Math 607: Applied Math II: Statistical Learning

University of Oregon, Winter 2024

Time: TΘ, 12-1:20pm
Place: Columbia Hall 44
Instructor: James Murray
Instructor Email: `jmurray9@uoregon.edu`
Teaching assistant: Halley Fritze (`hfritze@uoregon.edu`)
Office hours: Tue, 1:30-3pm
Office hours location: Knight Campus lobby

Prerequisites: Prerequisites for this course include multi-variable calculus and linear algebra. Some prior experience with probability and statistics is helpful but not strictly necessary. Students should have prior experience with coding at least at the level of an introductory undergraduate course, or from the first course in the Applied Math sequence.

Textbook: This course will follow the book Pattern Recognition and Machine Learning by Bishop (2006 edition). A free PDF version of the book is available on the author's website. There will also be a limited number of copies available from the campus bookstore.

Description: This course will cover statistical and machine learning theory using foundational approaches (*i.e.* not neural-network based). These will include probability theory, regression, classification, kernel methods, mixture models and expectation maximization, as well as inference for sequential data using hidden Markov models and linear dynamical systems.

Learning outcomes: By the end of this course, students will have obtained the following skills:

- Approach problems involving data from the perspective of probabilistic inference, using statistical thinking to draw conclusions about patterns and relationships in data, as well as to quantify the degree of confidence about these conclusions.
- Derive the equations underlying standard machine-learning algorithms mathematically.
- Implement standard machine-learning algorithms from scratch in Python.
- Use standard built-in implementations of machine-learning algorithms in Python.

- Use simulated data that is designed to contain some interesting structure to develop understanding of machine-learning algorithms.
- Apply machine-learning algorithms to real data to quantify the patterns and relationships that it contains.

Class format: This class will be held in person. Class time will consist of blackboard lectures combined with interactive coding demonstrations. The coding demonstrations will be based on tutorial Python notebooks, which can be found on the following webpage:
https://github.com/murray-lab/statistical-learning

Homework: Homework assignments will be posted on the webpage given above and will consist of a mix of pen-and-paper calculations together with implementations and applications of machine learning algorithms to real and synthetic data using Python notebooks. Homework assignments will be submitted via Canvas, and solutions to homework assignments will be posted to Canvas approximately one week after each assignment due date. Please do not share these solutions with anyone who is not currently enrolled in the class. Details about how to complete and turn in homework assignments can be found on the course's Canvas page.

During the term, each student will be granted a total of 7 days to be late on homework assignments. This means that one assignment may be turned in one week late without penalty during the term, or seven assignments may be turned in one day late, or any other combination. After the 7 days have been used up, late assignments will no longer be accepted and will receive a grade of zero.

Students are welcome and encouraged to discuss and work together on their homework assignments in small groups. So that all students have a chance to participate, this works best in groups of 2-4 students. Each student must complete and submit their work independently, however, and under no circumstances should a student copy work directly from another student.

Attendance: Attendance, while strongly encouraged, is not required for this course.

Expected classroom behavior: Students are expected to behave respectfully toward each other and toward the instructor during class time. This includes refraining from using cell phones during lectures.

Evaluation: Homework assignments will be a mix of pen-and-paper calculations and Python notebooks applying algorithms to real or synthetic data. Midterm and final exams will be taken on paper and in person. Final grades will be weighted as 60% homework, 20% midterm, and 20% final exam. The final take-home exam will be due on the date of the formally assigned exam date during finals week. If a student is not able to complete either exam at the scheduled time, they must provide the instructor with advance notice and arrange a make-up period for the exam to be taken.

Schedule: This course will cover the following topics, following the textbook fairly closely. The following schedule is approximate and subject to change.

