Course Syllabus

GCB537 Advanced Computational Biology

Term: Spring 2018
Instructors: Yousef Barash, R. Babak Faryabi
Guest lecturers: Golnaz Vahedi, Li-San Wang, Casey Brown.
TA: Anupama Jha & Scott Norton

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, genomics 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 12:30-15:00 262 BRB
Thursdays 13:30-16:00 262 BRB

Office Hours:
Yousef - Monday 4-5:30pm, Richards D205
Anupama - Tuesday 4-5:30pm, Richards C269
Scott - TBD.

Setting Computer Cluster Access: [PMACS HPC for GCB 537](http://hpcwiki.genecommons.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 topic can span multiple classes. Paper discussions and exercise reviews are not listed.

- Intro: Course goals & mindset, grading, programming exercises. Review good/bad summary examples. Intro to GCB cluster, Python tools/IDE, Python scientific libraries. Review relevant basics from Prob, Statistics and ML.
- Motif finding.
- Optimization methods: Gradient ascent, EM.
- RNA-Seq: its usage, and related computational challenges.
- Clustering, unsupervised learning.
- Regression & Classification: Linear models, GLMs, Naive Bayes, Logistic Regression, Sparseness control in regression and classification.
- ChIP-Seq: its usage, and related computational challenges.
- HMM, Viterbi.
- Intro to Human Genetics Research and related computational challenges: GWAS, Phasing/Imputation.
- eQTL mapping and related challenges.
- MCMC.
- Dimensionality reduction.
- Information theory (only the relevant bits :).
- Decision Trees.
- Cancer Genomics.
- SVMs, mixture models.
- Systems Biology - current topics 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 led by instructors followed by discussions led by students. Students will team up to work towards a guided evaluation project of algorithm/metrics 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 discussions:
The course is divided into units covering current topics in comp bio and 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 slides 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) end work with the instructors to select a specific topic to evaluate tools on a particular biological/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_557: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 student 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

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 any 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/academicIntegrity.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:

