BINF 760
Machine Learning for Bioinformatics

Fall 2024

School of Systems Biology
George Mason University
Manassas, VA

Instructor: Chris Lockhart
Email: clockha2@gmu.edu (preferred)
Phone: Microsoft Teams
Office: Virtual
Office Hours: By appointment (https://calendly.com/chrislockhart/office-hours)

Meeting Place: Online via Canvas
Meeting Time: Asynchronous
Course Website: Canvas

Credits: 3

Course Description

This course explores machine learning and data mining methods relevant to applications and problems in computational biology. Methods include logistic regression, Naïve Bayes, support vector machines, neural networks, decision trees, and random forests. Applications include cancer prediction, gene finding, protein function classification, and other recent bioinformatics applications selected from the literature. In addition to lectures from the instructor, students will present papers from the literature and complete a machine learning project.

Recommended Prerequisites: BINF 630, BINF 631, and BINF 634, or permission of instructor. Student experience in programming is expected.

Course Material

Three optional reference texts are listed below. They are available to download without charge directly from the authors.

Bishop (2006). *Pattern Recognition and Machine Learning*. Springer.
Download at https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/

Hastie, Tibshirani, & Friedman (2017). *The Elements of Statistical Learning.* Springer.

Download at https://web.stanford.edu/~hastie/ElemStatLearn/

Shalev-Shwartz & Ben-David (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.
Download at https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/

Tentative Course Topics

Lecture	Topic	Recommended Reading
1	Introduction to machine learning	UML – 1.1-1.4
2	Linear regression	ESL – 2.3.1, 3.1, 3.2
3	Naïve Bayes	PRML – 1.2 ESL – 6.6.3
4	Logistic regression	ESL – 4.4
5	Support vector machines	ESL – 12.1-12.3
6	Neural networks	PRML – 5.1, 5.3
7	Model and feature selection	ESL – 7.1-7.3, 7.10
8	Decision trees	UML – 18.2
9	Random forests	ESL – 8.7, 10.1, 15.1-15.3
10	Genetic algorithms	
11	Hidden Markov models	PRML – 13.2
12	Principal component analysis	UML – 23.1
13	Clustering	PRML – 9.1
14	Prospective on deep learning	

Learning Outcomes

By the end of this course, students will be able to:

1. Discuss various machine learning algorithms and identify their strengths and weaknesses
2. Design a machine learning pipeline, which includes dataset preparation, testing/training, performance evaluation, and hyperparameter tuning
3. Apply machine learning techniques to bioinformatics and computational biology problems
4. Appraise recent literature in bioinformatics and computational biology that apply machine learning

Course Policies

Grading scale (points): A+ (≥100), A (90-99.9), B+ (87-89.9), B (80-86.9), C (<80).

Grading policy: Students will be graded on homework (30%), participation (30%), and a final project (40%).

- There will be weekly homework assignments, which will feature theoretical questions and programming exercises related to that week's lecture. Each assignment will be equally weighted.
- Class participation will be met by (1) once during the semester finding a recent peer-reviewed article that applies machine learning to biological data, writing 1-2 paragraphs on the article, and posting this information to Canvas for discussion and (2) asking critical questions of the papers posted by other students. To earn full credit, students must contribute to the discussion of 4 articles throughout the semester.
- The final project will require students to complete their own machine learning project from topics prepared by the instructor. Students may pursue their own topics with permission. The project will be presented in a written report.

Late assignments and make-up work: In case of illness or quarantine, please contact the instructor to set up a plan for make-up work. Late assignments will not be accepted unless due to emergency, illness, quarantine, work-related, or other documented reasons.

Course recordings: All lectures in this class will be recorded to provide necessary information for students in this class. Recordings will be stored on Canvas and will only be accessible to students taking this course during this semester.

Other considerations: If there are any schedule issues related to religious holidays, please inform the instructor the first week of class.

Course Logistics

Content distribution: The course uses Canvas for distributing lecture materials, submission of homework, and grading. Canvas can be accessed by visiting https://canvas.gmu.edu/ and logging in with your MasonID and password.

Communication: I will use Mason email to distribute class updates and communicate with students (see Email section in Student Responsibilities). If you wish, please share your name and gender pronouns with me and how best to address you in class and via email. I use he/him for myself, and you may address me as Chris or Dr. Lockhart in email and verbally. Communication over email is largely preferred, and I will respond to student emails promptly within 48 hours.

