COMP683: Computational Biology

Instructor: Natalie Stanley

Spring 2024

E-mail: natalies@cs.unc.edu
Web: https://compcylab.squarespace.com/
Class Meeting: Tuesday/Thursday 9:30am-10:45am
Office Hours: Monday 10:45am-noon
Office: Sitterson Hall 305
Class Room: Fred Brooks 007

Course Description

Modern, high-throughput assays allow us to efficiently profile a variety of biological processes to gain a systems-level understanding of health and disease. Recent technologies and experimental assays generate an abundance of detailed information that needs to be extracted, summarized, and interpreted. In this course we will discuss the methodology used to extract signal from (e.g. process, engineer features from, combine, etc.) data generated by some of the most cutting-edge technologies, such as single-cell and imaging assays. We will go into detail about the methods and theory underlying bioinformatics algorithms, originating from numerical linear algebra, graph-signal processing, and machine learning. While computational biology is a very broad field, we will focus here on applications in single-cell biology (CyTOF, single-cell RNA sequencing), multiomics/multi-modal analysis, systems immunology, and benchmarking. For each class of algorithms introduced for some task on biological data, we will also go over necessary theory and mathematical intuition. The course covers the foundations for biomedical data science and does not assume any biological knowledge. **This course is a semester-long 3-credit course, appropriate for graduate students and advanced undergraduate students with an interest in applying machine learning techniques to biomedical applications.**

Course Goals and Learning Outcomes

As a result of the course, students should gain the following conceptual understandings:

- Fluency in constructing and analyzing graph-based data structures.
- Competency in the critical aspects of single-cell analysis, including, pre-processing and normalization, imputation, clustering and automated cell-population discovery, robust experimental design, multi-sample analysis, linking particular cells to external variables, feature engineering, and visualization
- Fluency in statistical and graph-based methods for integrating multiple data modalities

- Developed intuition in how to objectively evaluate and benchmark various bioinformatics approaches.

Furthermore, students will gain hands-on experience with the following:

- Automated cell-population discovery, feature engineering from single-cell data, all aspects of CyTOF (cytometry by time of flight) data analysis
- Integrating and interpreting datasets acquired across multiple biological modalities
- Formulating and carrying out a novel research question and in computational biology through a course project and paper-style write-up.

Schedule, Notes, Readings, Code, Homework

I post a highly dynamic schedule containing readings, lecture notes, code and homework in a Git Repository found here, https://github.com/natalies-teaching/Comp683_CompBio_2024. Please be sure to check the Git Repo before every class meeting to get a heads-up about what will be covered on the particular day. I also use Canvas to make announcements, post grades, and collect homework.

Please find an example specific schedule of lectures appended at the end of this document.

Prerequisites

This course is meant to be applicable to a broad audience. Students should be strong in programming in Python or Julia or R, and be comfortable with linear algebra and basic probability. **I do not assume any prior biological knowledge.** Any relevant concepts will be introduced. In addition, students should have taken the following courses with a grade of C or better,

- **MATH577**: Linear Algebra or MATH347 Linear Algebra for Applications.
- One of → **COMP562**: Machine Learning or **STOR520**: Statistical Computing for Data Science or **STOR565**: Machine Learning

Course Structure

This course will be mostly lecture-based with two homework assignments and a course project. I will provide ideas for several publicly available biological datasets and open problems for you to work on for these projects. Overall, the project is intended to give you an opportunity to implement/apply methodology discussed in the papers that we will discuss together. The final project writeup will also give you practice writing up results and communicating ideas. You are welcome to work on teams for this project. Students will also pick any two days during the semester to answer a set of reading questions about one of the assigned papers.

- **Important**: This is a 3-credit full-semester course and fulfills the 'Applications' category for CS students. It is a lecture-style class (I will teach the lectures) and includes two homework assignments and a course project.

Topics

For a detailed, daily list of topics, please see the Git Repository from last year. Note that I will be making an updated course page soon for the 2024 version of the class. However, the general themes of the course can be summarized as follows:

- Linear algebra review and computations on graphs
- Graph Signal Processing
- Automated Cell-Population Discovery
- Imputation for Single-Cell Data
- Differential Abundance Analysis and Identifying Condition-Specific Prototypical Cells
- Trajectory Inference and Pseudotime Estimation
- Multimodal Data Integration in Biomedicine [multiomics]
- Spatial Profiling and Imaging Modalities
- Graph Neural Networks and Applications in Spatial Omics
- Technical Writing in Computational Biology and Writing and Communicating to an Interdiscplinary Audience.

Homework

There will be two homework assignments to practice implementing particular concepts. Often, things can become a bit easier to understand and use when they are implemented by you. I will provide code and hints in Python, but will be happy to read/run code written in Python, R, Julia, or Matlab. **Only homework submitted as a single PDF will be graded**.

Project

I will provide you with several examples of publicly available biological datasets and problems (Datasets and Problems). Half-way through the semester, you will submit your project proposal and present your idea to the class. The proposal will be a short document describing 1) The problem 2) A background on other people's attempts to solve this problem and 3) A background on your idea of a solution and 4) the data you will use to test your method. At the end of the semester you will write a short paper explaining your method and results and present your results. See the following for a successful project that resulted in a workshop paper `https://icml-compbio.github.io/2021/papers/WCBICML2021_paper_42.pdf`.