<table>
<thead>
<tr>
<th>Date</th>
<th>Details</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Details</th>
</tr>
</thead>
</table>
| Thu Jan 11, 2018 | Publish: Ex2 Project  
[https://canvas.upenn.edu/calendar/event?event_id=2264753&include_contexts=course_1390731]  
12pm |
| Tue Jan 16, 2018 | Intro meeting + Review  
[https://canvas.upenn.edu/calendar/event?event_id=2244760&include_contexts=course_1390731]  
1:30pm to 3pm |
| Thu Jan 18, 2018 | Review: Topics in Prob. Stts, ML  
[https://canvas.upenn.edu/calendar/event?event_id=2244729&include_contexts=course_1390731]  
1:30pm to 3pm |
| Thu Jan 18, 2018 | How to work with HPC/remote  
[https://canvas.upenn.edu/calendar/event?event_id=2264754&include_contexts=course_1390731]  
4pm to 5pm |
| Tue Jan 23, 2018 | ML Topics 1  
[https://canvas.upenn.edu/calendar?event_id=2264738&include_contexts=course_1390731]  
1:30pm to 3pm |
| Thu Jan 25, 2018 | Paper Presentation: Moph Fmding (MEME) -  
Yoseph  
[https://canvas.upenn.edu/calendars/event?event_id=2264756&include_contexts=course_1390731]  
1:50pm to 3pm |
|             | Questions: MEME Paper  
[https://canvas.upenn.edu/courses/1390731/assignments/6099227]  
due by 11:59pm |
| Tue Feb 6, 2018 | Paper Presentation: RNA-Seq (DESSeq)  
[https://canvas.upenn.edu/calendar?event_id=2264728&include_contexts=course_1390731]  
1:30pm to 3pm |
| Thu Feb 8, 2018 | Paper Presentation: RNA-Seq (EXPress)  
[https://canvas.upenn.edu/calendar?event_id=2264755&include_contexts=course_1390731]  
1:30pm to 3pm |
|             | ExPress Paper Questions  
[https://canvas.upenn.edu/courses/1390731/ assignments/6099169]  
due by 11:59pm |
|             | Feedback for DESSeq Paper Presentation  
[https://canvas.upenn.edu/courses/1390731/ assignments/6099169]  
due by 11:59pm |
| Thu Feb 15, 2018 | ML Topics 2 Continue + Clustering  
[https://canvas.upenn.edu/calendar?event_id=2264733&include_contexts=course_1390731]  
1:30pm to 3pm |
|             | Review: Ex1 Project  
[https://canvas.upenn.edu/calendar?event_id=2264768&include_contexts=course_1390731]  
1:30pm to 3pm |
|             | Feedback for aXpress Paper Presentation  
[https://canvas.upenn.edu/courses/1390731/ assignments/6099168]  
due by 11:59pm |
| Tue Feb 20, 2018 | Publish: Ex2 Project  
[https://canvas.upenn.edu/calendar?event_id=2264750&include_contexts=course_1390731]  
12am |
|             | ML Topics 2.6: Clustering  
[https://canvas.upenn.edu/calendar?event_id=2264738&include_contexts=course_1390731]  
1:30pm to 3pm |
| Thu Feb 22, 2018 | Paper Presentation: Peak Calling (ZINBA)  
[https://canvas.upenn.edu/calendar?event_id=2264754&include_contexts=course_1390731]  
1:30pm to 3pm |
|             | ChiP Peak Calling Paper Questions  
[https://canvas.upenn.edu/courses/1390731/ assignments/6099218]  
due by 11:59pm |
| Mon Feb 26, 2018 | Ex2 Project  
[https://canvas.upenn.edu/courses/1390731/ assignments/6099169]  
due by 11:59pm |
| Tue Feb 27, 2018 | Publish: Ex2 Project  
[https://canvas.upenn.edu/calendar?event_id=2264737&include_contexts=course_1390731]  
12am |
|             | Introduction to human genetics research  
[https://canvas.upenn.edu/calendar?event_id=2264750&include_contexts=course_1390731]  
1:30pm to 3pm |
| Wed Feb 28, 2018 | Ex1 Project  
[https://canvas.upenn.edu/courses/1390731/ assignments/6099163]  
due by 11:59pm |
| Thu Mar 1, 2018 | Publish: Final Project  
[https://canvas.upenn.edu/calendar?event_id=2264736&include_contexts=course_1390731]  
12am |
|             | Review: Ex2 Project  
[https://canvas.upenn.edu/calendar?event_id=2264741&include_contexts=course_1390731]  
12am |
|             | Review: Ex1 Project Analysis  
[https://canvas.upenn.edu/calendar?event_id=2264757&include_contexts=course_1390731]  
1:30pm to 3pm |
|             | Zinba Paper Presentation Feedback  
[https://canvas.upenn.edu/courses/1390731/ assignments/6099214]  
due by 1:30pm |
| Tue Mar 6, 2018 | NO CLASS - SPRING BREAK  
[https://canvas.upenn.edu/calendar?event_id=2264749&include_contexts=course_1390731]  
12am |
| Thu Mar 8, 2018 | NO CLASS - SPRING BREAK  
[https://canvas.upenn.edu/calendar?event_id=2264746&include_contexts=course_1390731]  
12am |
| Tue Mar 13, 2018 | Snow Day - Class Cancelled Paper Presentation: GWAS (SKAT)  
[https://canvas.upenn.edu/calendar?event_id=2264746&include_contexts=course_1390731]  
1:30pm to 3pm |
| Thu Mar 15, 2018 | Review: eQTL mapping - Casey Brown  
[https://canvas.upenn.edu/calendar?event_id=2264748&include_contexts=course_1390731]  
1:30pm to 3pm |
| Tue Mar 20, 2018 | Paper Presentation: eQTL  
[https://canvas.upenn.edu/calendar?event_id=2264784&include_contexts=course_1390731]  
1:30pm to 3pm |
<table>
<thead>
<tr>
<th>Date</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thu Mar 22, 2018</td>
<td>Review Ex1 Proposal Analytic (<a href="https://canvas.upenn.edu/calendar?event_id=2284778&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284778&amp;include_contexts=course_1390731</a>) due by 1:30pm</td>
</tr>
<tr>
<td></td>
<td>ML Topics 3 - MCNC (<a href="https://canvas.upenn.edu/calendar?event_id=2284774&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284774&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td>Tue Mar 27, 2018</td>
<td>Publish: Ex3 Prop (<a href="https://canvas.upenn.edu/calendar?event_id=2284767&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284767&amp;include_contexts=course_1390731</a>) 12am</td>
</tr>
<tr>
<td></td>
<td>ML Topics 3 (cont.) - Dimensionality Reduction (<a href="https://canvas.upenn.edu/calendar?event_id=1390731&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=1390731&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td>Thu Mar 29, 2018</td>
