo, MCMC Diagnostics	BDA3 Ch 17
Mon, Apr 19		“MCMC using Hamiltonian dynamics” (Neal, 2012), BDA3 Ch 11.4-11.5
Wed, Apr 21	Lap 4: Bayesian Mixture Models and (Collapsed) Gibbs Sampling	BDA3 Ch 22
Mon, Apr 26		BDA3 Ch 23
Wed, Apr 28	Lap 5: Topic Models, Matrix Factorization, and Coordinate Ascent Variational Inference	"Probabilistic topic models" (Blei, 2012)
Mon, May 3		“Variational Inference: A Review for Statisticians” (Blei et al, 2017)
Wed, May 5	Lap 6: Deep Generative Models, Black Box Variational Inference	"Stochastic backpropagation and approximate inference in deep generative models" (Rezende et al., 2014)
Mon, May 10		“An Introduction to Variational Autoencoders” (Kingma and Welling, 2019)
Wed, May 12	Lap 7: State Space Models, Message Passing, and Sequential Monte Carlo	BRML Ch 23
Mon, May 17		“Elements of Sequential Monte Carlo” (Naesseth et al, 2019) Ch. 1 and 2.

Wed, May 19	Lap 8: Gaussian Processes, Elliptical Slice Sampling, Bayesian Optimization	BDA3 Ch 21, “Elliptical slice sampling” (Murray et al, 2010)
Mon, May 24		“Taking the Human out of the Loop: A Review of Bayesian Optimization.” (Shahriari et al, 2016)
Wed, Apr 28	Victory Lap	The great beyond
Mon, May 31	*Memorial Day. No class.*	
Wed, June 2	Project poster session	

Assignments:
- 7 weekly homework assignments (implementation focused)
- 1 final project proposal
- 1 final project report (with code)

Note: the homework assignments will involve coding up the model and algorithm and applying it to a given dataset. You can code in Julia, Python, or R (i.e. as long as it runs in a Jupyter notebook), but the TAs and I can't offer “tech support.” We're happy to help with conceptual questions and will help think through implementations, but can't necessarily go through your code line by line.

Grading:
- We will drop the lowest homework grade.
- Top 6 assignments: 10% each = 60%
- Project proposals: 5%
- Final project: 35%

*Note to C/NC students: passing is >65% total, so you **must** do a project to pass.*

Policies:
This is an extraordinary year and we're all struggling to do our best despite the challenges of COVID19. Though we're studying remotely, I do expect students to engage in lecture with their camera on, if possible. Please be respectful of other students, especially since we're coming from so many different backgrounds. Different people know different things, but everyone has something to contribute. I'll be understanding of the extenuating circumstances and allow each student 7 late days total to be used over the quarter, and I'll drop the lowest assignment.

Honor Code:

See https://communitystandards.stanford.edu/policies-and-guidance/honor-code