Course Syllabus

Jump to Today ✎ Edit

GCB537 Advanced Computational Biology

Term: Spring 2019
Instructors: Yoseph Barash
Guest lecturers: Pablo Camera, Casey Brown.

TA: Barry Slaff

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.

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

NOTE:

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

Time and Location:

Tuesdays 13:30-15:00 252 BRB

Thursdays 13:30-15:00 252 BRB

Office Hours:

Yoseph - Monday 4-5:30pm, Richards D205

Barry - TBD, Richards C209

Setting Computer Cluster Access: PMACS HPC for GCB 537 (https://hpcwiki.genomics.upenn.edu/index.php/GCB_537:Main_Page)

Topics:

A list of tentative topics is given below. It does not necessarily represent chronological order and each topics can span multiple classes. Paper discussions and exercise reviews are not listed.

1. Intro: Course goals & mindset, grading, programming exerices. Review good/bad summary examples. Intro to GCB cluster. Python tools/IDE, Python scientific libraries. Review relevant basics from Prob, Statistics and ML.
2. Motif finding.
3. Optimization methods: Gradient ascent, EM.
4. RNA-Seq: its usage, and related computational challenges.
5. Clustering, unsupervised learning.
6. Regression & Classification: Linear models, GLMs, Naive Bayes, Logistic Regression, Sparseness control in regression and classification.
7. ChIP-Seq: its usage, and related computational challenges.
8. HMM, Viterbi.
9. Intro to Human Genetics Research and related computational challenges: GWAS, Phasing/Imputation.
10. eQTL mapping and related challenges.
11. MCMC.
12. Dimensionality reduction.
13. Information theory (only the relevant bits ;).
14. Decision Trees.
15. Cancer Genomics.
16. SVMs, mixture models.
17. Systems Biology - current topics and challenges.
18. 3D Genomes - Technology and challenges.

Course format:

The course is comprised of lectures, assignments, paper discussions, and a final project presentation. Lectures cover 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 and students will need to submit their answers before the class starts. Student teams will volunteer or be picked at random to answer these questions. Students will be graded for the answers they submitted and that will count as part of the homework assignments (see below). 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. See separate announcements regarding regarding coding environments.

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

Note: All pen & paper assignments are to be submitted as PDF using Latex

Grading of each programming exercise is composed of the following:

40% 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.

The complete code of academic integrity by the University can be found **here (http://www.upenn.edu/academicintegrity/ai_codeofacademicintegrity.html)**

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.

Course Summary:

Date	Details	
Thu Jan 17, 2019	**Publish: Ex0 Prog (https://canvas.upenn.edu/calendar?event_id=2405429&include_contexts=course_1431043)**	12am
	Intro+Review1 (https://canvas.upenn.edu/calendar?event_id=2405428&include_contexts=course_1431043)	1:30pm to 3pm
Tue Jan 22, 2019	**Intro+Review4: Topics in Stats/ML for CompBio +Motif Finding**	12am

Date	Details	
	(https://canvas.upenn.edu/calendar?event_id=2447218&include_contexts=course_1431043)	
	Publish: Ex1 PnP (https://canvas.upenn.edu/calendar?event_id=2405398&include_contexts=course_1431043)	12am
	Intro+Review2 (https://canvas.upenn.edu/calendar?event_id=2405427&include_contexts=course_1431043)	1:30am to 3am
Thu Jan 24, 2019	Intro+Review3: Topics in Stats/ML for CompBio +Motif Finding (https://canvas.upenn.edu/calendar?event_id=2405426&include_contexts=course_1431043)	1:30pm to 3pm
	How to work with HPC/remotely (https://canvas.upenn.edu/calendar?event_id=2405425&include_contexts=course_1431043)	4pm to 5pm
Fri Jan 25, 2019	Preference for paper presentation topics (https://canvas.upenn.edu/courses/1431043/assignments/6643211)	due by 11:59pm
Thu Jan 31, 2019	Publish Ex1prog (https://canvas.upenn.edu/calendar?event_id=2405397&include_contexts=course_1431043)	12am
	ML Topics: EM (https://canvas.upenn.edu/calendar?event_id=2405424&include_contexts=course_1431043)	1:30pm to 3pm
	Ex0 (https://canvas.upenn.edu/courses/1431043/assignments/6643174)	due by 11:59pm
Tue Feb 5, 2019	Paper Presentation: Motif Finding (MEME) - Yoseph (https://canvas.upenn.edu/calendar?event_id=2405423&include_contexts=course_1431043)	1:30pm to 3pm
	Questions: MEME Paper (https://canvas.upenn.edu/courses/1431043/assignments/6643217)	due by 1:30pm
Thu Feb 7, 2019	Review: RNA-Seq (https://canvas.upenn.edu/calendar?event_id=2405422&include_contexts=course_1431043)	1:30pm to 3pm
Tue Feb 12, 2019	ML Topics 2 (https://canvas.upenn.edu/calendar?event_id=2405421&include_contexts=course_1431043)	1:30pm to 3pm
	Review Ex0 prog (https://canvas.upenn.edu/calendar?event_id=2405396&include_contexts=course_1431043)	1:30pm to 3pm
	Ex1PnP (https://canvas.upenn.edu/courses/1431043/assignments/6643176)	due by 1:30pm
Thu Feb 14, 2019	Paper Presentation: RNA-Seq (DESeq) (https://canvas.upenn.edu/calendar?event_id=2405420&include_contexts=course_1431043)	1:30pm to 3pm

