CSCI2950-C: Computation and Cancer: Spring 2013

Meeting time: Tu/Th 1-2:20pm CIT 241
Instructor: Ben Raphael, CIT 505, braphael@brown.edu
Website: http://cs.brown.edu/courses/csci2950-c/

Description

It has been known for decades that cancer is driven largely by mutations that accumulate in an individual's genome during their lifetime. However, it was only in 2008 that the first cancer genome was sequenced, an advance made possible by new DNA sequencing technologies. Cancer sequencing data is now being generated at an exponentially increasing rate. The challenge has shifted from producing cancer genome sequencing data to interpreting this data in order to advance cancer biology and treatment. This seminar will explore algorithms, statistical methods, and techniques from machine learning that address four important challenges in cancer genome sequencing and interpretation.

1) Identification of different classes of somatic mutations (single-nucleotide variants, copy number aberrations and rearrangements) from high-throughput DNA sequencing data.
2) Reconstruction of tumor evolution from DNA sequencing data using techniques from phylogenetics and population genetics.
3) Comparison of cancer genomes from different individuals to distinguish causal somatic mutations from random mutations.
4) Analysis of combinations of mutations in signaling pathways and cellular interaction networks.

The course will be organized in seminar style where students will read and present recent research papers on the topics listed above. Each topic will be introduced with introductory lectures. Students will undertake a project to further study one of the topics. To the extent possible, projects will be adjusted to the background/interest of the student and could range from theoretical (e.g. designing a new algorithm and proving its correctness), to the practical (a software implementation). The project will include a written proposal, midterm report, and final presentation.

Prerequisites

- Undergraduate-level knowledge of probability: random variables, distributions, etc.
- Undergraduate-level knowledge of algorithms and/or statistics

No biology background is assumed. Necessary background will be introduced in lectures and reading.

Grading

Papers: 45% total
- Reviews 20%
- Presentations 15%
- Discussion Summaries 10%

Project: 45% total
- 10% Written proposal.
- 10% Written midterm report.
- 15% Final written report.
- 10% Final presentation.
Computer Assignment: 5%
Participation: 5%

**Papers, Presentation, and Discussion**

**Presentations:** Each student will present at least two papers from the reading list. The exact number will depend on class size. Presentations should be designed to facilitate a class discussion of the paper, and may use slides, whiteboard, or a combination of both. A long presentation is expected to be about 40 minutes, and a short presentation 20 minutes with the remaining class time for questions and discussion. Students are strongly encouraged to discuss presentation with instructor beforehand.

**Reviews:** Each student (except presenter) will submit a written review for each discussion paper, with the exception of those that you present. Reviews will be submitted online (details to come) by 10am on the day that the paper is presented. Reviews will be graded on a 3-point scale: 2 = mastered the key issues introduced in the paper 1 = read the paper and understood the basics; 0 otherwise.

**Discussion Summary:** One student will be the appointed scribe for each presentation. Following the paper discussion, the presenter and scribe will prepare a written (document or slide) summary of the merits and deficiencies of the paper. This summary should incorporate material from the presentation (e.g. slides), the class paper reviews, and the discussion. The summary will be posted on the course webpage.

**Project**

The project is a semester-long effort to further study one of the class topics. The project could range from theoretical (e.g. designing a new algorithm and proving its correctness), to the practical (a software implementation) depending on the interest of the student. Projects can be undertaken in groups of 1-3 students. Students are encouraged to propose a project matching their research interests and discuss with instructor. A list of suggested projects will also be distributed.

**Participation**

All students are expected to contribute to paper discussions by asking questions, making observations, identifying strengths and weaknesses of the approaches under discussion. Critical discussion and analysis are key prerequisites for research.

**Course Credits**

PhD: Area T (Theory) [Pre-2012: Area B (Algorithms)]
ScM: “Theory” or “Practice”* course.
Significant Programming*
*With appropriate class project.