<td>Publish: Ex3 Prop (<a href="https://canvas.upenn.edu/calendar?event_id=2284773&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284773&amp;include_contexts=course_1390731</a>) 12am</td>
</tr>
<tr>
<td></td>
<td>ML Topics 4 - Dimensionality Reduction (Cont.) (<a href="https://canvas.upenn.edu/calendar?event_id=2284776&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284776&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td>Tue Apr 3, 2018</td>
<td>Feedback for QTL Paper Proposal (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099160">https://canvas.upenn.edu/courses/1390731/assignments/6099160</a>) due by 1:30pm</td>
</tr>
<tr>
<td></td>
<td>Phaslon Paper Questions (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099167">https://canvas.upenn.edu/courses/1390731/assignments/6099167</a>) due by 1:30pm</td>
</tr>
<tr>
<td>Wed Apr 4, 2018</td>
<td>Submit: Ex2 Prop (<a href="https://canvas.upenn.edu/calendar?event_id=2284747&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284747&amp;include_contexts=course_1390731</a>) 12am</td>
</tr>
<tr>
<td></td>
<td>Ex2 Prop (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099167">https://canvas.upenn.edu/courses/1390731/assignments/6099167</a>) due by 11:50pm</td>
</tr>
<tr>
<td>Thu Apr 5, 2018</td>
<td>Review: Ex2 Prop (<a href="https://canvas.upenn.edu/calendar?event_id=2284748&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284748&amp;include_contexts=course_1390731</a>) 12am</td>
</tr>
<tr>
<td></td>
<td>ML Topics 4 Dim Reduction (final) + Information Theory Basics (<a href="https://canvas.upenn.edu/calendar?event_id=2284774&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284774&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td>Fri Apr 6, 2018</td>
<td>Phaslon Paper Presentation Feedback (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099169">https://canvas.upenn.edu/courses/1390731/assignments/6099169</a>) due by 11:50pm</td>
</tr>
<tr>
<td>Tue Apr 10, 2018</td>
<td>Review: Cancer Genomics - Babak Farzam (<a href="https://canvas.upenn.edu/calendar?event_id=2284775&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284775&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td></td>
<td>Publish: Ex4 Prop (<a href="https://canvas.upenn.edu/calendar?event_id=2284752&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284752&amp;include_contexts=course_1390731</a>) 12am</td>
</tr>
<tr>
<td></td>
<td>Submit: Ex3 Prop (<a href="https://canvas.upenn.edu/calendar?event_id=2284772&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284772&amp;include_contexts=course_1390731</a>) 1:30pm to 3am</td>
</tr>
<tr>
<td>Thu Apr 12, 2018</td>
<td>Paper Presentation: Cancer Genomics 1 (<a href="https://canvas.upenn.edu/calendar?event_id=2284771&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284771&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td></td>
<td>Ex3 Prop (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099188">https://canvas.upenn.edu/courses/1390731/assignments/6099188</a>) due by 1:30pm</td>
</tr>
<tr>
<td></td>
<td>Questions for PennCNV Paper (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099200">https://canvas.upenn.edu/courses/1390731/assignments/6099200</a>) due by 3pm</td>
</tr>
<tr>
<td>Tue Apr 17, 2018</td>
<td>Review: System Biology - Gregory Schwartz (<a href="https://canvas.upenn.edu/calendar?event_id=2284744&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284744&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td></td>
<td>Feedback for PennCNV Paper Presentation (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099187">https://canvas.upenn.edu/courses/1390731/assignments/6099187</a>) due by 1:30pm</td>
</tr>
<tr>
<td>Thu Apr 19, 2018</td>
<td>Machine Learning 5 - Kernel Methods, SVM (<a href="https://canvas.upenn.edu/calendar?event_id=2284743&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284743&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td>Tue Apr 24, 2018</td>
<td>Paper Presentation - System Biology (<a href="https://canvas.upenn.edu/calendar?event_id=2284754&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284754&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td>Wed Apr 25, 2018</td>
<td>Ex3 Prop (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099170">https://canvas.upenn.edu/courses/1390731/assignments/6099170</a>) due by 11:50pm</td>
</tr>
<tr>
<td></td>
<td>Ex4 Prop (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099174">https://canvas.upenn.edu/courses/1390731/assignments/6099174</a>) due by 11:50pm</td>
</tr>
<tr>
<td>Thu Apr 26, 2018</td>
<td>Review: Ex3 prop, Ex4 Prop (<a href="https://canvas.upenn.edu/calendar?event_id=2284768&amp;include_contexts=course_1390731">https://canvas.upenn.edu/calendar?event_id=2284768&amp;include_contexts=course_1390731</a>) 1:30pm to 3pm</td>
</tr>
<tr>
<td>Fri Apr 27, 2018</td>
<td>Systems Biology paper review (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099213">https://canvas.upenn.edu/courses/1390731/assignments/6099213</a>) due by 1:30pm</td>
</tr>
<tr>
<td>Wed May 2, 2018</td>
<td>Final Project (<a href="https://canvas.upenn.edu/courses/1390731/assignments/6099193">https://canvas.upenn.edu/courses/1390731/assignments/6099193</a>) due by 9am</td>
</tr>
<tr>
<td></td>
<td>System Biology Paper (ARACNE) Feedback (<a href="https://canvas.upenn.edu/courses/120731/assignments/6099216">https://canvas.upenn.edu/courses/120731/assignments/6099216</a>) due by 11:50pm</td>
</tr>
</tbody>
</table>