CSE3800/BME4800/CSE5800: BIOINFORMATICS
Fall 2013

Lectures: TTh 3:30-4:45pm, Laurel Hall 305

Instructor: Ion Mandoiu
Office: ITEB 261
Phone: 486-3784
E-mail: ion@engr.uconn.edu
Office hours: MW 12:15-1:30pm or by appointment

Textbook: Neil C. Jones and Pavel A. Pevzner, *An Introduction to Bioinformatics Algorithms*, MIT Press, 2004. Electronic version available at

http://www.netLibrary.com/urlapi.asp?action=summary&v=1&bookid=125977

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

Course outline: 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.

Tentative Schedule:

Week	Lecture dates	Topics
1	Aug 27 & 29	Intro to molecular biology (Ch. 3), Exhaustive search (Ch. 4)
2	Sept 3 & 5	Exhaustive search (Ch. 4), Integer programming (handout)
3	Sept 10 & 12	Greedy algorithms (Ch. 5)
4	Sept 17 & 19	Dynamic programming (Ch. 6)
5	Sept 24 & 26	Dynamic programming (Ch. 6)
6	Oct 1 & 3	Divide and conquer (Ch. 7)
7	Oct 8 & 10	Graph algorithms (Ch. 8)
8	Oct 15 & 17	Graph algorithms (contd.)
9	Oct 22 & 24	Pattern matching (Ch. 9)
10	Oct 29 & 31	Pattern matching (Ch. 9)
11	Nov 5 & 7	Hidden Markov models (Ch. 11)
12	Nov 12 & 14	Hidden Markov models (Ch. 11)
13	Nov 19 & 21	Clustering and Phylogenetic trees (Ch. 10)
14	Dec 3 & 5	Randomized algorithms (Ch. 12)

Grading: Grading will be based on:

- In-class quizzes (10%)
- Homework assignments (30%)
- Programming assignments (30%)
- Final project (30%)

Assignment submission: Homework assignments must be submitted in electronic format via HuskyCT. Programming assignments must be submitted electronically via the Rosalind site at rosalind.info (you must first enroll for the appropriate Rosalind class using the enrollment links posted under "Web links" on HuskyCT). 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.

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. You will be required to submit several progress reports and a written final report, and give a short presentation at the end of the semester.

Late policy: There will be no make-ups for missed quizzes or assignments. Late submissions for homework and programming assignments are allowed for up to three days after the original submission deadline, with a 10% penalty subtracted for each late day.

HuskyCT & Piazza: We have a HuskyCT site for the class; you can access it by logging in with your NetID and password at learn.uconn.edu. Please check this site regularly for class materials, grades, assignment clarifications, changes in class schedule, and other class announcements. For class discussions we will be using Piazza. Rather than emailing questions to the instructor you are strongly encouraged to post them on Piazza at piazza.com/uconn/fall2013/cse3800bme4800cse5800/home. The system is highly catered to getting you help fast and efficiently from both the instructor and classmates.

Academic honesty: You are expected to adhere to the highest standards of academic honesty. All submitted solutions must be your own work. You may discuss ideas and concepts with other people, but *must not share written solutions or computer code*. Use of published materials is allowed, but the sources should be explicitly stated in your submission. 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.