- Week 1: Introduction to machine learning and probability theory (Chapter 1, pages 1-33). Introduction and overview of machine learning. Introduction to probability theory. Bayes's theorem, statistical models, and curve fitting.
- Week 2: Probability distributions (Chapter 2, pages 67-102). Discrete and continuous probability distributions. Multivariate Gaussian distribution. Prior distributions.
- Week 3: Linear regression (Chapter 3, pages 137-165). Linear regression as a statistical model. Bias-variance decomposition. Bayesian linear regression.
- Week 4: Classification, part 1 (Chapter 4, pages 179-196). Introduction to classification. Fisher linear discriminant. Fisher linear discriminant coding example.
- Week 5: Classification, part 2 (Chapter 4, pages 196-210). Probabilistic generative models for classification. Linear/quadratic discriminant analysis. Logistic regression.
- Week 6: Kernel methods, part 1 (Chapter 6, pages 291-315; Chapter 7, pages 325-344). Introduction to kernel approaches. Gaussian processes. Support vector machines.
- Week 7: Kernel methods, part 2; Mixture models and expectation maximization, part 1 (Chapter 9, pages 423-439). Support vector machine coding example. K-means clustering and K-means coding example. Introduction to Gaussian mixture models.
- Week 8: Mixture models and expectation maximization, part 2 (Chapter 9, pages 430-455). Gaussian mixture models. Expectation maximization. Expectation maximization for Bernoulli mixtures.
- Week 9: Sequential data, part 1 (Chapter 13, pages 605-635). Hidden Markov models, part 1. Hidden Markov models, part 2. Hidden Markov model coding example.
- Week 10 (tentative): Sequential data, part 2; Principal components analysis and continuous latent variables (Ch. 13, pages 635-644; Ch. 12, pages 559-586). Linear dynamical systems. Principal components analysis. Probabilistic principal components analysis.

Artificial Intelligence: Generative AI algorithms such as Chat-GPT have recently emerged as a powerful technology that can be highly useful for the sorts of coding problems that we will be doing in class. Students are welcome and encouraged to make use of such tools in completing homework assignments, but they may not be used during the midterm and final exams.

What follows is a general description of various University of Oregon policies, not specific to this course.

Academic Disruption due to Campus Emergency: In the event of a campus emergency that disrupts academic activities, course requirements, deadlines, and grading percentages are subject to change. Information about changes in this course will be communicated as soon as possible by email, and on Canvas. If we are not able to meet face-to-face, students should immediately log onto Canvas and read any announcements and/or access alternative assignments. Students are also expected to continue coursework as outlined in this syllabus or other instructions on Canvas. In the event that the instructor of this course has to quarantine, this course may be taught online during that time.

Academic Misconduct: The University Student Conduct Code, available at https://conduct.uoregon.edu, defines academic misconduct. Students are prohibited from committing or attempting to commit any act that constitutes academic misconduct. By way of example, students should not give or receive (or attempt to give or receive) unauthorized help on assignments or examinations without express permission from the instructor. Students should properly acknowledge and document all sources of information (e.g. quotations, paraphrases, ideas) and use only the sources and resources authorized by the instructor. If there is any question about whether an act constitutes academic misconduct, it is the students' obligation to clarify the question with the instructor before committing or attempting to commit the act. Additional information about a common form of academic misconduct, plagiarism, is available at https://researchguides.uoregon.edu/citing-plagiarism.

Accessible Education: The University of Oregon is working to create inclusive learning environments. Please notify me if there are aspects of the instruction or design of this course that result in disability-related barriers to your participation. You are also encouraged to contact the Accessible Education Center in 360 Oregon Hall at 541-346-1155 or uoaec@uoregon.edu.

Basic Needs: Any student who has difficulty affording groceries or accessing sufficient food to eat every day, or who lacks a safe and stable place to live and believes this may affect their performance in the course is urged to contact the Dean of Students Office (346-3216, 164 Oregon Hall) for support. This UO webpage includes resources for food, housing, healthcare, childcare, transportation, technology,finances, and legal support: https://blogs.uoregon.edu/basicneeds/food.

Inclement Weather: It is generally expected that class will meet unless the University is officially closed for inclement weather. If it becomes necessary to cancel class while the University remains open, this will be announced on Canvas and by email. Updates on inclement weather and closure are also communicated in other ways described here: https://hr.uoregon.edu/about-hr/campus-notications/inclement-weather.

Mental Health and Wellness: Life at college can be very complicated. Students often feel overwhelmed or stressed, experience anxiety or depression, struggle with relationships, or just need help navigating challenges in their life. If you're facing such challenges, you don't need to handle them on your own—there's help and support on campus.

As your instructor, if I believe you may need additional support, I will express my concerns, the reasons for them, and refer you to resources that might be helpful. It is not my intention to know the details of what might be bothering you, but simply to let you know I care and that help is available. Getting help is a courageous thing to do—for yourself and those you care about.

University Health Services help students cope with difficult emotions and life stressors. If you need general resources on coping with stress or want to talk with another student who has been in the same place as you, visit the Duck Nest (located in the EMU on the ground floor) and get help from one of the specially trained Peer Wellness Advocates. Find out more at health.uoregon.edu/ducknest.

University Counseling Services (UCS) has a team of dedicated staff members to support you with your concerns, many of whom can provide identity-based support. All clinical services are free and confidential. Find out more at counseling.uoregon.edu or by calling 541-346-3227 (anytime UCS is closed, the After-Hours Support and Crisis Line is available by calling this same number).