Course Syllabus

Jump to Today

Course Overview

Recent breakthroughs in high-throughput genomic and biomedical data are transforming biological sciences into "big data" disciplines. In parallel, progress in deep neural networks are revolutionizing fields such as image recognition, natural language processing and, more broadly, AI. This course explores the exciting intersection between these two advances. The course will start with introduction to deep learning and overview the relevant background in genomics and high-throughput biotechnology, focusing on the available data and their relevance. It will then cover the ongoing developments in deep learning (supervised, unsupervised and generative models) with the focus on the applications of these methods to biomedical data, which are beginning to produced dramatic results. In addition to predictive modeling, the course emphasizes how to visualize and extract interpretable, biological insights from such models. Recent papers from the literature will be presented and discussed. Students will work in groups on a final class project using real world datasets.

Prerequisites

College calculus, linear algebra, basic probability and statistics such as CS109, and basic machine learning such as CS229. No prior knowledge of genomics is necessary.

Lecture Venue and Times

09/25/2017 - 12/09/2017 Mon, Wed 3:00 PM - 4:20 PM at Hewlett Teaching Center 201 (http://campus-map.stanford.edu/?srch=Hewlett+Teaching+Center+201)

Recitation. Fridays 10:30am - 11:20am at 380-380D (http://campus-map.stanford.edu/?srch=380-380D)

Instructors

Anshul Kundaje (https://sites.google.com/site/anshulkundaje/), Assistant Professor (akundaje@stanford.edu (mailto:akundaje@stanford.edu))

James Zou (https://sites.google.com/site/jamesyzou/), Assistant Professor (jamesz@stanford.edu (mailto:jamesz@stanford.edu))

Office hours

James Zou (https://sites.google.com/site/jamesyzou/) **:** Mondays 4:30-6pm (Packard 253).

Anshul Kundaje (http://anshul.kundaje.net) **:** Mondays 4.30-5.30 pm (Lane Med School Building L301 (https://www.google.com/maps?q=Lane+building+Stanford&hl=en&ll=37.432675,-122.175546&spn=0.002461,0.002733&sll=42.361864,-71.090563&sspn=0.012954,0.021865&t=h&hnear=Lane+Bldg,+
)

Assignments

Course project (60%): the students will form teams of 3-5 and choose from one of the suggested projects or select their own project. Teams will be given Microsoft Azure credits to implement algorithms and perform analysis. Teams are expected to work on the research project throughout the second half of the quarter and produce conference-style papers. Each team will present the paper to the entire class at the end of the semester. The course project will consists of the following **milestones**:

1. Project proposal in class (3 minute talk).
2. First draft of the paper for peer review.
3. Poster presentation (12/11 3:30-6:30pm).
4. Final paper (due at noon on Friday 12/15).

A significant portion of the class will be based on reading and discussing the latest literature. Every student should read the assigned papers before class and participate in discussions.

Paper presentation (20%): each team selects one of the suggested papers to present in detail to the class. The presentation should be 20 mins + 5 mins for Q&A. Each team will also write a concise review of the paper. The review will be published on bioRxiv and the Zou group blog (https://zou-group.github.io/blog/archive/) .

Peer project review (10%): each team will be assigned two other groups' paper drafts to review. The review should concisely summarize the key findings of the paper, highlight interesting ideas, weaknesses and give suggestions.

Class participation (10%): every student should actively engage in paper discussions in class.

Schedule

The first few lectures will cover the basics of deep learning---convolutional and recurrent architectures, generative models, and optimization/regularization. We will also study the applications of deep learning in several biomedical domains---genomics, protein structure, imaging and medical records.

