Comp790-166: Computational Biology

Instructor: Natalie Stanley

Spring 2023

E-mail: `natalies@cs.unc.edu`
Class Meeting: M/W 11:00-12:15
Office: Sitterson Hall 305

Web: **https://compcylab.squarespace.com/**
Office Hours: Monday 2:15-3:15
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 assays and imaging. 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.

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/CompBio2023`. 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 sakai to make announcements, post grades, and collect homework (`https://sakai.unc.edu/portal/directtool/efa6e56c-9b54-4bf8-bdb9-4f28e0e3baf9/`).

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.

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. Please make sure you selected the 3-credit option when you enrolled.

Topics

For a detailed, daily list of topics, please see the Git Repository. 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 Inter-discplinary 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`.

Reading Questions

You may pick any two days to answer reading questions for. This means that on a particular day, you may choose one of the assigned papers to answer the following questions about:

1. Please explain in 2 sentences or less what the problem being solved is.

2. What were the main contributions of the authors in this work? (You can answer in a few bullet points).

3. Please describe 1-2 computational experiments that the authors implemented to test their method.

4. Were the authors the first to attempt this particular problem? If not, did they compare their results to other baselines? Do you think that their evaluation was objective?

5. Do you think that the authors provided enough evidence for why their developed method is an important contribution? If yes, please describe their reasoning here. If you do not think they adequately justified why they worked on this particular problem, please describe your thoughts on that here.

6. What is one follow-up idea or extension from this work?

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%

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.

Accessibility Statement

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.

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.