Information for ECE1508F: Machine Learning and Genomic Medicine

1 Course objectives

The cost of sequencing a genome is expected to drop to $20 within two years and it is now possible to edit DNA in living tissues. These technologies create the possibility of transforming medicine globally, saving lives and preventing or curing disease. However, achieving this requires that we make accurate computational predictions of how genetic alterations, whether they are naturally occurring disease-causing mutations or therapeutic genetic modifications, will change the cellular processes involved in disease. *In this course, we will study one of the most powerful approaches, which is to use machine learning to infer models that relate genetic sequence to cellular processes and then use these models to predict the effects of genetic mutations and gene-altering therapies.* Companies like Google, 23andMe, Illumina and Roch are making huge investments in this field. This course will review crucial techniques, such as deep learning and high throughput experimental biology, and drill down on how these techniques can be used to transform medicine. The focus will be on defining machine learning challenges that will disrupt the field of genomic medicine, and on implementing and testing machine learning algorithms that address these challenges.

2 Instructor

Brendan J. Frey, BA 4136, frey@psi.utoronto.ca, www.genes.toronto.edu

3 Prerequisites

Students should have already taken a course on machine learning and have a good working knowledge of basic machine learning methods, such as regression, classification, clustering, and neural networks. Students should have a background in genome biology and be familiar with concepts such as DNA, RNA, replication, transcription, translation, introns, exons, and experimental methods. Students should be proficient in programming, including using Python. Contact me if you have questions.

4 Learning Materials

- Genome biology primer: http://www.ebi.ac.uk/microarray/biology_intro.html.
- Machine learning: Chris Bishop's book Pattern Recognition and Machine Learning.
- Genome biology: Tom Strachan, Andrew Read, Human Molecular Genetics, Fourth Edition 4th Edition.

5 Assignments, midterm test and final test

There will be two 2-hour tests (not formal exams), plus three assignments, which will involve implementing algorithms taught in class and applying them to genome biology datasets that will be posted on the course website. The distribution of grades and due dates is as follows:

Project	60%
Midterm test	20%
Final test	20%