Date	Topic	Papers
9/25	Overview. Intro to machine learning	1. Deep Learning **http://www.nature.com/nature/journal/v521/n7553/full/nature14539.html** **(http://www.nature.com/nature/journal/v521/n7553/full/nature14539.html)** **(http://www.nature.com/nature/journal/v521/n7553/full/nature14539.html)** 2. Neural Nets and Deep learning primer **http://neuralnetworksanddeeplearning.com/** **(http://neuralnetworksanddeeplearning.com/)** (See other readings in Files Section)
9/27	Genomics	1. Neural Nets and Deep learning primer **http://neuralnetworksanddeeplearning.com/** **(http://neuralnetworksanddeeplearning.com/)** 2. Deep learning for computational biology **http://msb.embopress.org/content/12/7/878** **(http://msb.embopress.org/content/12/7/878)** (See other readings in Files Section)
10/2	DenseNets + Convolutional Nets for Genomics	1. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks **http://genome.cshlp.org/content/26/7/990.full** **(http://genome.cshlp.org/content/26/7/990.full)** 2. Denoising genome-wide histone ChIP-seq with convolutional neural networks **https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btx243** **(https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btx243)** **(https://www.ncbi.nlm.nih.gov/pubmed/27084946)** (See other readings in Files Section)
10/4	Recurrent NN	1. DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences **https://www.ncbi.nlm.nih.gov/pubmed/27084946** **(https://www.ncbi.nlm.nih.gov/pubmed/27084946)**
10/9	Autoencoders + representation learning	
10/11	Optimization + regularization + Azure Demo	
10/16	Generative models	GANs for Biological Image Synthesis **https://arxiv.org/abs/1708.04692** **(https://arxiv.org/abs/1708.04692)**
10/18	Instructor led paper presentations	1. Basenji: Sequential regulatory activity prediction across chromosomes with convolutional neural networks **https://www.biorxiv.org/content/early/2017/07/10/161851** **(https://www.biorxiv.org/content/early/2017/07/10/161851)** 2. FIDDLE: An integrative deep learning framework for functional genomic data inference **https://www.biorxiv.org/content/early/2016/10/17/081380** **(https://www.biorxiv.org/content/early/2016/10/17/081380)** 3. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning **https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1189-z** **(https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1189-z)**
10/23	Interpretation of black-box models	1. The Mythos of Model Interpretability **https://arxiv.org/abs/1606.03490** **(https://arxiv.org/abs/1606.03490)** 2. DeepLIFT: Learning Important Features Through Propagating Activation Differences **https://arxiv.org/abs/1704.02685v1** **(https://arxiv.org/abs/1704.02685v1)** 3. Deep Motif Dashboard: Visualizing and Understanding Genomic Sequences Using Deep Neural Networks **https://arxiv.org/abs/1608.03644** **(https://arxiv.org/abs/1608.03644)**
10/25	Paper presentations	1. Attend and Predict: Understanding Gene Regulation by Selective Attention on Chromatin **https://arxiv.org/abs/1708.00339** **(https://arxiv.org/abs/1708.00339)** 2. Deep learning for population genetics