Date	Details	
	DESeq Paper Questions (https://canvas.upenn.edu/courses/1431043/assignments/6643172)	due by 1:30pm
Fri Feb 15, 2019	Feedback for DESeq Paper Presentation (https://canvas.upenn.edu/courses/1431043/assignments/6643193)	due by 11:59pm
Tue Feb 19, 2019	Paper Presentation: RNA-Seq (EXPress) (https://canvas.upenn.edu/calendar?event_id=2405394&include_contexts=course_1431043)	1:30pm to 3pm
	Review Ex1 PnP (https://canvas.upenn.edu/calendar?event_id=2405419&include_contexts=course_1431043)	1:30pm to 3pm
	ExPress Paper Questions (https://canvas.upenn.edu/courses/1431043/assignments/6643191)	due by 1:30pm
Thu Feb 21, 2019	Publish: Ex2 PnP (https://canvas.upenn.edu/calendar?event_id=2405418&include_contexts=course_1431043)	12am
	ML Topics 2 Continue + Clustering (https://canvas.upenn.edu/calendar?event_id=2405393&include_contexts=course_1431043)	1:30am to 3pm
	Feedback for eXpress Paper Presentation (https://canvas.upenn.edu/courses/1431043/assignments/6643195)	due by 11:59pm
Tue Feb 26, 2019	Review: ChIP-Seq (Golnaz Vahedi) (https://canvas.upenn.edu/calendar?event_id=2405417&include_contexts=course_1431043)	1:30pm to 3pm
Thu Feb 28, 2019	Paper Presentation: Peak Calling (ZINBA) (https://canvas.upenn.edu/calendar?event_id=2405416&include_contexts=course_1431043)	1:30pm to 3pm
	Zinba Paper Questions (https://canvas.upenn.edu/courses/1431043/assignments/6643230)	due by 1:30pm
Fri Mar 1, 2019	Zinba Paper Presentation Feedback (https://canvas.upenn.edu/courses/1431043/assignments/6643228)	due by 11:59pm
Mon Mar 4, 2019	Ex2PnP (https://canvas.upenn.edu/courses/1431043/assignments/6643181)	due by 11:59pm
Tue Mar 5, 2019	NO CLASS - SPRING BREAK (https://canvas.upenn.edu/calendar?event_id=2405410&include_contexts=course_1431043)	12am
	Publish: Ex2 Prog (https://canvas.upenn.edu/calendar?event_id=2405415&include_contexts=course_1431043)	12am
Wed Mar 6, 2019	Ex1 Prog (https://canvas.upenn.edu/courses/1431043/assignments/6643178)	due by 11:59pm

Date	Details	
Thu Mar 7, 2019	NO CLASS - SPRING BREAK (https://canvas.upenn.edu/calendar?event_id=2405409&include_contexts=course_1431043)	12am
Tue Mar 12, 2019	Introduction to human genetics research (https://canvas.upenn.edu/calendar?event_id=2405414&include_contexts=course_1431043)	1:30pm to 3pm
Thu Mar 14, 2019	Publish: Final Project (https://canvas.upenn.edu/calendar?event_id=2405413&include_contexts=course_1431043)	12am
	Review: Ex2PnP (https://canvas.upenn.edu/calendar?event_id=2405412&include_contexts=course_1431043)	12am
	ML Topics 2.5: Clustering (https://canvas.upenn.edu/calendar?event_id=2405392&include_contexts=course_1431043)	1:30pm to 3pm
	Review Ex1 Prog Analysis (https://canvas.upenn.edu/calendar?event_id=2405411&include_contexts=course_1431043)	1:30pm to 3pm
Tue Mar 19, 2019	Review: eQTL mapping - Casey Brown (https://canvas.upenn.edu/calendar?event_id=2405385&include_contexts=course_1431043)	1:30pm to 3pm
Thu Mar 21, 2019	Paper Presentation: eQTL (https://canvas.upenn.edu/calendar?event_id=2405384&include_contexts=course_1431043)	1:30pm to 3pm
	Questions: eQTL Paper (https://canvas.upenn.edu/courses/1431043/assignments/6643213)	due by 1:30pm
Tue Mar 26, 2019	Review Ex1Prog Analysis (https://canvas.upenn.edu/calendar?event_id=2405382&include_contexts=course_1431043)	1:30am to 3pm
	ML Topics 3 - MCMC (https://canvas.upenn.edu/calendar?event_id=2405381&include_contexts=course_1431043)	1:30pm to 3pm
	Feedback for QTL Paper Presentation (https://canvas.upenn.edu/courses/1431043/assignments/6643199)	due by 1:30pm
Thu Mar 28, 2019	Paper Presentation - GWAS (https://canvas.upenn.edu/calendar?event_id=2405383&include_contexts=course_1431043)	1:30pm to 3pm
	Questions for PrediXcan (https://canvas.upenn.edu/courses/1431043/assignments/6643215)	due by 1:30pm
Tue Apr 2, 2019	Publish: Ex3Prog (https://canvas.upenn.edu/calendar?event_id=2405408&include_contexts=course_1431043)	12am
	ML Topics 3 (cont), 4 - Dimensionality reduction (https://canvas.upenn.edu/calendar?event_id=2405407&include_contexts=course_1431043)	1:30pm to 3pm

