CSE6301 Algorithms for Bioinformatics and Computational Biology — Course Syllabus (Fall 2025)

Instructor Information

Instructor	Email	Office Hours & Location
Xiuwei Zhang	xiuwei.zhang@gatech.edu	See Canvas

General Information

Time and Location
Mon, Wed: 9:30-10:45
Instructional Center 215

Description
This graduate-level course focuses on the intersection between algorithms and bioinformatics. We will cover algorithms used in classical bioinformatics problems, as well as some machine learning techniques. Students will learn the fundamental principles, underlying mathematics, and implementation details of these methods. Through reading and critiquing published research papers, students will learn the applications of algorithms to a variety of biological problems. In the final project, students will apply what they have learned to real-world data by exploring these concepts with a biological problem that they are passionate about.

This course is appropriate for graduate students or advanced undergraduate students in computer science, bioinformatics, biomedical engineering and mathematics. Familiarity with basic linear algebra, statistics, probability, and discrete math is expected. Students are also expected to have programming experience in Python.

Pre- &/or Co-Requisites
While there are no formal prerequisites for this course, it is intended as a graduate-level course, and as such there is recommended background necessary to keep up with the material covered. Programming skills (specifically Python) are necessary to complete the assignments. Students must have a strong mathematical background (linear algebra, calculus especially partial derivatives, and probabilities & statistics).

Learning Objectives

- Learn how to formulate computational biological questions as algorithmic or machine learning problems.
- Understand some fundamental algorithms and machine learning models and their underlying mathematical principles, and what types of problems each model is appropriate for.
- Gain experience in reading and presenting research papers in algorithms/machine learning for computational biology.

Course Materials

Course Text
While the slides are not developed based on any specific textbook, you may find the following books helpful.

Gusfield, Dan. Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology. Cambridge, UK: Cambridge University Press, 1997. ISBN: 0521585198.

Waterman, Michael. Introduction to Computational Biology: Maps, Sequences, and Genomes. Boca Raton, FL: CRC Press, 1995. ISBN: 0412993910.

Durbin, Richard, Graeme Mitchison, S. Eddy, A. Krogh, and G. Mitchison. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge, UK: Cambridge University Press, 1997. ISBN: 0521629713.

Additional Materials/Resources

All additional reading materials, if any, will be available on Canvas.

Course Website and Other Classroom Management Tools

Canvas will be used as the course website.

Tentative Topics

1. Genome assembly
   Next generation sequencing; Sequence assembly; Genomes are several billion letters long, and yet sequencing technologies only read a few hundred nucleotides at a time. Assembling the pieces into a complete genome is especially challenging, given the highly repetitive nature of the human genome, and the error-prone nature of sequencing reads.
2. Sequence alignment
   Pairwise alignment of biomolecular sequences. Global alignment, Needleman-Wunsch algorithm. Local alignment, Smith-Waterman algorithm. Derivation of scoring system for amino acid substitutions in sequence evolution. Dayhoff's series of scoring matrices (PAM matrices). Estimation of parameters of mutation matrices using alignments of closely related sequences. Derivation of scoring functions for amino acid substitutions observed in BLOCKS database. Series of BLOSUM matrices.
3. Markov models of DNA sequence evolution
   Jukes-Cantor's and Kimura's models. Rate of mutation matrices and matrices of transition probabilities. Amino acid classification. Markov model of protein sequence evolution.
4. Pattern discovery/motif finding
   Detecting genes and motifs on the genome. In response to environmental changes, the cell reads a complex code of regulatory signals which dictate gene usage. These sequence patterns are extremely short (10 nucleotides), often degenerate, and occur at varying distances from the gene start. We will develop algorithms for detecting these elements amidst oceans of non-functional nucleotides.
5. Phylogenetics
   Construction of a tree by using pairwise distances. UPGMA clustering and neighbors joining clustering algorithm. Notion of parsimony. Probabilistic approaches to phylogeny. Random genetic drift. Molecular clock. Synonymous and non-synonymous substitutions. Using maximum likelihood approach for phylogenetic inference. Orthologs and paralogs. Evolutionary and comparative genomics.
6. Biological network
   Random graphs, small world model; scale free model, hierarchical model. Network properties:

degree distribution, clustering coefficient, distance related measures. Network motifs, feed forward loop, motif conservation and evolution, noise in real life networks. Regulation of gene expression. Transcriptional networks. The feed-forward loop network motif. The single-input module network motif. Network motifs in developmental, signal transduction and neuronal networks. Genetic interaction networks: high order genetic interactions, modules and module finding, reconstruction of genetic networks from gene expression data. Protein interaction networks: protein complexes, scoring schemes, confidence assignment, RoC curves, inference of protein interactions based on genomic context and comparative genomics. Two-hybrid assays for mapping protein-protein interactions. Signal transduction networks: basic components and mechanisms, integration with gene expression networks.

