Please carefully read this and all other course related instructions, particularly those pertaining to the Academic Honesty Policy!

Course Mechanics:

Pre-requisites: A course in probability/statistics, fluency in a programming language such as C/C++ (or instructor permission)

Credit Hours: 4

Class Sessions: Tue, Thu 9:40-10:55 a.m., Hylan 306

Course Description:

The widespread availability of computation in the 21st is changing the landscape for research in Biology by bringing computational and high-throughput tools to the forefront. The seeds for this change have been laid over the past 50 years via fundamental discoveries in molecular biology. Specifically, the advances in molecular biology have led to a richer understanding of the chemistry of life and enabled experimental quantitative characterization of biological processes and systems. When coupled with modern automation tools, these experimental techniques generate large datasets, often ranging in the hundreds of gigabytes for a single experiment. Computational tools of bioinformatics are essential to make sense of the large datasets and to exploit the unprecedented resolution that they offer. The need for bioinformatics tools spans a wide spectrum, including, for example, methods designed for: (a) digesting raw data into information useful for biologists, (b) discovery of inter-relations through matching and comparison, (c) modeling molecular structure, (d) formulating hypotheses for experimental validation, and (e) higher level statistical inference across datasets gathered with diverse modalities. Collectively, these tools offer the tantalizing possibility of unraveling the complexity of biological systems to not only further our scientific understanding but also to allow targeted design of therapeutic interventions for overcoming diseases.

The course provides an introduction to the computational methods used in Bioinformatics with a particular emphasis on scalability to large scale datasets that are increasingly becoming the hallmark of twenty-first century Bioinformatics. The coverage includes a sampling of algorithmic techniques that is motivated by and presented in the context of specific applications such as high throughput sequencing, genome alignment and annotation, RNA secondary structure prediction, coding and noncoding gene discovery, estimation of genome/protein interaction networks, and flow cytometry data analysis. Topics such as dynamic programming algorithms, probabilistic and graphical models, clustering, mixture models, and statistical estimation/inference are introduced in the context of these applications along with specific approximation methods directed at improving the scalability. The class will feature several guest lectures by researchers working on individual problem areas. Students will complete three implementation mini-projects, a paper review, and a final term project.

Grading:

- Assignments 25%
- Class Presentations 25%
- Course Project 50%
Instructor:

Prof. Gaurav Sharma
Office hours: by appointment.
Office: Hopeman 417
Email: gsharma@ece.rochester.edu
Homepage: http://www.ece.rochester.edu/~gsharma

Course Web Site:

http://www.ece.rochester.edu/~gsharma/teaching/bioinformatics/

Text:
The course does not have a required text. Assigned readings related to each of the topics will be posted on the course website. Several books [1][2][3][4] will be available on reserve in the library for the course.

Academic Honesty Policy:
The University of Rochester academic honesty policy applies to all assignments and exams for this class. The full-text of the academic honesty policy can be found at: http://www.rochester.edu/College/CCAS/AdviserHandbook/AcadHonesty.html

In addition to the general guidelines mentioned in the above policy, for this course I require that: Any external material used should be clearly cited. In your own writings (example project reports, assignment solutions, proposals etc), no more than one or two sentences may be used verbatim from any source. READ THESE INSTRUCTIONS CAREFULLY! If any aspect of the academic honesty policy and guidelines for this course are unclear, please ask me for clarifications. Lack of awareness or understanding of this policy will not be an acceptable excuse or defense against disciplinary action.

References