BINF 760
Machine Learning for Bioinformatics

Fall 2025

School of Systems Biology
George Mason University
Manassas, VA

Instructor: Dr. Chris Lockhart
Email: clockha2@gmu.edu (preferred)
Phone: Microsoft Teams
Office: Virtual
Office Hours: By appointment (Microsoft Bookings)

Meeting Place: Online (https://gmu.zoom.us/j/93521969668)
Meeting Time: Wednesdays, 7:20-10pm
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.

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 Material

Three freely available optional reference texts are listed below.

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	Date	Topic
1	Aug 27	Introduction to machine learning
2	Sept 3	Linear regression ***Homework 1 assigned***
3	Sept 10	Naïve Bayes
4	Sept 17	Logistic regression ***Homework 2 assigned***
5	Sept 24	Support vector machines ***Course project assigned***
6	Oct 1	Neural networks ***Homework 3 assigned***
7	Oct 8	Model and feature selection
8	Oct 15	Decision trees ***Homework 4 assigned***
9	Oct 22	Random forests ***Project topic must be decided***
10	Oct 29	Unsupervised learning ***Homework 5 assigned***
11	Nov 5	Genetic algorithms
12	Nov 12	Hidden Markov models ***Homework 6 assigned***
13	Nov 19	No-code and AutoML
14	Dec 3	Prospective on deep learning

Each lecture is a 2½ hour presentation with a 10-minute break.

Course Policies

Grading scale (points): A+ (≥100), A (94-99), A- (90-93), B+ (87-89), B (84-86), B- (80-83), C (60-79), F (<60). Final grades will be rounded to the nearest whole number to assign letter grades.

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

- There will be 6 homework assignments, which will feature theoretical questions and/or programming exercises related to that week's lecture. Each assignment will be equally weighted.
- Journal club 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 apply machine learning to a biological dataset of interest. Students will present their work in a written report.

Assignment resubmissions: Grades posted for submitted assignments are final. Assignments cannot be revised or resubmitted after grading.

Late assignments: Late assignments will be penalized based on the number of days late but will not be accepted after assignments have been reviewed in class or after answers have been posted online. Extensions may be granted due to emergency, illness, quarantine, work-related, or other documented reasons. Except in emergency situations, extension requests should be made before the assignment due date.

Course recordings: All synchronous meetings will be recorded 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.

Virtual classroom and office hours: Zoom will be used for online lectures and office hours. Zoom lecture links will be distributed by the instructor.

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. Communication over email is largely preferred, and I will respond to student emails promptly within 24 hours on weekdays. I do not typically respond to emails after 8pm or on weekends.

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). Activities and assignments in this course will use web-conferencing software (Zoom). In addition to the requirements above, students are required to have a device with a functional camera and microphone.

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 standards: Some kinds of participation in online study sites violate GMU's Academic Standards: these include accessing exam or quiz questions for this class; accessing exam, quiz, or assignment answers for this class; uploading of any of the instructor's materials or exams; and uploading any of your own answers or finished work. Always consult your syllabus and your professor before using these sites. See https://academicstandards.gmu.edu/ for additional information.

Use of AI tools: Use of generative AI (e.g., ChatGPT) is strongly discouraged because these tools do not have access to the specific data used in this course. If students use these tools, they must follow GMU's academic standards. This includes being honest about the use of generative AI for submitted work and giving credit through accurate citations.

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.

Common Course Policies

This course adheres to the common course policies set by George Mason University, which includes polices about Academic Standards, Accommodations for Students with Disabilities, FERPA, and Title IX. These policies are described in more detail at the following link: https://stearnscenter.gmu.edu/home/gmu-common-course-policies/.