Welcome

CSCI 1850: Deep Learning in Genomics
Spring 2021

http://cs.brown.edu/courses/csci1850

Instructor: Ritambhara Singh
Format: Online (Synchronous)
Time: TTh 10:30-11:50 AM
About the instructor

Research Interests
Machine Learning, Data Mining, Computational Biology, Health Sciences

Office Location
Room 313, Center for Computational Molecular Biology (CCMB)
3rd Floor, 164 Angell Street

Office Hours
Thursdays, 12-1 PM EST or by appointment (will be held over Zoom)

Email: ritambhara@brown.edu
Website: www.ritambharasingh.com
About the TAs

Daniel Ben-Isvy (HTA)

Office Hours
Wednesdays: 4-6 PM EST
(will be held over Zoom)

Giselle Garcia (UTA)

Office Hours
Mondays: 8-10 PM EST
(will be held over Zoom)

Email: cs1850tas@lists.brown.edu
Pre-requisites

**Required:**
- Programming
- Machine learning (courses or research experience)
  and/or
- Deep Learning (courses or research experience)

**Good to have:**
- Experience of working with deep learning frameworks
- Familiarity with genomics datasets
Why take this course?

- Molecular signatures of tumor / blood sample
- Signs & Symptoms
- Patient Medical History & Demographics
- Public Health Data
- Genetic Data
- Mobile medical sensor data

Too much data to analyze. We need to go deeper.

Information Explosion

- Volume
- Variety
- Velocity

Analysis Gap

Ability to Analyze

Deep Learning
Important questions

• Is it appropriate to use deep learning for our application?

• What model should we use?

• Will our approach improve our understanding of the data or the problem?

• What are the different genomics tasks and deep learning models? How they fit together?
Course objectives

• **Connect** different state-of-the-art models to applications in genomics

• **Extract** key ideas from research papers when *solving* homework assignments

• **Think** critically about using a deep-learning method for a new task

• **Collaborate** with classmates on a team project to *apply* deep learning models to a genomics task

• **Communicate** your findings clearly by writing a report and through oral presentations.
How will you accomplish that?

- HOMEWORK (30%)
- CLASS DISCUSSIONS (10%)
- COURSE-PROJECT (60%)
4 Homework assignments

- 5-6 conceptual questions based on papers related to the topics covered during the lectures

- 1 programming assignment to familiarize you with genomics datasets and deep learning frameworks

Refer to course syllabus sheet for due dates
So what topics are we covering?

- Predicting from DNA Sequences
- Predicting from other genomic signals
- Imputation and single-cells
- Interpretation of deep learning models in genomics
- Other interesting applications
Class discussions

- Actively ask questions during the class
- Engage in discussions during the class
- Answer questions raised by your peers
- Post questions and comments on Piazza outside the class

Keep it concise, be respectful, and Always, always provide constructive criticism! 😊
Questions?
Ice-breaker
(a.k.a “please-don’t-make-me-do-this” activity)

“Hello my name is ...”
One thing on your bucket list that you hope to do once the pandemic gets over
Course project (Kaggle competition)

- Form a team of 2

**Round 1 (Mid-term)**
- We will release the details of the prediction task and datasets
- Apply deep learning model to solve the task
- **Mid-term**: Submit trained model, present the work, and write a report
- We will release score board on held out dataset

**Round 2 (Final)**
- We will update the task and introduce new datasets
- Utilizing previous results + new data, update the model for the task
- **Final**: Submit trained model, present the work, and write a report
- We will release score board on held-out dataset

Refer to course syllabus sheet for due dates
Team-work is critical!

Project plan, team check-ins, and code repository would be used to monitor contributions and progress.
Assignment submission policies

- All submissions are due at 11:49 PM EST of the due date
- You can get a 3-day extension for at most 3 deadlines without penalty
- 20% of points will be deducted for late submission and missed assignment will get 0 score
- Extension for any other circumstance can only be granted by the instructor
- If needed, you can switch presentations with another student/team and inform the instructor
How is this different from CSCI 2952-G (Fall)?

Graduate seminar course (research based):

- Cover around ~20 research papers (similar topics)
- Paper reading and student presentations
- Writing paper summaries
- Class discussions on papers
- Course research project – literature review, idea presentation, workshop paper writing, final project presentations

Will be offered in Fall 2021!
Questions?
Class Exercise: Think-pair-share

• Think [5 mins]:

(a) Broadly speaking, what advantages do you think deep neural networks have in genomics tasks over traditional machine learning methods?

(b) What do you think could be some of the challenges of working with genomics data?

• Pair (Break out rooms)
• Share [5 mins]

https://docs.google.com/document/d/1_QkQnie1j0B1rT_GiSOkK1W2DGrn4U8_jo
de29TQjRG0/edit?usp=sharing
Questions?
Next week

Course website: http://cs.brown.edu/courses/csci1850

Section I: Predicting from DNA sequences

• Jan 26, Tuesday:
  Transcription Factor Binding Prediction: task, biology, previous works

• Jan 28, Thursday:
  Transcription Factor Binding Prediction: deep learning models

• Jan 29, Friday:
  Course setup form due at 11:59 PM EST
Wrap up

What was the clearest point today?

What was the muddiest point today?

https://forms.gle/nh3EZL9oAH4cRSaj8