CSE3800/BME4800/CSE5800: Bioinformatics – Fall 2019

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

Instructor:

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

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 algorithms for the latter. Covered algorithmic techniques will include exhaustive search, greedy algorithms, dynamic programming, divide-and-conquer, graph algorithms, combinatorial pattern matching, clustering, and randomized algorithms. Biological applications covered will include motif finding, sequence assembly, pairwise sequence alignment, genome rearrangement analysis, gene expression analysis, and evolutionary tree reconstruction.

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

Textbook: *Bioinformatics Algorithms: an Active Learning Approach* by Phillip Compeau and Pavel Pevzner, 3rd edition. The book website, http://bioinformaticsalgorithms.com/, includes video lectures, slides, and answers to frequently asked questions. Additional materials will be distributed on Moodle for some of the topics.

Grade breakdown: Grading will be based on in-class and online quizzes given throughout the semester, theoretical homework assignments and programming assignments reinforcing the material covered in lectures, and a final project, according to the following breakdown:

Quizzes	10%
Theoretical homeworks	20%
Programming assignments	30%
Final project	40%

Assignment submission: Solutions for theoretical homework assignments and deliverables for the final project must be submitted in electronic format via Moodle (see below). Programming assignments must be submitted electronically via the Rosalind site at 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.

Late policy: All assignments are due by midnight on the specified due date. Late submissions are allowed for up to three days with a 10% penalty for each late day. Assignments that are more than three days late and make-up quizzes will not be allowed, however, to accommodate unforeseen circumstances that may prevent timely submission, the lowest quiz, homework assignment, and programming assignment scores will be dropped from the overall grade calculation.

Final project: The final project 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: Course announcements and class related materials including slides, handouts, assignments, grades, etc. will be distributed using Moodle. To get access to the Moodle site you must first create an account at https://dna.engr.uconn.edu/moodle/login/ then self-enroll in the Bioinformatics course using enrolment key "BIOINFO". *Note that this is a local Moodle installation and you will not be able to login using your netid credentials.* You can also use Moodle 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. Appropriate questions are general questions about the covered material and clarifications on the assignments. For questions that are specific to your own work you should contact the instructor directly.

Academic integrity: You are expected to adhere to the highest standards of academic integrity. 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 university policies.

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.