GCB 537 2016A Adv. Computational Biol

Jump to Today

GCB537 Advanced Computational Biology

Term: Spring 2016
Instructors: Yoseph Barash, Li-San Wang
TA: Anupama Jha

Objectives:
1. Learn important concepts/methods from computer science and statistical data analysis as they are applied in computational biology.
2. Learn about current topics in genomics and computational biology through in depth discussion of classic and recent papers.
3. Learn to evaluate, criticize, summarize, and present research papers in genomics and computational biology.
4. Gain hands on experience coding, experimenting, and evaluating tools/algorithms as they apply to topics covered in class and/or their final project. NOTE:

Requirement: Background in statistics, biology, genetics and genomics, and computer science.

This is NOT a bioinformatics lab.
Non-GCB students need to be approved by the instructors.

Time and Location:
Tuesdays 13:30-15:00 251 BRB
Thursdays 13:30-15:00 252 BRB

Office Hours:
Yoseph - Friday 3-4:30pm, Richards D205
Anupama - Friday 3-4:30pm, Richards D201


Course format:
The course is comprised of lectures, assignments, paper discussions, and a final project presentation. Lectures covers material in computer science and computational biology relevant for the topics covered in the course. Assignments will include written and programming assignments. In paper discussion classes, papers are selected to cover a range of important topics, with emphasis on algorithm design and best practice for data interpretation and presentation. Paper discussions are first led by instructors followed by discussions led by students. Students will team up to work towards a guided evaluation project of algorithms/tools in one of the topics chosen by the instructors. The students will survey the literature and present their research plan at mid-semester. At the end of the semester each team will present their findings and submit a written summary of their work as the final project.

Paper discussion:
The course is divided into units covering current topics in comp bio research and units that cover topics in ML and statistical data analysis. Each comp bio topic unit starts with a review lecture, followed by one or more paper discussion classes led by the instructors or students. Emphasis will be given to understanding the computational methods, model assumptions, evaluation process, overall significance and open issues/directions. To ensure the quality of the presentation, send PowerPoint files to the instructor or discuss with the instructor at least two days before the scheduled presentation.
The leading team will also submit questions on the paper to the instructors two days before the discussion. The question list will be circulated to the entire class before the class starts. Student teams will volunteer or be picked at random to answer these questions. After the class, all other students at the presentation will send a grade (between 1=unprepared and 5=excellent) + constructive feedback to the instructor by email by end of the day. Comments will be forwarded to the presenters anonymously.
The instructors will tally the number and quality of questions and answers presented and include as part of the grade.

Term project -- Experiment/Evaluate/Improve in a topic of choice:
The students will team up (2-3 students per team, 7 teams maximum) and:
Work with the instructors to select a specific topic to evaluate tools on a particular bioinformatic/computational topic by end of February. Preference would be given to
extending in a topic already presented in the course or related ones. Other topics may be approved as well on a special case basis if there is a strong drive from a specific group.
Submit a 1-2 page proposal for final project by end of February: Topic, what questions you aim to answer, datasets and experimental/analysis plan to answer those questions. Discuss with the instructors. Make a 30-minute presentation towards the end of the semester.
After the class, all other students at the presentation will send a grade (between 1=unprepared and 5=excellent) and any constructive feedback to the instructor by email by end of the day; comments will be forwarded to the presenters anonymously.
Submit a summary report that includes: Background, the main issues they addressed/evaluated, their results, and conclusions. Summary is no longer than 4 pages 12pt font, 1’ margin.

**Coding Environment:**

Students will program in Python. All solutions are expected to run on the [PMACS HPC](https://hpcwiki.genomics.upenn.edu/index.php/GCB_537:Main_Page) within the virtual cluster environment created for the course. In addition, students are encouraged to use Python coding environments/IDEs for their code development. To aid in this two different Python coding environments were installed on the GCB students computers: [Canopy](https://www.enthought.com/products/canopy/) and [PyCharm](https://www.jetbrains.com/pycharm/). These two Python coding environments are free for academic use so students should be able to also install these on their own computers if they wish. The Canopy academic license includes online lectures and coding exercises that may help students less familiar with Python, NumPy, and SciPy.

**Grading:**

16% paper presentation (8% graded by other students and 8% graded by instructors)
9% paper answers
10% participation in paper discussion during the class (graded by instructors)
50% written/programming assignments following lectures
10% class participation (graded by instructors)
15% Term presentation and written summary

Grading of each programming exercise is composed of the following:

45% analysis
40% correctness
10% good coding practices
10% efficiency

**Policies**

**Collaboration**

You are allowed and encouraged to discuss the homework with other people to understand the problem and reach a solution. Moreover, since the analysis questions aim to mimic a research scenario, there is not necessarily a single/specific solution and students can take several different approaches (Note: even if there are multiple solutions/approaches there can still exist many things which would be wrong to do). You also have Canvas Announcements and Assignments pages where you can post/view questions/answers. However, for non-coding homework, each student must write down the solution independently, without referring to written notes from others. I.e., you must understand the solution well enough in order to construct it by yourself. Similarly, coding is done by each student independently. Students can only use libraries we supply or their own code. Students can not use any other library/code snippet without specific permission by instructors. Students should report at the end of their exercise submission which resources they consulted (e.g. a paper describing a method).

