CS 618

This page contains the course syllabus and plan for CS 618 Computational Biology (run also as CS 459 Topics in CS for undergrads) for the summer of 2024. The basic plan is to look at a number of projects that I have worked on in the past, and to look at key tools and algorithms used in computational biology and bioinformatics.

The top 3 goals for the course - (1) being able to use programming, tools, etc. to work on biology-related projects and data, (2) understanding some of the key algorithms, statistics, etc. used in this area, and (3) understanding as much of the biology as we can, in particular related to where the data comes from, what it means, etc. And tying all of these together will be working on some projects.

Contents

General Information

Course website' - https://cs.indstate.edu/wiki/index.php/CS_618

Your Instructor

Jeff Kinne (https://kinnejeff.com), jkinne@cs.indstate.edu (http://mailto:jkinne@cs.indstate.edu)
Office: Root Hall A-142 and in Microsoft Teams, phone 812-237-2126
Instructor Office Hours: MTWR 11am-noon
Meeting: https://cs.indstate.edu/jkinne-meeting

Lecture, Exam

Lecture: Course is being run asynchronously, so the regular lecture hour time is being used as office hours. If you want to meet you can schedule a meeting with the meeting link, or join the zoom meeting that is in the course (and if I am available I will join as well)
Mid-term exam: TBA
Final exam: TBA

Prerequisites - no prerequisite.

CRN numbers - 30693 for the 001 face to face section, 30695 for the 301 online section

Required text We will use all online sources. This section will be updated as we go through them.

- Bioinformatics

Class notes - Notes during class will mostly be kept in the documents in **this OneDrive folder (https://sycamoresin dstate-my.sharepoint.com/:f:/g/personal/jeffrey_kinne_indstate_edu/EiJJbQVND3FMr25IAoHydnUBA DrC7z-tZH_OTaD1h_A8Ow?e=y81xWu)**. Note that you will need to authenticate with your ISU account to view the notebook.

Programming/Tools

R programming, including commonly used packages.

Python programming, including commonly used packages.

Other programming - javascript/node, bash.

Software/tools - BLAST, NCBI.

Algorithms/Statistics

Statistics

Clustering techniques

Sequence alignment algorithms

Biology

Central dogma of genetics

Biological data - different types of assays, etc. - how the data is produced, what the data looks like, etc.

Projects

Gene expression - determining key genes from gene expression datasets. Project is in R, uses Shiny, Datatables, ShinyProxy, Docker. Poster - https://cs.indstate.edu/info/posters/bd4isu2022-bartlett.pdf

Protein topology prediction - finding potential transmembrane proteins in genomes. Project is in Python, uses Javascript, NCBI, BLAST. Poster - https://cs.indstate.edu/info/posters/bd4isu2022-hoffman.pdf

Transcription factors - finding mutations to disable a transcription factor while still preserving others. Project is in Python, R, and/or C. Poster - https://cs.indstate.edu/info/posters/bd4isu2020-bennett.pdf

Gene expression - determining key genes in a particular dataset from fish. Poster - https://cs.indstate.edu/info/posters/bd4isu2021-gosnell.pdf

Mass spectrometry data - keeping a database of mass spec data and searching through databases for new samples. Potential new project.

Genome sequencing - doing whole genome sequencing for species that have not yet had this done. New project with one of the students in the course.

Sequencing - different sequencing technologies (RNA seq, CHIP seq, single cell RNA seq, etc.), pros/cons/costs/what-used-for/etc., for a study looking cancer in a model organism.

Other requested topics...

Drug discovery/modeling - modeling/simulating drug interactions with the body and the drug discovery process.

Resources

Bioinformatics

Watch list

Watch starting from the bottom.

- single cell seq (https://www.youtube.com/watch?v=k9VFNLLQP8c)
- ATAC seq (https://www.youtube.com/watch?v=L2Kxaq9yRE4)
- high throughput sequencing (StatQuest) (https://www.youtube.com/watch?v=tlf6wYJrwKY&list=PLblh5JKOoLUJo2Q6xK4tZElbIvAACEykp)
- genetics (https://www.youtube.com/watch?v=9zwq8N4Ufd8&list=PL8dPuuaLjXtPW_ofbxdHNciuLoTRLPMgB&index=33) (and continue watching subsequent videos up through mutations

- cell (https://www.youtube.com/watch?v=jsDxw63QqK0&list=PL8dPuuaLjXtPW_ofbxdHNciuLoTRLPMgB&index=24)

Data files

- GSE85331 Liu et al (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85331) - see file
  GSE85331_all.gene.FPKM.output.replicates.txt.gz at the bottom.

Programs to install

- 7-zip (https://www.7-zip.org/) - for extracting zip files, if your OS cannot unzip them already (e.g., Windows 10).

Course Description and Content

Course Description

The catalog description for this course is: "An introduction to computational biology. Topics may include principles and methods used for sequence alignment, motif finding, structural modeling, structure prediction and network modeling, as well as currently emerging research areas. A focus is placed on the computational cost of solving problems in terms of CPU time, memory, and disk space. Study of the core algorithms used to solve problems."