		http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845 (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845) 3. Integrative deep models for alternative splicing https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btx268 (https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btx268) (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845) (https://www.biorxiv.org/content/early/2017/02/03/10595
10/30	Proposal presentations	
11/1	Guest lecture: Google Brain team	1. DeepVariant: Creating a universal SNP and small indel variant caller with deep neural networks https://www.biorxiv.org/content/early/2016/12/21/092890 (https://www.biorxiv.org/content/early/2016/12/21/092890) 2. Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning https://www.biorxiv.org/content/early/2017/09/12/179531 (https://www.biorxiv.org/content/early/2017/09/12/179531) 3. Adaptive Somatic Mutations Calls with Deep Learning and Semi-Simulated Data https://www.biorxiv.org/content/early/2016/10/04/079087 (https://www.biorxiv.org/content/early/2016/10/04/079087)
11/6	Drug discovery + protein structure	1. Deep learning for computational chemistry (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845) http://onlinelibrary.wiley.com/doi/10.1002/jcc.24764/abstract (http://onlinelibrary.wiley.com/doi/10.1002/jcc.24764/abstract) 2. (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845) MoleculeNet (https://arxiv.org/pdf/1703.00564v1.pdf) (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845) (https://arxiv.org/pdf/1703.00564v1.pdf) 3. (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845) One shot learning drug discovery (http://pubs.acs.org/doi/pdf/10.1021/acscentsci.6b00367)
11/8	Paper presentations	1. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005324 (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005324) 2. De novo peptide sequencing by deep learning http://www.pnas.org/content/114/31/8247 (http://www.pnas.org/content/114/31/8247) 3. Molecular De Novo Design through Deep Reinforcement Learning https://arxiv.org/abs/1704.07555 (https://arxiv.org/abs/1704.07555)
11/13	Paper presentations	1. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico http://pubs.acs.org/doi/abs/10.1021/acs.molpharmaceut.7b00346 (http://pubs.acs.org/doi/abs/10.1021/acs.molpharmaceut.7b00346) 2. Molecular Graph Convolutions: Moving Beyond Fingerprints https://arxiv.org/abs/1603.00856 (https://arxiv.org/abs/1603.00856) , 3. Convolutional Networks on Graphs for Learning Molecular Fingerprints http://papers.nips.cc/paper/5954-convolutional-networks-on-graphs-for-learning-molecular-fingerprints (http://4.%20Convolutional%20Networks%20on%20Graphs%20for%20Learning%20Molecular%20Fingerprints%20http://papers.nips.cc/paper convolutional-networks-on-graphs-for-learning-molecular-fingerprints)
11/15	Imaging + Electronic Medical Records	1. Deep learning for healthcare: review, opportunities and challenges https://academic.oup.com/bib/article-abstract/doi/10.1093/bib/bbx044/3800524/Deep-learning-for-healthcare-review-opportunities (https://academic.oup.com/bib/articl abstract/doi/10.1093/bib/bbx044/3800524/Deep-learning-for-healthcare-review-opportunities) 2. Deep EHR: A Survey of Recent Advances on Deep Learning Techniques for Electronic Health Record (EHR) Analysis https://arxiv.org/abs/1706.03446 3 (https://arxiv.org/abs/1706.03446) . Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning https://arxiv.org/abs/1708.09843v2 (https://arxiv.org/abs/1708.09843v2)
11/27	Paper presentations	1. Privacy-preserving generative deep neural networks support clinical data sharing https://www.biorxiv.org/content/early/2017/07/05/159756.1 (https://www.biorxiv.org/content/early/2017/07/05/159756.1) 2. Dermatologist-level classification of skin cancer with deep neural networks http://www.nature.com/nature/journal/v542/n7639/abs/nature21056.html (http://www.nature.com/nature/journal/v542/n7639/abs/nature21056.html) 3. Medical Concept Representation Learning from Electronic Health Records and its Application on Heart Failure Prediction https://arxiv.org/ftp/arxiv/papers/1602/1602.03686.pdf (https://arxiv.org/ftp/arxiv/papers/1602/1602.03686.pdf)
11/29	Paper presentations	1. Learning to Detect Sepsis with a Multitask Gaussian Process RNN Classifier

		http://proceedings.mlr.press/v70/futoma17a/futoma17a.pdf (http://proceedings.mlr.press/v70/futoma17a/futoma17a.pdf) 2. RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism https://arxiv.org/abs/1608.05745 (https://arxiv.org/abs/1608.05745) 3. Leveraging uncertainty information from deep neural networks for disease detection https://www.biorxiv.org/content/early/2017/08/02/084210 (https://www.biorxiv.org/content/early/2017/08/02/084210)
12/4	Paper presentations	1. Learning Sleep Stages from Radio Signals:A Conditional Adversarial Architecture http://sleep.csail.mit.edu/files/rfsleep-paper.pdf (http://sleep.csail.mit.edu/files/rfsleep-paper.pdf) 2. Interpretable Deep Models for ICU Outcome Prediction http://www-scf.usc.edu/~zche/papers/amia2016.pdf (http://www-scf.usc.edu/~zche/papers/amia2016.pdf) 3. Reconstructing cell cycle and disease progression using deep learning https://www.nature.com/articles/s41467-017-00623-3 (https://www.nature.com/articles/s41467-017-00623-3%20)
12/6	Wrap up	Opportunities/Obstacles Deep Learning Biomedicine https://www.biorxiv.org/content/early/2017/05/28/142760 (https://www.biorxiv.org/content/early/2017/05/28/142760)
12/11	Finals week: poster presentation	

Course Summary:

Date	Details
	Paper Presentation (https://canvas.stanford.edu/courses/70852/assignments/100969)