**Honor code**

The purpose of problem sets in this class is to help you think about the material, not just give us the good answers (see note above about possibly multiple solutions/approaches). You are encouraged to use online resources for learning more about the material covered in class; however, you should not look for or use found solutions to questions in the problem sets. Specifically, you must not look at any code that has been created to solve the assignment, including solutions found on the internet to questions in the problem sets, code created by a student in a previous class or code created by a current classmate. Cheating will be punished according to university regulations as determined by the Office of Student Conduct.

If one student shares code with another on a different team, both the donor and the recipient of the code are in violation of the Penn honor code and will be referred to the Office of Student Conduct.

**Late Policy**

Any homework turned in late will be penalized per late day or fraction of day. The exact penalty per late day will be announced before the exercise submission but you should always check with the course instructors what is the penalty and what is the last possible date for a late submission which will still count as completing the assignment. Completion of all assignments is mandatory to complete the course.
Intro meeting + Review (https://canvas.upenn.edu/calendar?event_id=1970940&include_contexts=course_1306777)

Tue Jan 19, 2016
Review: Topics in Prob, Stats, ML (https://canvas.upenn.edu/calendar?event_id=1979627&include_contexts=course_1306777)

Thu Jan 21, 2016
Motif Finding Review + ML Topics 1 (https://canvas.upenn.edu/calendar?event_id=1970935&include_contexts=course_1306777)

Tue Jan 26, 2016
ML Topics 1 (https://canvas.upenn.edu/calendar?event_id=1979628&include_contexts=course_1306777)

Thu Jan 28, 2016
Publish: Ex1 PnP, Ex1 Prog (https://canvas.upenn.edu/calendar?event_id=1981641&include_contexts=course_1306777)

Paper Presentation: Motif Finding (MEME) - Yoseph (https://canvas.upenn.edu/calendar?event_id=1970928&include_contexts=course_1306777)

Ex0 Prog (https://canvas.upenn.edu/courses/1306777/assignments/5108798) due by

Tue Feb 2, 2016
Review: RNA-Seq (https://canvas.upenn.edu/calendar?event_id=1970939&include_contexts=course_1306777)

Thu Feb 4, 2016
Paper Presentation: RNA-Seq (DESeq) (https://canvas.upenn.edu/calendar?event_id=1970927&include_contexts=course_1306777)

Ex1 PnP (https://canvas.upenn.edu/courses/1306777/assignments/5108814) due by

Tue Feb 9, 2016
Review Ex0 prog, Ex1 PnP + ML Topics 2 (https://canvas.upenn.edu/calendar?event_id=1979630&include_contexts=course_1306777)

Thu Feb 11, 2016
Publish: Ex2 PnP (https://canvas.upenn.edu/calendar?event_id=1981643&include_contexts=course_1306777)

ML Topics 2 (https://canvas.upenn.edu/calendar?event_id=1970948&include_contexts=course_1306777)

Tue Feb 16, 2016
Paper Presentation: RNA-Seq (EXPRESS) (https://canvas.upenn.edu/calendar?event_id=1970926&include_contexts=course_1306777)

Thu Feb 18, 2016
Review: ChIP-Seq (Golnaz Vahedi) (https://canvas.upenn.edu/calendar?event_id=1970950&include_contexts=course_1306777)

Ex1 Prog (https://canvas.upenn.edu/courses/1306777/assignments/5108819) due by

Ex2 PnP (https://canvas.upenn.edu/courses/1306777/assignments/5108823) due by

Tue Feb 23, 2016
Paper Presentation: Peak Calling (ZINBA) (https://canvas.upenn.edu/calendar?event_id=1970949&include_contexts=course_1306777)

Thu Feb 25, 2016
Publish: Ex2 Prog (https://canvas.upenn.edu/calendar?event_id=1981644&include_contexts=course_1306777)

Review Ex1 Prog, Ex2 P&P, start ML Topics 3 (https://canvas.upenn.edu/calendar?event_id=1970924&include_contexts=course_1306777)

Tue Mar 1, 2016
Introduction to human genetics research (https://canvas.upenn.edu/calendar?event_id=1970941&include_contexts=course_1306777)

Thu Mar 3, 2016
Paper Presentation: GWAS (SKAT) (https://canvas.upenn.edu/calendar?event_id=1970942&include_contexts=course_1306777)
Tue Mar 8, 2016
NO CLASS - SPRING BREAK

Thu Mar 10, 2016
NO CLASS - SPRING BREAK

Tue Mar 15, 2016
Review: eQTL mapping - Casey Brown

Thu Mar 17, 2016
Paper Presentation: eQTL

Tue Mar 22, 2016
ML Topics 3

Thu Mar 24, 2016
Publish: Ex3 PnP, Ex3 Prog

ML Topics 4

Ex2 Prog due by

Tue Mar 29, 2016
Review: TBD - Babak

Thu Mar 31, 2016
Review Ex2 Prog, Ex3 PnP

Ex3 PnP due by

Tue Apr 5, 2016
Paper Presentation: Babak's Topic

Thu Apr 7, 2016
Publish: Ex4 PnP

ML Topics 5

Tue Apr 12, 2016
Paper Presentation: Phasing/Imputation

Thu Apr 14, 2016
Paper Presentation - Phasing/Imputation 2

Ex4 PnP due by

Tue Apr 19, 2016
Guest Lecture - TBD

Thu Apr 21, 2016
Guest Lecture - TBD

Ex3 Prog due by

Tue Apr 26, 2016
Review Ex3 prog, Ex4 PnP
Guest Lecture - TBD (https://canvas.upenn.edu/calendar?event_id=1970922&include_contexts=course_1306777)