

CSE3800/BME4800/CSE5800: Bioinformatics – Fall 2018

Lecture: TuTh 3:30PM-4:45PM, ITE 119



Instructor:

Ion Măndoiu
ion@engr.uconn.edu
Office: ITE 261
Office Hours:
Tu/We/Th 12pm-1pm

Course Description: This course is an introduction to the fundamental mathematical models and algorithmic techniques used in bioinformatics. Emphasis will be placed on modeling computational problems arising in biology as graph-theoretic, statistical, or mathematical optimization problems, and on designing, analyzing, and implementing efficient combinatorial algorithms for the latter. Covered algorithmic techniques will include exhaustive search, integer programming, greedy algorithms, dynamic programming, divide-and-conquer, graph algorithms, combinatorial pattern matching, clustering, hidden Markov models, and randomized algorithms. Biological applications covered will include restriction mapping, DNA sequencing, motif finding, pairwise and multiple sequence alignment, gene prediction, evolutionary trees, and genome rearrangements.

Prerequisites: BIOL 1107, CSE 1100 or 1010 or 1729, and either STAT 3025Q or STAT 3345Q.

Textbook:

- Neil C. Jones and Pavel A. Pevzner, *An Introduction to Bioinformatics Algorithms*, MIT Press, 2004. Electronic version available (on campus network) at <http://www.netLibrary.com/urlapi.asp?action=summary&v=1&bookid=125977>

Grade breakdown:

In-class quizzes	10%
Theoretical homeworks	25%
Programming assignments	25%
Final project	40%

The lowest quiz, homework, and programming assignment scores will be dropped from the final grade calculation.

Assignment submission: Solutions for theoretical homeworks must be submitted in electronic format via HuskyCT. Programming assignments must be submitted electronically via the Rosalind site at <http://rosalind.info>. Rosalind is a repository of intellectually stimulating problems of varying difficulty that are extracted from real challenges of molecular biology. Solutions can be prepared using any high-level programming language. You will be asked to process a dataset generated by Rosalind on your own computer and then upload or copy-paste the solution to Rosalind along with your source code. Each submitted solution is automatically checked for correctness, allowing you to fix potential problems before the due date. **To access programming assignments you must first enroll for the appropriate Rosalind class using the link corresponding to your section below:**

- CSE3800: <http://rosalind.info/classes/enroll/cbef2b2257/>
- BME4800: <http://rosalind.info/classes/enroll/5769e66f9e/>
- CSE5800: <http://rosalind.info/classes/enroll/0f0650a921/>

Late policy: All assignments are due at midnight on the specified due date. Late submissions for homework and programming assignments are allowed for up to three days after the original submission deadline, but 10% of the grade will be subtracted for each late day.

Final project: The final project, typically done in teams of 2-3 students, will give you the opportunity to study a bioinformatics problem in more depth. Suitable final project topics include surveys of bioinformatics topics not covered in the lectures, design and implementation of novel algorithms, and empirical comparisons of existing methods. A list of potential topics will be provided towards the middle of the semester, although you are strongly encouraged to devise your own. Project requirements will include submitting two intermediate progress reports and a written final report of 15-20 pages. You will also be required to give a short presentation on your project at the end of the semester.

HuskyCT: We have a HuskyCT site for the class; you can access it by logging in with your NetID and password at <https://learn.uconn.edu>. Please check this site regularly for class materials, grades, assignment clarifications, changes in class schedule, and other class announcements. You can also use HuskyCT to ask class-related questions and communicate with your peers and the instructor. Please observe basic etiquette by keeping your postings polite, concise, and on-topic. Before posting do take a look at the postings that are already there—it is possible that your question has already been answered. Appropriate questions are general questions about the material and clarifications on the assignments, for questions that are specific to your own work you should contact the instructor directly.

Academic honesty: You are expected to adhere to the highest standards of academic honesty. All submitted solutions must be your own work. For homework assignments and programming projects you may discuss ideas and concepts with others, but must not share written solutions or code. Use of published materials (including web resources) is allowed, but all sources should be explicitly acknowledged in your solutions. Violations will be reviewed and sanctioned according to the University Policy on Academic Integrity.

Students with disabilities: If you have a documented disability for which you are or may be requesting an accommodation, you are encouraged to contact the instructor and the Center for Students with Disabilities or the University Program for College Students with Learning Disabilities as soon as possible to better ensure that such accommodations are implemented in a timely fashion.

Tentative Schedule:

Week #	Lecture dates	Topics	Reading assignment
1	Aug 28 & 30	Intro to molecular biology	Ch. 3
2	Sept 4 & 6	Exhaustive search	Ch. 4
3	Sept 11 & 13	Greedy algorithms	Ch. 5
4	Sept 18 & 20	Dynamic programming	Ch. 6
5	Sept 25 & Sept 27	Dynamic programming	Ch. 6
6	Oct 2 & 4	Divide and conquer	Ch. 7
7	Oct 9 & 11	Graph algorithms	Ch. 8
8	Oct 16 & 18	Graph algorithms	Ch. 8
9	Oct 23 & 25	Pattern matching	Ch. 9
10	Oct 30 & Nov 1	Pattern matching	Ch. 9
11	Nov 6 & 8	Hidden Markov models	Ch. 11
12	Nov 13 & 15	Hidden Markov models	Ch. 11
13	Nov 27 & 29	Clustering and phylogenetic trees	Ch. 10
14	Dec 4 & 6	Randomized algorithms	Ch. 12