[ISYE 3803] Syllabus

Introduction for Bioinformatics

Instructor Information

Instructor	Email	Drop-in Hours & Location
Saurabh Sinha	saurabh.sinha@bme.gatech.edu	TBD

General Information

Description

In this course, students will be introduced to the field of bioinformatics, an interdisciplinary field that combines computer science, statistics, machine learning and mathematics to analyze biological data. The course will cover key problems of the field, such as sequence alignment, gene finding and gene expression analysis, presenting algorithms, probabilistic models and machine learning techniques used to solve them.

Pre- &/or Co-Requisites

BMED 2400 or ISyE 3770 or ISyE 3030.

Course Goals and Learning Outcomes

Students will

- understand how efficient algorithms enable practical software tools that tackle daunting computational challenges in biological sequence analysis,
- understand how biological and molecular knowledge can be codified in statistical and machine learning models to extract insights from large and noisy data sets,
- learn to apply algorithms and models to toy data sets,
- learn to work with an online bioinformatics server to analyze real data,

This course explores the subject of bioinformatics from a 'computational science' angle, i.e., understanding the various computational tasks undertaken by researchers and practitioners of the field, and appreciating the basic computational paradigms in vogue. The focus is not on training the student to be a bioinformatics specialist in the industry (e.g., someone who is well versed with a large number of available tools and their interfaces).

Course Requirements & Grading

Assignment	Date	Weight (Percentage, points, etc)
Homework (x 5)		30
Midterm + Final exam		50
Quizzes		20

Extra Credit Opportunities

None offered.

Description of Graded Components

Five homework assignments (30% of grade): solving problems and analyzing toy data sets using algorithms introduced in course. These will not involve programming.

In-class exams (midterm and final: 25% + 25%): Short answer questions and toy problems testing students' understanding of analytics concepts.

13-16 in-class quizzes (20% of grade): testing of conceptual understanding of materials discussed in prior lecture and/or current lecture.

Grading Scale

TBD

Course Materials

Course Text

None

Additional Materials/Resources

Suggested reading, but not required:

Bioinformatics. Polanski, Andrzej ; Kimmel, Marek; 2007 -- Edition: 1.

Bioinformatics: the machine learning approach. Baldi, Pierre ; Brunak, Søren. 2001 -- Edition: 2nd edition.

"Basics of Bioinformatics", Lecture Notes of the Graduate Summer School on Bioinformatics of China, Editors: Rui Jiang, Xuegong Zhang, Michael Q. Zhang.

An introduction to bioinformatics algorithms. Jones, Neil C.; Pevzner, Pavel.
Cambridge, MA : MIT Press; ©2004.

Course Website and Other Classroom Management Tools

TBD

Course Expectations & Guidelines

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. 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/.

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.

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.

Attendance and/or Participation

In-person attendance in lectures is strongly encouraged but not mandatory. In-class quizzes require attendance on those days.

Collaboration & Group Work

Students are required to complete the home assignments independently. However, discussion with others is allowed to the extent of helping understand the material. The purpose of student collaboration is to facilitate learning, not to circumvent it. The actual solution must be done by each student alone, and the student should be ready to reproduce their solution upon request. In any case, you must exercise academic integrity. See the University Policy on Academic Integrity, especially the section on plagiarism.

Use of internet sources, including *chatGPT*, to get ready solutions to homework problems, is strongly discouraged.

Extensions, Late Assignments, & Re-Scheduled/Missed Exams

Late submission of an assignment will result in a reduced grade for the assignment. An assignment is worth full credit at the due time and date. It is worth 80% credit if submitted up to 3 days (72 hours) after the due time and date. This default late policy is subject to change for specific assignments, and such change will be announced at the time of posting that assignment.

Student-Faculty Expectations Agreement

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

Student Use of Mobile Devices in the Classroom

You may use your mobile devices and/or laptop to take notes during class, as long as such usage does not disturb or distract other students or the instructor.

Additional Course Policies

Campus Resources for Students

Additional Syllabus Components

Course Schedule

Tentative list of topics:

- Mol bio primer. Protein, DNA, RNA, genes, transcription, translation, genetic code, gene structure (exon, intron, promoter), gene expression, gene regulation, genome, evolution, phylogenetic tree.
- Basic statistics: Probability, independence, conditional probability, Bayes theorem, random variables, discrete and continuous distributions, likelihood maximization.
- Dynamic programming
- Sequence Alignment
- Pattern Matching
- BLAST tool
- Motif finding
- Statistical testing and systems biology.
- Gene Ontology and Gene set analysis.
- Markov Chains and HMM
- Gene finding
- Sequencing and sequence assembly.
- Transcriptomics
- Clustering algorithms.
- Single-cell transcriptomics and cell type identification.
- Linear regression.
- Gene regulatory network reconstruction.
- Classification methods
- Artificial Neural Networks
- Convolutional Neural Networks
- Molecular Phylogenetics