6.8700/6.8701/HST.507
Machine Learning for Computational Biology
Fall 2023 Course Information
Lectures: Tue/Thu, 1-2:30PM, 32-144
Office hour: Fri 3PM, 32-144
Units: 12 (4-0-8)
Prerequisites: Algo (6.006), Prob (6.041), Bio (7.01)

Course information

Professor: Manolis Kellis: 32-D524, manoli@mit.edu; Regina Barzilay: 32-G468, regina@csail.mit.edu
Teaching Assistants: Jeremy Wohlwend (jw2016@mit.edu), Na Sun (nasun@mit.edu), Benjamin James (benjames@mit.edu)
Recitations/Mentoring/Office Hours: On Friday at 3pm in 32-144
Contact: MLCB@mit.edu, mlcb-tas@mit.edu
Course website: http://compbio.mit.edu/MLCB

Grading

Your grade in this course will be based on the following:

- Problem sets (30)
- Online quiz (25)
- Participation (5)
- Final project (40)
  - Self-introduction (2)
  - Paper selection and project ideas (3)
  - Proposal (10)
  - Midcourse report (5)
  - Final slides (5)
  - Final presentation (5)
  - Final report (10)

Problem Sets

There will be four problem sets during the semester, each including 3–5 problems for all students and a lab problem which is optional for 6.8701 students. The problem sets will include both theoretical and programming problems. For programming problems, we provide skeleton code in Python, but you are welcome to write solutions in any language. Submit all homeworks online from canvas page No solutions distributed.

Online quiz

There will be one quiz which will cover all material covered up to that point. There will be no final exam. The quiz will include true/false questions, short answer questions, practical problems using algorithms covered in class, and one or two problems extending ideas seen in class.
Paper presentation

There will be a short presentation on a paper of your choosing. We highly recommend that you choose a paper from a high-impact journal (Cell, Nature, Science, etc.) that is related to your intended project. This is a great opportunity to familiarize yourself with expectations for the final presentation and scientific presentations in general.

Final project

Students will work on a final project with deliverables due at several milestones during the term as marked on the course schedule. The first part of the term will be spent identifying a topic relevant to the course material, planning the project, writing an NIH-style proposal, and reviewing the proposals of your peers. The second part of the term will be focused on completing the project, writing the report, and presenting the results. Details of what is expected by each milestone will be posted on the course website.

We expect students to work in groups of three on the final project. However, we will consider groups of one or two if presented with a strong argument for it. Keep in mind that larger teams are expected to undertake more ambitious projects. Part of the final project grade will depend on the challenge and originality of your project.

We anticipate projects of a few types:

- Identify a biological problem, gather relevant datasets, design and implement new algorithms, apply the methods, and interpret the results
- Rigorously compare several algorithms which solve the same biological problem in terms of their performance and the quality of their outputs on synthetic and real data sets

Recitations/Mentoring/Office hours

Recitation/mentoring/office hours will be held on Fridays. During recitations, we will both review the lecture material and discuss additional aspects of it. Since there is only one recitation section, we will not be able to accommodate all scheduling conflicts. Therefore, attendance is not mandatory and recitation notes and videos will be made available on the course website. Material in the recitation notes may appear on the quiz. During mentoring section, we will provide feedback on project proposal, progress and reports, etc. During office hours, we will answer any questions related to problem sets and project.

Textbooks

The course textbook is comprised of scribe notes and is available from the course website. You may also find the following optional texts helpful:

- Richard Duda, Peter Hart, David Stork, *Pattern Classification*.

Collaboration Policy

You are welcome to collaborate on problem sets and the final project. However:

- You must work independently on each problem before you discuss it with others.
- You must write the solutions on your own.
- You must acknowledge outside sources and collaborators.

You may use ChatGPT/LLM, but you must:

- Acknowledge the way you would for a collaboration partner.
• Be transparent, save your chats, submit with homework.