CMSE 491/890 - Bioinformatics & Computational Biology

Arjun Krishnan

arjun@msu.edu | @compbiologist | thekrishnanlab.org

Lecture 1: Introduction and Overview

Course overview

- Introductions
- Course website
- Communication
- Course activities
- Tentative Schedule
- Course topics
- Wrap-up

Intro

- arjun@msu.edu | @compbiologist | thekrishnanlab.org
- Assistant Professor
  - Dept. Computational Mathematics, Science, and Engineering
  - Dept. Biochemistry and Molecular Biology
- Research Interests: Computational genomics, Biomedical data science, Biological networks, Data integration, Machine learning

What you should get out of this course

How to become a practicing computational biologist in 60+h?

- An understanding of the major research directions in computational biology.
- Familiarity with the algorithms and statistical techniques for addressing these problems.
- At the end you should be able to:
    - Critically read bioinformatics / computational-biology literature.
    - Apply the methods you have learned to other problems both within and outside of bioinformatics.

Modules & Topics

Module	Topics
Genome assembly, alignment, & annotation	de Bruijin graphs; Suffix trees; Hidden Markov models
Sequence alignment & pattern finding	Dynamic programming; Substitution matrices; BLAST
Comparative genomics; Phylogenomics	Molecular evolution; Tree construction
Genetic variation & quantitative genetics	LD, GWAS, Regularized linear regression; Statistical inference, Multiple testing
Regulatory genomics	Gibbs sampling; Expectation-Maximization
Functional genomics	Differential expression; Functional enrichment analysis; Clustering; Intro to ML
DataSci Primers	Data wrangling & visualization in R/Python; Exploratory data analysis
ML Primers	Machine learning; Deep learning; Applications
Single-cell genomics	Missing value imputaion; Dimensionality reduction; Trajectory inference; Spatial reconstruction
Molecular dynamics; Protein structure prediction	Molecular simulation; Maximum entropy modeling
Modeling cellular pathways; Digital evolution	Dynamical simulation, State Space, Bifurcation; Linear programming; Artificial life
Biological networks	Measuring associations; Network inference; Graph theory, Label propagation
Cancer genomics	Overview
Genome engineering	Overview
Personal genomics	Overview

Modules & Topics

Module	Topics
Genome assembly, alignment, & annotation	de Bruijin graphs; Suffix trees; Hidden Markov models
Sequence alignment & pattern finding	Dynamic programming; Substitution matrices; BLAST
Comparative genomics; Phylogenomics	Molecular evolution; Tree construction
Genetic variation & quantitative genetics	LD, GWAS, Regularized linear regressio
Regulatory genomics	Gibbs sampling; Expectation-Maximizati
Functional genomics	Differential expression; Functional enrich
DataSci Primers	Data wrangling & visualization in R/Pyth
ML Primers	Machine learning; Deep learning; Applic
Single-cell genomics	Missing value imputaion; Dimensionality
Molecular dynamics; Protein structure prediction	Molecular simulation; Maximum entropy
Modeling cellular pathways; Digital evolution	Dynamical simulation, State Space, Bifu
Biological networks	Measuring associations; Network inferer
Cancer genomics	Overview
Genome engineering	Overview
Personal genomics	Overview

Arjun presents

Each topic:

- Lecture
- Paper discussion

You present as team of two.

Prerequisites & Expectations

- CMSE 201 and two semesters of introductory biology (LB 144 and 145 OR BS 161 and 162 OR BS 181H and 182H, or equivalent).
  - Statistics at the level of STT 231 is strongly recommended.

- Basically, it would be assumed that you:
  - know how to code in one of the mainstream languages like Python or R,
  - have an understanding of basic statistics and probability, and
  - have studied basic genetics, molecular biology, and cellular biology.

Course website

github.com/krishnanlab/teaching/tree/master/2019-spring_compbio

- Contact information
- Course outline and materials
- Schedule, location, calendar, and office hours
- Website and communication
- Course activities
- Grading information
- Attendance, conduct, honesty, and accommodations

- Lecture slides
- Learning materials
- Assignments
- Scribe notes

Communication

cmse-compbio-ss19.slack.com

- The primary mode of communication in this course (including major announcements) will be the course Slack account.
- All of you should have invitations to join this account in your MSU email.

#syllabus-schedule	#assignments
#announcements	#semester-projects
#lectures	#random

Tell us about you

- Name
- Major or graduate program
- Research interests and/or topics you're especially interested in learning about
- Background in programming, algorithms, statistics, biology
- Office hours

Survey: bit.ly/ss19-compbio_incoming-survey

My office: 2507H Engineering Building (2nd floor)

Course activities

- Assignments: ~35%
- Class participation: ~15%
- Scribing: ~10%
- Project: ~40%

Assignments

- For each topic, you will be assigned a paper after the topic's "Lecture" class that you are required to read, summarize, and critique.
- Submit a report ***before*** the topic's "Paper discussion" class.
- See class website for details on what this report needs to contain.

Class participation

- Do the assignments and additional readings.
- Show up to class.
- Work in groups during in-class discussion sessions (Bring your laptops)
- No one will have the perfect background.
    - Ask questions about computational or biological concepts.
- Correct me when I am wrong.

- Paper presentation (see the class website).

Scribing

- Each lecture will have one dedicated scribe who will take notes on the lecture.
- Complete draft of scribe notes due in 3 days after lecture.
    - I will read and give comments/suggestions.
- Final scribe notes due in 6 days after lecture.
- We will use stackedit.io + Google Drive, and then publish on GitHub.
    - https://writing.stackexchange.com/a/32393

Project

- A major goal of this course is to prepare your ability to perform original research in computational biology, and to present your ideas and research.
- Can be one of:
  - Design and implement a new computational method for a task in biology
  - Improve an existing method
  - Perform an evaluation of several existing methods
  - Develop a fully-reproducible codebase for an existing analysis

Project

Item	Due date
Project profile	W Jan 16
Project topic	F Feb 01
Project pre-proposal	W Fri 08
Project proposal	W Feb 20
Proposal reviews	F Mar 01
Mid-term project proposal presentations	M Mar 11 W Mar 13 F Mar 15
Review response	Th Mar 14 Sa Mar 16 M Mar 18
Mid-course project report	F Mar 29
Final project report	F Apr 26
Final project presentations	M Apr 29

Course reading, External resources, Relevant seminars

- Mostly papers: primary articles and reviews
- Seminar
  - Science at the Edge https://web.pa.msu.edu/seminars/edge/
  - ML Seminar (CSE)
  - CSME Colloquium

Groups @ MSU doing (a sigf. amount of) Bioinfo & Compbio

- Jianrong Wang
- Yuying Xie
- Leslie Kuhn
- Alex Dickson
- Michael Feig
- George Mias
- Ana Vasquez
- Gustavo de los Campos
- Eran Andrechek
- Robin Buell
- Erik Goodman
- Kevin Liu
- Mark Reimers
- Sudin Bhattacharya
- Jiayu Zhou
- Shinhan Shiu
- MANY OTHERS

You are welcome to work with me as well!

What you need to do before the next class

- Read the course website: github.com/krishnanlab/teaching/tree/master/2019-spring_compbio
- Fill out the incoming survey: bit.ly/ss19-compbio_incoming-survey
- Sign-up for scribing and paper presentation: check Slack for link

Lecture 2: Introduction and Overview

Getting started

- Some history & high-level view of the field
- Choosing a good problem
- Organizing a computational biology project
- Reading journal articles | Supplementary materials
- Programming languages and other logistics
- Managing data and code
- Getting help