Course Technology Requirements

Software and hardware: This course uses Canvas as a learning management system available at https://canvas.gmu.edu/. Students are required to have regular, reliable

access to a computer with an updated operating system (recommended: Windows 10 or Mac OS X 10.15 or higher) and a stable broadband Internet connection (cable modem, DSL, satellite broadband, etc., with a consistent 1.5 Mbps download speed or higher).

Course-specific software: This course will use Python (the Python distribution from Anaconda https://www.anaconda.com/ is recommended).

Technical help: If you have difficulty with accessing Canvas, please contact the ITS Support Center at (703) 993-8870 or support@gmu.edu. If you have trouble with using the features in Canvas, email courses@gmu.edu.

Student Responsibilities

Email: Students must use their Mason email account to receive important University information, including communications related to this class. Per University policy, I will not respond to messages sent from or send messages to a non-Mason email address.

Academic integrity: The integrity of the University community is affected by the individual choices made by each of us. Mason has an Honor Code with clear guidelines regarding academic integrity. Three fundamental and rather simple principles to follow always are that: (1) all work submitted be your own; (2) when using the work or ideas of others, including fellow students or online sites, give full credit through accurate citations; and (3) if you are uncertain about the ground rules on a particular assignment, ask for clarification. No grade is important enough to justify academic misconduct. Plagiarism means using the exact words, opinions, or information from another person without giving the person credit. Writers give credit through accepted documentation styles, such as parenthetical citation, footnotes, or endnotes. Paraphrased material must also be cited, using the appropriate format for this class. A simple listing of books or articles is not sufficient. Plagiarism is the equivalent of intellectual robbery and cannot be tolerated in the academic setting. If you have any doubts about what constitutes plagiarism, please see me, or consult the Academic Integrity website at https://oai.gmu.edu/.

Generative AI policy: Use of generative AI (e.g., ChatGPT) is strongly discouraged. If students use these tools, they must follow the fundamental principles of the Honor Code. This includes being honest about the use of generative AI for submitted work and giving credit through accurate citations.

Disability accommodations: Disability Services at George Mason University is committed to providing equitable access to learning opportunities for all students by upholding the laws that ensure equal treatment of people with disabilities. If you are seeking accommodations for this class, please first visit http://ds.gmu.edu/ for detailed information about the Disability Services registration process. Then please discuss your approved accommodations with me. Disability Services is in Student Union Building I (SUB I), Suite 2500. Email: ods@gmu.edu | Phone: (703) 993-2474.

Student Services

University writing center: Take advantage of the Writing Center as you work on written assignments in this course. You can book a free 45-minute appointment to meet with a tutor on Zoom or to submit a draft for written feedback. Tutors will work with you on any phase of a writing project. They can help you develop your ideas, provide feedback on a draft, answer your questions, and show you strategies for brainstorming, organizing, drafting, revising, and editing. To schedule an appointment, go to writingcenter.gmu.edu, register with the center, and make an appointment using the online scheduler. Watch this short video (https://youtu.be/LA-B0Szoe28) for more detailed guidance on making an appointment and send any questions to wcenter@gmu.edu.

University Libraries: University Libraries provides resources for distance learning students (See the Library website: https://library.gmu.edu/for/online).

Counseling and psychological services: The George Mason University Counseling and Psychological Services (CAPS) staff consists of professional counseling and clinical psychologists, social workers, and counselors who offer a wide range of services (e.g., individual and group counseling, workshops and outreach programs) to enhance students' personal experience and academic performance (See the Counseling and Psychological Services website: https://caps.gmu.edu).

Family Educational Rights and Privacy Act (FERPA): The Family Educational Rights and Privacy Act of 1974 (FERPA), also known as the "Buckley Amendment," is a federal law that gives protection to student educational records and provides students with certain rights (See the Registrar's Office website: registrar.gmu.edu/ferpa/).

Course Materials and Student Privacy

Video recordings of class meetings that are shared only with the instructors and students officially enrolled in a class do not violate FERPA or any other privacy expectation. Video recordings that only include the instructor (no student names, images, voices, or identifiable texts) may be shared without violating FERPA (but see University Policies: Privacy, for some qualifications and recommendations). All course materials posted to Canvas or other course site are private to this class; by federal law, any materials that identify specific students (via their name, voice, or image) must not be shared with anyone not enrolled in this class.

Video conferencing or recordings: Video recordings – whether made by instructors or students — of class meetings that include audio, visual, or textual information from other students are private and must not be shared outside the class. Live video conference meetings (e.g., Zoom) that include audio, textual, or visual information from other students must be viewed privately and not shared with others in your household or recorded and shared outside the class.