CMSC858B: Computational Systems Biology and Functional Genomics

Hector Corrada Bravo

Spring 2012

Description

Major advances in technology for genomic studies are bringing the prospect of personalized and individualized medicine closer to reality. Many of these advances are predicated on the ability to generate data at an unprecedented rate, posing a significant need for computational data analysis that is clinically and biologically useful and robust.

This course will concentrate on the fundamental computational and statistical methods required to meet this need. It will cover topics in functional genomics, population genetics and epigenetics. Computational methods studied for this type of analysis include: supervised, unsupervised and semi-supervised learning, data visualization, statistical modeling and inference, probabilistic graphical models, sparse methods, and numerical optimization. Machine learning methods will be a core component of this class. No prior knowledge of biology is required.

Course Information

- **Instructor:** Hector Corrada Bravo
  Center for Bioinformatics and Computational Biology
  Department of Computer Science
  hcorrada@umiacs.umd.edu
  Office: 3114F Biomolecular Sciences Building
  Phone Number: 301–405–2481

- **Lecture Meeting times**
  Monday and Wednesday, 3:30pm–4:45pm Room BSB 3118

- **Office Hours:** Thursday 11:00am–12:00pm BSB 3114F and by appointment

- **Evaluation**
  - Assignments: Three short assignments including programming, data analysis and pen-and-paper problems (30%)
  - Project and its presentation (20%)
  - Exams: Midterm (20%) and Final (25%)
  - Class participation (5%)

- **Textbook:** There is no required textbook.
Tentative Course Calendar

- 1/25: Course introduction and administrivia
- 1/31: Molecular biology for computer scientists and statisticians
- 2/1: Statistical Learning: a whirlwind tour of statistical inference, machine learning and probabilistic models
- 2/6: The R statistical programming environment
- 2/8: The R/Bioconductor genomics analysis environment
- 2/13: Gene expression analysis: Overview of microarray technology, preprocessing methods and algorithms
- 2/15: Differential expression analysis (1)
- 2/20: Empirical Bayes, SAM and Multiple testing
- 2/22: Clustering methods HW 1 due
- 2/27: Classification and prediction methods
- 2/29: Sparse methods in genomics
- 3/5: Overview of second generation sequencing technology
- 3/7: RNA sequencing analysis
- 3/12: De-novo transcript assembly and alternative splicing analysis, HW 2 due
- 3/14: Geneset Enrichment Analysis
- 3/19: No class: Spring Break
- 3/21: No class: Spring Break
- 3/26: Gene regulation Regulation of gene expression, Project proposal due
- 3/28: Transcription factor analysis (ChIPSeq)
- 4/2: Midterm exam
- 4/4: Regulatory network discovery and analysis
- 4/9: Genetics: Genomic variant discovery with sequencing technology (1)
- 4/11: Structural variant analysis with sequencing technology (2)
- 4/16: Statistical methods and algorithms for discovering genetic disease associations
- 4/18: Epigenetics: Intro to epigenetics: Chromatin modifications, DNA methylation and the CpG genomic landscape
- 4/23: Analysis of differential methylation with microarrays and sequencing
- 4/25: Chromatin modifications and conformation
- 4/30: Data integration: Methods and algorithms for genomic data integration, HW 3 due
- 5/2: Approaching the promise of individualized medicine
- 5/7: Project presentations
- 5/9: Wrap-up, Final project writeup due
- 5/13: Final Exam

The final (mutable) calendar is in the course homepage.
Preliminary Project Guidelines

A substantial amount of the evaluation in this course will consist of a project. In general you can choose from four types of projects:

1. Algorithm/method design: Design and provide a preliminary implementation of an algorithm or method to analyze a particular type of genomic assay
2. Algorithm/method application: Apply existing algorithms or methods to existing genomic datasets
3. Literature review: Review the existing methods for a particular analysis task
4. Other: Any other approved project analyzing high-throughput genomic assays

The project can be done in pairs, but please let the instructor know if you plan to do this as soon as possible.

Policies

- There will be reading assignments. Students are expected to have read the material before class.
- Students are expected to attend lectures. Active participation is expected.
- Assignments are to be handed-in electronically or in class as instructed on their due date. Late assignments will be penalized.
- Students may discuss homeworks in groups. However, each student must write solutions independently.
- Programming exercises must be completed using R. We will spend time in class going over enough R that you can complete your assignments.
- Cell phone usage is prohibited during lecture, laptop use will be allowed to the extent that students use it to follow along an in-class analysis or demonstration.
- Using or referencing any materials from the web without proper citation is a violation to the honor code.
- In this course you are responsible for both the University’s Code of Academic Integrity and the University of Maryland Guidelines for Acceptable Use of Computing Resources. Any evidence of unacceptable use of computer accounts or unauthorized cooperation on tests, quizzes, or projects will be submitted to the Student Honor Council, which could result in an XF for the course, suspension, or expulsion from the University.
- Students claiming an excused absence for an exam must apply in writing and furnish documentary support (such as from a health care professional who treated the student). No make-up exams are given.
- Any student eligible for and requesting reasonable academic accommodations due to a disability is requested to provide, to the instructor in office hours, a letter of accommodation from the Office of Disability Support Services (DSS) within the first two weeks of the semester.
- Any student who must miss a class due to religious holidays should also notify the instructor during the first two weeks of class.