Grading

Grading will be based on the following:

- Reading Questions : 15% [complete two per semester]
- Homework 1: 20% [Due February 24]
- Homework 2: 20% [Due April 21]
- Project Proposal : 10% [Due March 8]
- Project final writeup: 30% [Due May 1]
- Class Participation and Attendance : 5%

The following cutoff will be used to assign grades.

- 97-100 $\rightarrow$ **A+**
- 93-95 $\rightarrow$ **A**
- 90-92 $\rightarrow$ **A-**
- 87-89 $\rightarrow$ **B+**
- 83-85 $\rightarrow$ **B**
- 80-82 $\rightarrow$ **B-**
- 77-79 $\rightarrow$ **C+**
- 73-75 $\rightarrow$ **C**
- 70-72 $\rightarrow$ **C-**
- 67-69 $\rightarrow$ **D+**
- 63-65 $\rightarrow$ **D**
- 60-62 $\rightarrow$ **D-**
- $< 60 \rightarrow$ **F**

Note all due-dates imply that the item should be turned in by 11:59pm that day.

Background Resources

Most of what we discuss in class will come from papers. However, I suggest the following textbooks as background references. Conveniently, they are also available for free.

- **PRML** Pattern Recognition and Machine Learning– Chris Bishop [Link]
- **SLMP** Spectral Learning on Matrices and Tensors – Majid Janzamin et al. [Link]
- The Matrix Cookbook [Link]
- **PML** Probabilistic Machine Learning: An Introduction. – Kevin Murphy [Link]
- **GRL** Graph Representation Learning – William Hamilton [Link]

Lateness Policy

10% of the total points will be duducted on homework for every day late. There will be no late projects or project proposals accepted. Reading assignments must be turned in by the last day of class to be accepted for credit.

Honor Code Statement

Students are bound by the Honor Code in taking exams and in written work. The Honor Code of the University is in effect at all times, and the submission of work signifies understanding and acceptance of those requirements. Plagiarism will not be tolerated. Please consult with me if you have any questions about the Honor Code.

Accessibility Resources and Services

The University of North Carolina at Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in barriers to fully accessing University courses, programs and activities. Accommodations are determined through the Office of Accessibility Resources and Service (ARS) for individuals with documented qualifying disabilities in accordance with applicable state and federal laws. See the ARS Website for contact information: `https://ars.unc.edu` or email `ars@unc.edu`.

Counseling and Psychological Services

CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to their website: `https://caps.unc.edu/`or visit their facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more.

Title IX Resources

Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Reports can be made online to the EOC at https://eoc.unc.edu/report-an-incident/. Please contact the University's Title IX Coordinator (Elizabeth Hall, interim –titleixcoordinator@unc.edu), Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu), Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (gvsc@unc.edu; confidential) to discuss your specific needs. Additional resources are available at safe.unc.edu.

Policy on Non-Discrimination

The University is committed to providing an inclusive and welcoming environment for all members of our community and to ensuring that educational and employment decisions are based on individuals' abilities and qualifications. Consistent with this principle and applicable laws, the University's Policy Statement on Non-Discrimination offers access to its educational programs and activities as well as employment terms and conditions without respect to race, color, gender, national origin, age, religion, creed, genetic information, disability, veteran's status, sexual orientation, gender identity or gender expression. Such a policy ensures that only relevant factors are considered and that equitable and consistent standards of conduct and performance are applied. If you are experiencing harassment or discrimination, you can seek assistance and file a report through the Report and Response Coordinators (see contact info at safe.unc.edu) or the Equal Opportunity and Compliance Office, or online to the EOC at https://eoc.unc.edu/report-an-incident/.

University Attendance Policy

University Policy: No right or privilege exists that permits a student to be absent from any class meetings, except for these University Approved Absences: Authorized University activities Disability/religious observance/pregnancy, as required by law and approved by Accessibility Resources and Service and/or the Equal Opportunity and Compliance Office (EOC) Significant health condition and/or personal/family emergency as approved by the Office of the Dean of Students, Gender Violence Service Coordinators, and/or the Equal Opportunity and Compliance Office (EOC). Class Policy: Note: Instructors may work with students to meet attendance needs that do not fall within University approved absences. For situations when an absence is not University approved (e.g., a job interview or club activity), instructors determine their own approach to missed classes and make-up assessments and assignments. Please provide your approach on the course syllabus. Example: Please communicate with me early about potential absences. Please be aware that you are bound by the Honor Code when making a request for a University approved absence.

Diversity Statement

I value the perspectives of individuals from all backgrounds reflecting the diversity of our students. I broadly define diversity to include race, gender identity, national origin, ethnicity, religion, social class, age, sexual orientation, political background, and physical and learning ability. I strive to make this classroom an inclusive space for all students. Please let me know if there is anything I can do to improve, I appreciate suggestions.

Syllabus Changes

The instructor reserves the right to make changes to the syllabus, including project due dates and test dates. These changes will be announced as early as possible.