Course Outline

The top 3 goals for the course - (1) being able to use programming, tools, etc. to work on biology-related projects and data, (2) understanding some of the key algorithms, statistics, etc. used in this area, and (3) understanding as much of the biology as we can, in particular related to where the data comes from, what it means, etc. And tying all of these together will be working on some projects.

Learning Outcomes

Biology - Can explain fundamental concepts in biology and apply these to problems/algorithms covered - experimental errors, experiment design, basic statistics, central dogma of genetics, sequencing technologies and data formats.

Programming - Basic understanding of programming languages commonly used in computational biology - what each is used for, ability to compile/run programs in each, able to make changes to programs given to you.

Algorithms - Can explain and use a variety of classification and clustering algorithms (strengths, weaknesses, uses) as well as key concepts in machine learning.

Software - Can run bioinformatics tools to process biological data, perform analysis, and interpret the results.

Statistics - Can apply statistics to data used in the course.

Course Policies, Grading

See Jeff Kinne Course Policies for course policies and how your overall letter grade will be determined.

Assignments

Start Assignments and Quiz Studying Early - I suggest attempting an assignment the day it is given, or the day after, so that if you have a problem you can ask early. If you continue to have problems in trying to complete the assignment, you will have time to ask again. Many of the assignments require thought and problem solving, which takes "time on the calendar" not just "time on the clock". By that I mean that spending an hour on 3 consecutive days is likely to be more productive than trying to spend 3 hours at once on the assignment.

Expected Amount of Work - My expectation is that an average student will spend about 5-10 hours OUTSIDE of class each week (that is in addition to class time or viewing lecture videos) WORKING PRODUCTIVELY/EFFICIENTLY (not just staring at the computer) to complete their coursework for this class. Some students may spend less time than this, and some students will spend more.

Note - please find a way to spend enough time on this class (the investment will pay off in terms of skills, being able to get a job, etc.).

Grade Meanings

Letter grades in this course are intended to have the following rough meaning.

- A+/A: You understand everything and probably could teach the course yourself.
- B+/A-: You understand nearly everything, and should be all set to use this knowledge in other courses or in a job.
- C/C+/B-/B: Some things you understand very well and others you don't (more towards the former for a B and more towards the latter for a C).
- D-/D+/C-: You did put some effort in, and understand many things at a high level, but you haven't mastered the details well enough to be able to use this knowledge in the future. **Note that the lowest grade for grad courses is a C, so if you fall in the range below C then your letter grade will be an F.**
- F: Normally, students that get an F simply stopped doing the required work at some point.

The default assignment of letter grades will initially be the usual one (90 is an A-, 80 is a B-, etc.), but these cutoffs will be adjusted down (in your favor). I will adjust these cutoffs and let you know where you stand periodically throughout the term.

CS-Specific Items

This section contains items that are generally the same for all CS courses (and in particular those taught by this instructor).

Lab Help

Note - No lab help during summer courses.

We have a few lab assistants who are available to help students in beginning computer science courses. Please see https://cs.indstate.edu/wiki/index.php/Unix_Lab_and_Help for details. The lab hours are in a calendar on the CS homepage, at http://cs.indstate.edu/info/index.php#lab_hours. You can join the lab when working on your programs.

You can ask the lab assistants to look at your programs, and you can work with any other CS students that are there (you could use the lab as a regular meeting place to work with your classmates).

Course Announcements

Announcements regarding the course will be made in Canvas, with Announcements for the course. You should make sure your settings are such that you will be notified of these announcements (e.g., by email). You should regularly check your ISU email account or have it forwarded to an account that you check regularly. You can set the account to forward by logging into your indstate.edu email online (if you aren't able to find the option, try a different browser or search online for things like - outlook online forward email setting).

Classroom conduct

Note - online course, this section does not apply. You may not use cell phones, iPods/music players, etc. during class. You should be civil and respectful to both the instructor and your classmates, and you should arrive to class a few minutes before the scheduled lecture so you are ready for lecture to begin on time. You may use your computer during class if you are using it to follow along with the examples that are being discussed. You should avoid spending time on email, Facebook, work on other courses, etc. during the lecture for this class (be fully present wherever you are, make the most of each experience).

Academic Integrity

See also Jeff Kinne Course Policies for additional information for more specifics about how I am handling these things for this course.

Please follow these guidelines to avoid problems with academic misconduct in this course:

Homework: You may discuss the homework assignments, but should solve and finish them on your own. To make sure you are not violating this, if you discuss with someone, you should DESTROY any work or evidence of the discussion, go your separate ways, SPEND at least an hour doing something completely unrelated to the assignment, and then you should be able to RECREATE the program/solution on your own, then turn that in. If you cannot recreate the solution on your own, then it is not your work, and you should not turn it in.

Note on sources: if you use some other source, the web or whatever, you better cite it! Not doing so is plagiarism.

Exams: This should be clear no cheating during exams. Each instructor has different rules for what is allowed on exams in terms of notes, etc. If not noted otherwise, you should assume that a quiz or exam is closed notes, no computer, no calculator.