7. High-dimensional data
   Gene expression data, imaging data; dimensionality reduction methods.
8. Clustering
   K-means clustering, hierarchical clustering, spectral clustering, graph-based clustering.
9. Structure data
   Protein structure and RNA structure

Course Schedule

A tentative schedule can be found here (link to-be-added) and is subject to change depending on the enrollment size.

Important Dates

- 9/29/2025 Midterm Exam
- 10/23/2025 Project proposal due
- 10/25/2025 (4pm) Withdraw deadline
- 10/27/2025 (4pm) Grade mode change deadline
- 12/5/2025 Project report due

Course Requirements & Grading

Assignments

Assignments	Weight
Homework (3x)	25%
Midterm exam	20%
Paper presentation	20%
Project - proposal	5%
Project - final report	25%
Class participation	5%
Total	**100%**

Grading scale

Your final grade will be assigned as a letter grade according to the following scale:

A	90-100%
B	80-89%
C	70-79%
D	60-69%
F	0-59%
* If you are taking this course with a S/U (Pass/Fail) grade, you need to earn at least 70% of the grades to pass this course.	

Homework

Homework assignments will include written or programming problems. Major programming languages used are Python or R, which will be specified in the assignment. Students are expected to complete homework problems individually. All homework assignments should be submitted in PDF (together with source code if there is a programming problem) on Canvas/GradeScope.

Paper presentations

We will discuss cutting-edge research papers during the course. A paper list will be released by the instructor, and two papers will be discussed per lecture during the second half of this course ("Advanced topic" lecture series). For each paper will be presented by one or two students, depending on the size of the class.

- **Presentation**: Each team will present a paper, selected from the paper list provided by the instructor, with a 25-30 min time slot in the class and answer questions by the instructor and other students. The presentation slides should be made as visual (with videos, images, and animations) and clear as possible. Students should practice their talks ahead of time to make sure they are of appropriate length -- not shorter by more than a few minutes, and certainly not longer (we will set a timer that will go off). The presentations should be well-organized and polished. The presentation will be graded according to the following criteria:
    - Quality of Slides: How good are the design, organization, and content in the presentation slides? It is expected to include visuals, diagrams, and bullet points to effectively convey complex information and support the presentation narrative.
    - Presentation Clarity: How clearly and coherently did the presenter(s) communicate their ideas? It includes the ability to convey complex concepts in an accessible manner, maintain a logical flow of ideas, and engage the audience.
    - Question Addressing: How effective are the responses to questions from the audience? This involves demonstrating a thorough understanding of the topic and providing clear and concise answers.
    - Ending on Time: Whether the presenters managed their presentation time effectively, ensuring that all key points were covered within the given duration without rushing or overly exceeding the allotted time.

Course project

The course project is a group assignment comprising 3-4 members (depending on enrollment). The project group can have different members from the presentation group. The course project is meant for students to gain experience in implementing algorithms or models and applying them to real-world biological data or have a more in-depth understanding of a particular problem by literature review.

Examples of project ideas include (1) formulating a novel problem in biology as a computational problem and implementing methods you learned in or outside this class to address it; (2) developing novel computational methods and applying them to an existing computational biology problem; (3) creating a benchmarking dataset for comparing the performance of existing computational methods for a specific biology problem, which includes curating and unifying the biological data generated by different papers or databases and implementing or running existing methods for testing. Students are encouraged to be creative and come up with their own project ideas. Students who need help with project ideas should talk to the TA or the instructor. Every member of a group is expected to contribute a substantial part to the project. The contributions of each member should be clearly stated in the final project report.

Class participation

The participation points (5%) include in-class discussion points (4%) and online discussion points (1%). To earn the in-class discussion points, you should ask two questions to at least one presenting group. The question is expected to be in-depth and preferably prompt discussion, and simple clarification-based questions do not count for the point. You should log your question in a survey form released by the teaching team to receive the points. For online discussion: before each student presentation class, the teaching team will create a post on the online discussion platform. Within that post, you need to write a review of a paper that will be presented by other students in the next class, which includes a short summary (~100 words) and your comments (pros/cons of the paper). The review should be completed by the day before the presentation date. Each review counts as 1 participation point. You can earn at most 1 online discussion point.

Technology Requirements and Skills

Computer Hardware and Software

- Laptop or desktop computer with internet connection. Students do not have to have GPU-equipped computers for assignments or projects, as they can utilize free computational resources such as Google Colab to develop GPU-based deep learning models.
- This class will use Canvas to deliver course materials to students. All course materials and quiz assessments will take place on this platform. Gradescope will be used for the submission of assignments and the project. Ed will be used as the discussion platform. Zoom or Teams will be used for remote meetings if needed.

Course Expectations & Guidelines

University Use of Electronic Email

A university-assigned student e-mail account is the official university means of communication with all students at Georgia Institute of Technology. Students are responsible for all information sent to them via their university-assigned e-mail account. If a student chooses to forward information in their university e-mail account, he or she is responsible for all information, including attachments, sent to any other e-mail

account. To stay current with university information, students are expected to check their official university e-mail account and other electronic communications on a frequent and consistent basis. Recognizing that some communications may be time-critical, the university recommends that electronic communications be checked minimally twice a week.

