CMSE 491 - Bioinformatics & Computational Biology

Arjun Krishnan
arjun@msu.edu
@compbiologist

Lecture 1: Introduction and Overview

Course overview

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

Course website

bit.ly/cmse491-spring18

- 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
- Pre-class assignments
- Scribe notes

Communication

https://cmse491bioinfocompbio.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	#pre-class assignments
#announcements	#semester-projects
#slides-materials	#random

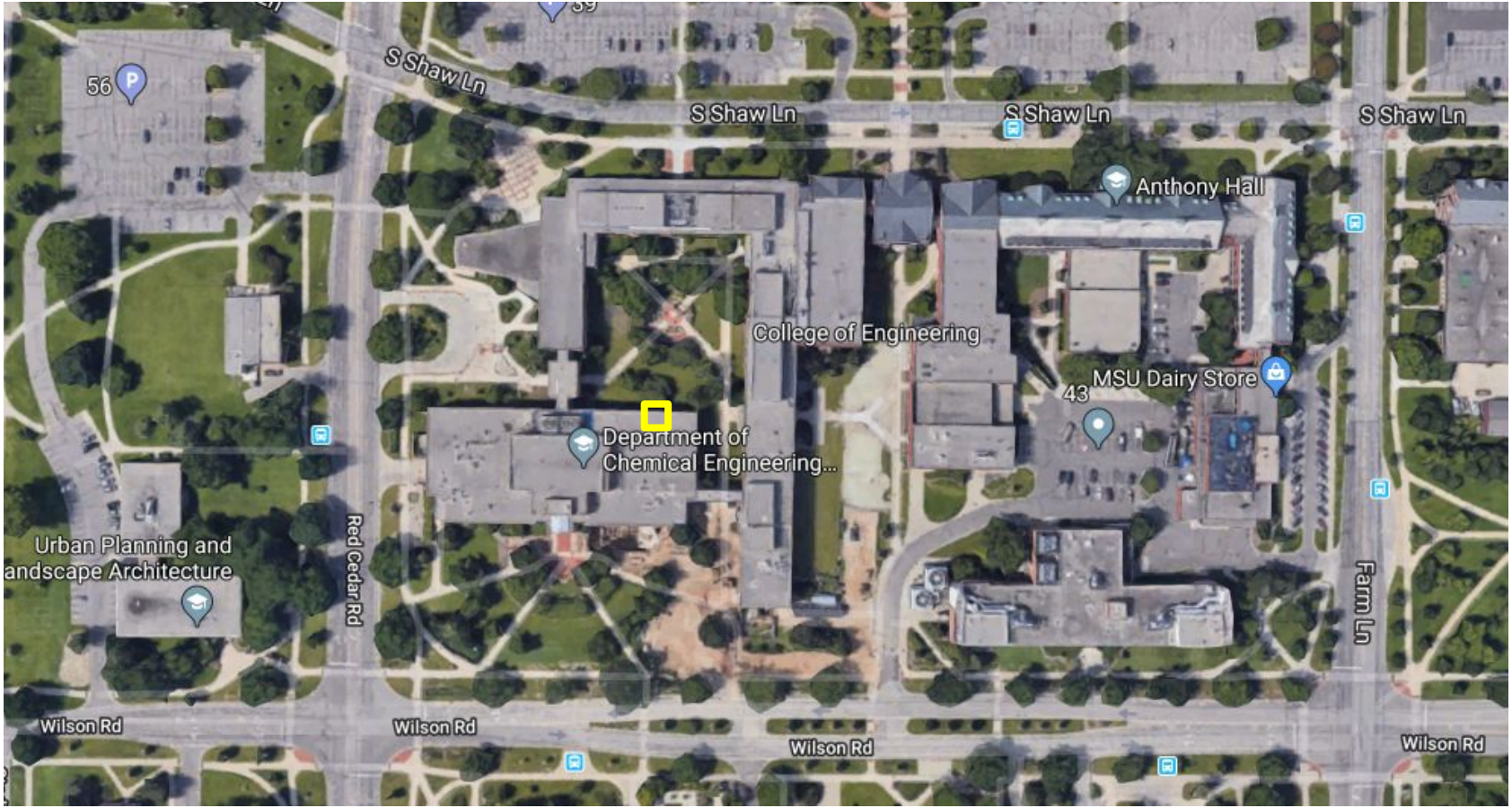
Instructor

Arjun Krishnan

- Email: arjun@msu.edu
- Office: 2507H Engineering Building
- 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, …

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



Tell us about you

- Name
- Major or graduate program
- Research interests and/or topics you're especially interested in learning about
- Favorite programming language

Survey: bit.ly/cmse491-spring18-survey

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 and R,
  - have an understanding of basic statistics and probability, and
  - have studied basic genetics, molecular biology, and cellular biology.

What you should get out of this course

How to become a practicing computational biologist in 40+h?

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

Major topics to be covered – Biological areas

- Genome assembly and annotation
- Sequence alignment and pattern finding
- Comparative genomics
- Genetic variation and quantitative genetics
- Regulatory genomics
- Functional genomics and data integration
- Molecular and digital evolution
- Molecular docking and molecular dynamics simulations
- Protein residue coupling and structure prediction
- Modeling cellular pathways
- Metabolomics and metabolic flux analysis
- Large-scale biological networks

Arjun presents

Each topic:

- Lecture
- Paper discussion

You present as team of two.

Major topics to be covered – Algorithms, techniques, approaches

- Dynamic programming
- Expectation Maximization, Gibbs sampling
- Distance measures, Mutual information
- Statistical modeling and inference, Multiple hypothesis correction
- Exploratory data analysis, Clustering, Dimensionality reduction
- Supervised machine learning, Deep neural networks
- Linear programming
- Hidden Markov Models
- Suffix trees
- Bayesian inference
- Stochastic context-free grammars
- Graph theory, Label propagation

Course activities

- Pre-class assignments: ~35%
- Class participation: ~15%
- Scribing: ~10%
- Project: ~40%

Pre-class 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 pre-class assignments and additional readings.
- Show up to class.
- Work in groups during in-class discussion sessions.
- 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 two dedicated scribes who will take notes on the lecture.
- Individually complete draft of scribe notes due in 3 days after lecture.
    - I will read and give comments/suggestions.
- The two scribes will then work together to combine their drafts+comments into a single final scribe notes due in 6 days after lecture.

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	Mon, Jan 29
Project topic/team	Wed, Feb 07
Project pre-proposal	Wed, Feb 14
Project proposal	Mon, Feb 26
Proposal reviews	Mon, Mar 05
Mid-term project proposal presentations	Mon, Mar 12
Review response	Mon, Mar 14
Mid-course project report	Wed, Apr 04
Final project report	Wed, Apr 25
Final project presentations 1	Wed, Apr 25
Final project presentations 2	Mon, Apr 30

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
- Yanni Sun
- 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: bit.ly/cmse491-spring18
- Fill out the class survey: bit.ly/cmse491-spring18-survey
- Sign-up for scribing and paper presentation: check Slack for link

Lecture 2: Introduction and Overview

Getting started

- 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