Projects: You should not copy from the Internet or anywhere else. The project should be your own work. It will be fairly obvious to me if you do copy code from the Internet, and the consequences will be at the least a 0 on the project. If cheating is observed, you will at the least receive a 0 for the assignment (and may receive an F for the course), and I will file a Notification of Academic Integrity Violation Report with Student Judicial Programs, as required by the university's policy on Academic Integrity. A student who is caught cheating twice (whether in a single course or different courses) is likely to be brought before the All University Court hearing panel, which can impose sanctions up to and including suspension/expulsion. See http://www.indstate.edu/sjp/docs/code.pdf and http://www.indstate.edu/academicintegrity/ for more information.

Please ask the instructor if you have doubts about what is considered cheating in this course.

Office hours (using Teams)

You can contact me by email or Teams. If you want to meet in person you should reserve an appointment using http://cs.indstate.edu/jkinne-meeting to reserve an in person meeting with Jeff Kinne. During the summer I am not normally in my office unless I have a meeting on campus that day.

Canvas

The course has a canvas site. Click https://indstate.instructure.com/ to go to canvas. You should see this course listed under your courses for the current term. If you don't you may need to click on the Courses icon and then click the "All courses" link. The canvas site is used for giving you your grades, for quizzes/exams, for getting to online lectures (which are done using Zoom), and for posting announcements. Links and such will be kept on this website.

Lectures (using Zoom)

Lectures are recorded using zoom and accessible from the course's Canvas page - go to our course, then Modules, then Zoom, then find "Cloud Recordings".

For ISU's links to information on getting started with Zoom, see https://indstate.teamdynamix.com/TDClient/1851/Portal/KB/ArticleDet?ID=107534. You can also see the information linked at https://www.indstate.edu/services/student-success/cfss. You will get to the lectures for this course by going to Canvas, select this course, click Modules on the menu on the left, and click on the Zoom module. Once there you should see a schedule of lectures and be able to view recorded lectures. Note that you should install the Zoom application for your computer, and you will need to be logged into to Zoom with your ISU credentials to be able to connect. Also note that the lectures are recorded and only available to those in our class. Recorded lectures normally appear later the same day as the lecture.

Note that if you have not used Zoom with your ISU account previously, you need to go to https://indstate-edu.zoom.us and login with your ISU email address and password to get it setup.

Participating online

If you are participating online, please see the information at https://www.indstate.edu/services/student-success/cfss about participating in online courses. You are expected to watch the recordings once they are available.

All assignments and quizzes will be due around 1 week after assigned. Quizzes will normally be available for around 1 week.

ISU Required Syllabus Items

The items in this section are required and are the same for every ISU course.

Special Needs / Disability Services

Standard ISU language required in all syllabi...

Indiana State University recognizes that students with disabilities may have special needs that must be met to give them equal access to college programs and facilities. If you need course adaptations or accommodations because of a disability, please contact us as soon as possible in a confidential setting either after class or in my office. All conversations regarding your disability will be kept in strict confidence. Indiana State University's Student Support Services (SSS) office coordinates services for students with disabilities: documentation of a disability needs to be on file in that office before any accommodations can be provided. Student Support Services is located on the lower level of Normal Hall in the Center for Student Success and can be contacted at 812-237-2700, or you can visit the ISU website under A-Z, Disability Student Services and submit a Contact Form. Appointments to discuss accommodations with SSS staff members are encouraged.

Once a faculty member is notified by Student Support Services that a student is qualified to receive academic accommodations, a faculty member is obligated to provide or allow a reasonable classroom accommodation under ADA.

Non-Discrimination, Harassment, and Sexual Misconduct

Standard ISU language required in all syllabi...

Indiana State University is committed to inclusive excellence. To further this goal, the university does not tolerate discrimination in its programs or activities on the basis of: race, color, national origin, gender, age, sexual orientation, gender identity or expression, disability, veteran status, or any other protected class. Title IX of the Educational Amendments of 1972 in particular prohibits discrimination based on sex in any educational institution that receives federal funding. This includes sexual violence, sexual misconduct, sexual harassment, dating violence, domestic violence, and stalking. If you witness or experience any form of the above discrimination, you are asked to report the incident immediately to Public Safety: 812-237-5555 or to The Office of Equal Opportunity & Title IX: 812-237-8954. With respect to sexual discrimination, instructors, faculty, and some staff are required by law and institutional policy to report what you share with them to The Office of Equal Opportunity & Title IX. You do, however, have the option of sharing your information with the following confidential resources on campus:

- Student Counseling Center: 812-237-3939; Gillum Hall, 2nd Floor
- Victim Advocate: 812-237-3849 or 812-243-7272 (cell); HMSU 8th Floor

For more information about discrimination and the support resources available to you visit the Office of Equal Opportunity and Title IX website. Please direct any questions or concerns to: Title IX Coordinator; 812-237-8954; Rankin Hall 426; ISU-equalopportunity-titleix@indstate.edu.