Date	Details	
Thu Apr 4, 2019	Publish: Ex3 PnP (https://canvas.upenn.edu/calendar?event_id=2405406&include_contexts=course_1431043)	12am
	ML Topics 4-5: Dimensionality Reduction (Cont.) (https://canvas.upenn.edu/calendar?event_id=2405405&include_contexts=course_1431043)	1:30pm to 3pm
Tue Apr 9, 2019	Paper Presentation: Phasing/Imputation (https://canvas.upenn.edu/calendar?event_id=2405391&include_contexts=course_1431043)	1:30pm to 3pm
	New ExExtra (https://canvas.upenn.edu/courses/1431043/assignments/6643205)	due by 1:30pm
	Phasing Paper Questions (https://canvas.upenn.edu/courses/1431043/assignments/6643209)	due by 1:30pm
Wed Apr 10, 2019	Submit: Ex2Prog (https://canvas.upenn.edu/calendar?event_id=2405404&include_contexts=course_1431043)	12am
	Ex2Prog (https://canvas.upenn.edu/courses/1431043/assignments/6643183)	due by 11:59pm
Thu Apr 11, 2019	Review: Ex2Prog (https://canvas.upenn.edu/calendar?event_id=2405403&include_contexts=course_1431043)	12am
	ML Topics 4 Dim Reduction (finish) + Infrmation Theory Basics (https://canvas.upenn.edu/calendar?event_id=2405390&include_contexts=course_1431043)	1:30pm to 3pm
	Phasing Paper Presentation Feedback (https://canvas.upenn.edu/courses/1431043/assignments/6643207)	due by 11:59pm
Tue Apr 16, 2019	Review: System Biology/Single Cell - Pablo Camera (https://canvas.upenn.edu/calendar?event_id=2405388&include_contexts=course_1431043)	1:30pm to 3pm
Thu Apr 18, 2019	Publish: Ex4 PnP (https://canvas.upenn.edu/calendar?event_id=2405402&include_contexts=course_1431043)	12am
	Submit Ex3PnP (https://canvas.upenn.edu/calendar?event_id=2405401&include_contexts=course_1431043)	1:30am to 3am
	Paper Presentation - Single Cell (https://canvas.upenn.edu/calendar?event_id=2405387&include_contexts=course_1431043)	1:30pm to 3pm
	Single Cell Methods paper questions (ZIFA) (https://canvas.upenn.edu/courses/1431043/assignments/6643223)	due by 1:30pm
	Ex3PnP (https://canvas.upenn.edu/courses/1431043/assignments/6643185)	due by 11:59pm

Date	Details	
Tue Apr 23, 2019	Review: Genome Organization - Babak Faryabi (https://canvas.upenn.edu/calendar?event_id=2405389&include_contexts=course_1431043)	1:30pm to 3pm
	Single Cell Paper Presentation Feedback (https://canvas.upenn.edu/courses/1431043/assignments/6643225)	due by 1:30pm
Thu Apr 25, 2019	Machine Learning 5 - Kernel Methods, SVM (https://canvas.upenn.edu/calendar?event_id=2405400&include_contexts=course_1431043)	1:30pm to 3pm
	Feedback for PrediXcan Paper Presentation (https://canvas.upenn.edu/courses/1431043/assignments/6643197)	due by 1:30pm
Tue Apr 30, 2019	Paper Presentation: Genome Organization (https://canvas.upenn.edu/calendar?event_id=2405386&include_contexts=course_1431043)	1:30pm to 3pm
	Feedback on Genome Organization Paper Presentation (https://canvas.upenn.edu/courses/1431043/assignments/6643201)	due by 11:59pm
Wed May 1, 2019	Ex4PnP (https://canvas.upenn.edu/courses/1431043/assignments/6643189)	due by 11:59pm
	3D paper questions (HiCSeg) (https://canvas.upenn.edu/courses/1431043/assignments/6643170)	due by 11:59pm
Thu May 2, 2019	Review Ex3 prog, Ex4 PnP (https://canvas.upenn.edu/calendar?event_id=2405399&include_contexts=course_1431043)	1:30pm to 3pm
Fri May 3, 2019	Ex3Prog (https://canvas.upenn.edu/courses/1431043/assignments/6643187)	due by 11:59pm
Wed May 8, 2019	Final Project (https://canvas.upenn.edu/courses/1431043/assignments/6643203)	due by 2am
	Roll Call Attendance (https://canvas.upenn.edu/courses/1431043/assignments/6643220)	