WILLIAM & MARY
CHARTERED 1693

Bayesian Reasoning in Data Science (Spring 2024, DATA 644)

Schedule

Tuesday and Thursday 3:30PM to 4:50PM ISC 0280

Instructor

Cristiano Fanelli
cfanelli@wm.edu
ISC 1265
Office hours: Friday 9:00 to 11:00am or by appointment

Course Description

No data scientist can work without a solid grasp of conditional probability and Bayesian reasoning. Bayes'theorem allows to update our beliefs based on the occurrence of new events, steering the inference towards the truth and assessing uncertainty in predictions. This course offers an in-depth examination of Bayesian reasoning in data science, allowing you to grasp the fundamental components of this approach through practical examples spanning various domains. Throughout the course, you will acquire hands-on experience with concrete computational implementations, enabling students to bridge the gap between theoretical concepts and their programming applications, thereby solidifying their understanding of the material. The DATA 644 is offered to graduate students, and compared to DATA 340 includes topics such as deep learning and networks of beliefs.

Prerequisite(s):

Basic knowledge of Python and statistics.

Credit Hours:

3

Grade Distribution

- **Assignments** (5 Total): 60% — Comprising problem sets, programming tasks, and data analysis projects.
- **Final Project**: 40% — A significant research project or application of Bayesian methods. Involves a written report and potentially a presentation or poster session. Each assignment contributes 12% to the final grade.

Course Grading Policy

Letter Grade Distribution			
≥ 93.00	A	$73.00 - 76.99$	C
$90.00 - 92.99$	A-	$70.00 - 72.99$	C-
$87.00 - 89.99$	B+	$67.00 - 69.99$	D+
$83.00 - 86.99$	B	$63.00 - 66.99$	D
$80.00 - 82.99$	B-	$60.00 - 62.99$	D-
$77.00 - 79.99$	C+	≤ 59.99	F

Attendance

This class does not have an attendance policy. However, it will be difficult to learn enough to pass the class without regular participation, as the majority of course content relevant to tests and assignments will be covered in class.

Classroom Behavior

Please remain civil during discussions to promote the open exchange of ideas and foster a culture of open dialogue. Please bear in mind that all students are en- titled to their own opinion.

Late / Poor Performance Policy

Assignments will not be accepted late, excepting in documented circumstances

Collaboration

All five assignments should be done individually. For the final project, you may work in groups of up to three. The expectations for the project scope will increase depending on the number of group members. Each group will submit a report, and for groups of more than one, a short paragraph should be included along with the report that explains the role of each group member.

Communication

Our preferred method of communication will be Blackboard discussion forum. Please post your questions about the course material here. Although you may post privately to your instructor, you are strongly encouraged to post publicly as it is often the case that if one person has a question, many others share the same question.

Final Project

The final project will include a report and potentially a poster presentation based on a data analysis you conduct within your own domain(s) of study.

Resource/Software

Python / PyTorch

Academic Integrity

Academic dishonesty is taken very seriously. Make sure to cite all of your work, and do not turn in work that is not yours! Cases of academic dishonesty will be evaluated and acted upon in accordance with William and Mary policies, which can be found at http://www.wm.edu/offices/deanofstudents/services (student-conduct/)

Use of Artificial Intelligence

The following three statements (in no particular order) are part of the AI Writing Tools Working Group convened by the President and the Dean of the Faculty last year.

- The use or incorporation of any AI-generated content (from ChatGPT, Dall-e, etc.) in assignments is not allowed.
- All work submitted in this course, whether in draft or final form, must be your own and must be cited appropriately. You may incorporate AI-generated content or ideas in assignments, but you must indicate your use of such tools by:
  - including a formal citation to ChatGPT, or
  - including an acknowledgement, or
  - providing the specific prompt(s) used to generate content, or
  - providing a full log of prompts, outputs, and revisions.

  You remain responsible for the validity of all material generated, noting that AI-generated content often contains falsehoods and fictional sources. I strongly recommend that you carefully check all AI-generated content before submitting it for review.
- In this course, we could explore the use of AI-generated content as an educational tool. You will analyze the quality/ethics/bias/etc. of this content. Ideas and content generated by you, and those that are AI-generated, should be clearly delineated.

Schedule

Week	Class Topics	Other
Week 1	• Tue, Jan 30: Bayes Theorem, priors, likelihood, posterior, distributions; subjective and objective interpretations of probability • Thu, Feb 01: Thinking Probabilistically: Causes and Effects / True and Measured; Credible Intervals, ROPE, HDI	Fri, Feb 02: Add/drop deadline
Week 2	• Tue, Feb 06: Inference, Uncertainty and Probability • Thu, Feb 08: Intro to probabilistic programming (with very light intro to samplers)	ASSIGNMENT 1 (assigned Feb 10; due Feb 16)
Week 3	• Tue, Feb 13: Gaussian Inferences • Thu, Feb 15: Bayesian Linear Regression	
Week 4	• Tue, Feb 20: Bayesian Polynomial Regression • Thu, Feb 22: Bayesian Logistic Regression	ASSIGNMENT 2 (assigned Feb 24; due March 3)
Week 5	• Tue, Feb 27: Bayesian Softmax Regression • Thu, Feb 29: Uncertainties: aleatoric and epistemic	Mon, Mar 04: Advising + Midterm grading period starts
Week 6	• Tue, Mar 05: Models Comparison • Thu, Mar 07: Information Criteria, Bayesian Factors	ASSIGNMENT 3 (assigned Mar 08; due Mar 21) Sat, Mar 09: Spring Break starts
Spring Break	• Tue, Mar 12: No class • Thu, Mar 14: No class	Sun, Mar 17: Spring Break ends
Week 7	• Tue, Mar 19: Back to sampling: Markov Chain Monte Carlo; MCMC diagnostics and applications • Thu, Mar 21: Bayesian A/B testing	Sun, Mar 24: Midterm grading period ends
Week 8	• Tue, Mar 26: Gaussian Processes • Thu, Mar 28: Gaussian Processes	Mon, Mar 25: Withdrawal deadline ASSIGNMENT 4 (assigned Mar 29; due Apr 15)

		Fri, Mar 29: Advising period ends
Week 9	• Tue, Apr 02: Bayesian Optimization • Thu, Apr 04: Bayesian Optimization	
Week 10	• Tue, Apr 09: Bayesian Networks • Thu, Apr 11: Bayesian Networks	ASSIGNMENT 5 (assigned Apr 16; due Apr 27)
Week 11	• Tue, Apr 16: HPC at W&M • Thu, Apr 18: Deep Learning in a Bayesian Framework: what is Bayesian in all this?	
Week 12	• Tue, Apr 23: Deep Learning in a Bayesian Framework • Thu, Apr 25: Deep Learning in a Bayesian Framework	
Week 13	• Tue, Apr 30: Review of topics • Thu, May 02: Review of topics	Fri, May 03: Last day of classes: Review of topics
Finals	May 06-10, 13-14: Final Exams start FINAL Final grades due for graduating students: May 16, 9am Final grades due for continuing students: May 21, 9am	