Late and Make-up Work Policy

The late submission policy for homework assignments is as follows: full credit is given if submitted before the due date, 50% credit is given for submissions within 24 hours after the due date, no credit is given for submissions after 24 hours past the due date.

There will be no make-up work provided for missed assignments. If you are unable to present during one of the pre-defined presentation days please contact the instructor to coordinate a solution. Of course, emergencies (illness, family emergencies) will happen. In those instances, please contact the Dean of Students office. The Dean of Students is equipped to verify emergencies and pass confirmation on to all your classes. For consistency, we ask all students to do this in the event of an emergency. Do not send any personal/medical information to the instructor or TAs; all such information should go through the Dean of Students.

Plagiarism & Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. All students enrolled at Georgia Tech, and all its campuses, are to perform their academic work according to standards set by faculty members, departments, schools, and colleges of the university; and cheating and plagiarism constitute fraudulent misrepresentation for which no credit can be given and for which appropriate sanctions are warranted and will be applied. For information on Georgia Tech's Academic Honor Code, please visit http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/.

Each student (or project group) must write their own solutions, in their own words, and must properly credit all sources. You are encouraged to discuss problems and papers with others as long as this does not involve the copying of code or solutions. After discussions, all materials that are part of a submission should be wholly your own. Any public material that you use to gain an understanding of the materials (open-source software, help from a textbook, or substantial help from a friend, etc.) should be acknowledged explicitly in anything you submit to us. To re-emphasize, no matter what the source you cannot copy any existing code, from other students, online, or otherwise, and all code must be wholly your own code that you wrote by yourself. If you have any doubts about whether something is legal or not, please do check with the class Instructor or TA. Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Policy on the uses of AI tools

We treat AI-based assistance, such as ChatGPT and Copilot, the same way we treat collaboration with other people: you are welcome to talk about your ideas and work with other people, both inside and outside the class, as well as with AI-based assistants.

However, all work you submit must be your own. You should never include in your assignment anything that was not written directly by you without proper citation (including quotation marks and in-line citation for direct quotes).

Including anything you did not write in your assignment without proper citation will be treated as an academic misconduct case. If you are unsure where the line is between collaborating with AI and copying AI, we recommend the following heuristics:

Heuristic 1: Never hit "Copy" within your conversation with an AI assistant. You can copy your own work into your own conversation but do not copy anything from the conversation back into your assignment. Instead, use your interaction with the AI assistant as a learning experience, then let your assignment reflect your improved understanding.

Heuristic 2: Do not have your assignment and the AI agent open at the same time. Similar to the above, use your conversation with the AI as a learning experience, then close the interaction down, open your assignment, and let your assignment reflect your revised knowledge. This heuristic includes avoiding using AI directly integrated into your composition environment: just as you should not let a classmate write content or code directly into your submission, so also you should avoid using tools that directly add content to your submission.

Deviating from these heuristics does not automatically qualify as academic misconduct; however, following these heuristics essentially guarantees your collaboration will not cross the line into misconduct.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404)-894-2563 or http://disabilityservices.gatech.edu/, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Collaboration & Group Work

- **Homework**: Students can discuss the homework with any other students in the class but should write their solutions individually.

- **Presentation and project**: The paper presentation and project will be completed as a group. You will sign up as a team and work together throughout the semester. Your team is welcome (and encouraged) to discuss your presentation/project with other members of the class. Research is highly collaborative and exchanging ideas is expected. However, you must implement your project ideas, write project reports, and create paper presentations within your group.

- **External resources**: We allow and encourage any outside reading material, blog posts, related work, and the use of open-source software for use within your project. Your proposed problem, approach, experiment implementation, and project presentation slides should be the original work of your project group. Any used resources should be cited or acknowledged in your report.

Student-Faculty Expectations Agreement

At Georgia Tech we believe that it is important to strive for an atmosphere of mutual respect, acknowledgment, and responsibility between faculty members and the student body. See http://www.catalog.gatech.edu/rules/22/ for an articulation of some basic expectations that students can have of the instructor and that the instructor has of students. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, the instructor encourages students to remain committed to the ideals of Georgia Tech while in this class.

Student Use of Mobile Devices in the Classroom

As research on learning shows, unexpected noises, and movement automatically divert and capture people's attention, which means you are affecting everyone's learning experience if your cell phone, pager, laptop, etc. makes noise or is visually distracting during class. That said, many students find it useful to have a mobile device on hand to access course materials. With this in mind, we allow you to take notes on your laptop but request that students turn the sound off so that they do not disrupt other students' learning. In addition, if you are doing anything other than taking notes or looking at course materials on your laptop, please sit in the back row so that other students are not distracted by your screen.

Subject to Change Statement

The syllabus and course schedule may be subject to change. Changes will be communicated via the Canvas announcement tool. It is the responsibility of students to check Ed Discussions, email messages, and course announcements to stay current in their courses.