CSE3800/BME4800/CSE5800: Bioinformatics – Fall 2015

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

Instructor:

Ion Măndoiu
ion@engr.uconn.edu
Office: ITE 261
Office Hours:
Tu/Th 5pm-6pm
We 11:15am-1:15pm

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, 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:

Theoretical homework assignments	30%
Programming assignments	30%
Final project	40%

Assignment submission: Homework assignments will be mostly theoretic in nature and will be assigned bi-weekly. There will also be 3-4 individual programming assignments requiring you to implement complete solutions to specified bioinformatics problems in the programming language of your choice. All solutions must be submitted electronically using Moodle.

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. The lowest take-home assignment score and lowest programming assignment score will be dropped from the overall grade calculation.

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.

Moodle site: A Moodle site is set up for the class at http://dna.engr.uconn.edu/moodle/. You are strongly encouraged to check it frequently for course materials, grades, and announcements related to the class. You can also use it to ask general 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 and reading assignments**
1	Sept 1 & 3	Intro to molecular biology (Ch. 3)
2	Sept 8 & 10	Exhaustive search (Ch. 4, integer programming handout)
3	Sept 15 & 17	Greedy algorithms (Ch. 5)
4	Sept 22 & 24	Dynamic programming (Ch. 6)
5	Sept 29 & Oct 1	Dynamic programming (Ch. 6)
6	Oct 6& 8	Divide and conquer (Ch. 7)
7	Oct 13 & 15	Graph algorithms (Ch. 8)
8	Oct 20 & 22	Graph algorithms (Ch. 8)
9	Oct 27 & 29	Pattern matching (Ch. 9)
10	Nov 3 & 5	Pattern matching (Ch. 9)
11	Nov 10 & 12	Hidden Markov models (Ch. 11)
12	Nov 17 & 19	Hidden Markov models (Ch. 11)
13	Dec. 1 & 3	Clustering and phylogenetic trees (Ch. 10)
14	Dec. 8 & 10	Randomized algorithms (Ch. 12)