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

09/26/2016 - 12/09/2016 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 Hewlett 102

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/)** **:** Wednesdays 5-7pm (Packard 253).

Anna Shcherbina and Nadine Hussami: Mondays 5-7pm (Lane L339)

Jayanth and Alon: Thursdays 10:30am-12:30pm (Huang Basement)

Assignments

Course project (50%): the students will form teams of 4-6 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.

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.

Paper review (20%): each team selects 2 other papers to review. The review concisely summarize the key findings of the paper, highlight interesting ideas, weaknesses and potential extensions.

Class participation and quizzes (10%): every student should actively engage in paper discussions in class and in the online forum. We will also have a few in class quizzes.

Tentative outline

[10 weeks of instruction; 20 classes]

Module 1: introduction to deep learning and demos (7 classes).

Module 2: Applications of deep learning to regulatory genomics, variant scoring and population genetics (4 classes)

Module 3: Applications of deep learning to predicting protein structure and pharmacogenomics (3 classes)

Module 4: Applications of deep learning to electronic health records and medical imaging data (4 classes)

Project presentations (Exam period)

Date	Topic	Primary instructor	papers	paper URLs	Other relevant link
9/26	Intro to neural networks, backprop	James	Stegle Review	http://msb.embopress.org/content/12/7/878 , http://neuralnetworksanddeeplearning.com/	
9/28	Convolutional neural network + intro to func genomics	Anshul	DeepBind	http://www.nature.com/nbt/journal/v33/n8/full/nbt.3300.html,	
10/3	Conv nets for genomics and imaging (contd)	Anshul, Serafim	DeepCpG	http://biorxiv.org/content/early/2016/05/27/055715	
10/5	Interpretation of deep learning models	Avanti		See Files Section	
10/10	Recurrent neural network + autoencoders + EHR data	James	DeepNano: Deep Recurrent Neural Networks for Base Calling in MinION Nanopore Reads	http://arxiv.org/abs/1603.09195	
10/12	Training deep neural networks + protein structures	James			
10/17	Azure, Tensorflow and Keras demo	James			
10/19	Func Genomics/Variant/PopGen	James	DeepSEA Basset	**http://www.nature.com/nmeth/journal/v12/n10/full/nmeth.3547.html (http://www.nature.com/nmeth/journal/v12/n10/full/nmeth.3547.html)** **http://genome.cshlp.org/content/26/7/990 (http://genome.cshlp.org/content/26/7/990)**	
10/24	Func Genomics/Variant/PopGen	James	DanQ The human splicing code reveals new insights into the genetic determinants of disease	**http://nar.oxfordjournals.org/content/44/11/e107 (http://nar.oxfordjournals.org/content/44/11/e107)** **http://sites.utoronto.ca/intron/xiong2015.pdf (http://sites.utoronto.ca/intron/xiong2015.pdf)**	
10/26	Func Genomics/Variant/PopGen	Anshul	DeepGDashboard Learning structure in gene expression data using deep architectures, with an application to gene clustering	**http://arxiv.org/abs/1608.03644 (http://arxiv.org/abs/1608.03644)** **http://biorxiv.org/content/biorxiv/early/2015/11/16/031906.full.pdf (http://biorxiv.org/content/biorxiv/early/2015/11/16/031906.full.pdf)**	
10/31	PopGen, Small molecules	James, Serafim	Deep Learning for Pop Gen Inference Automatic chemical design using a data-driven continuous	**http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845 (http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004845)** **https://arxiv.org/pdf/1610.02415v1.pdf (https://arxiv.org/pdf/1610.02415v1.pdf)**	

			representation of molecules		
11/2	Protein Structure	James	Protein secondary structure prediction using deep convolutional neural fields	**https://arxiv.org/abs/1512.00843 (https://arxiv.org/abs/1512.00843)**	
11/7	Project proposal	James, Serafim	project proposal lightning talks		**https://github.com (https://github.com/**
11/9	Pharmacogenomics	James	Convolutional LSTM Networks for Subcellular Localization of Proteins Protein contact map prediction using ultra deep residual nets	**https://arxiv.org/pdf/1503.01919.pdf (https://arxiv.org/pdf/1503.01919.pdf)** http://biorxiv.org/content/early/2016/09/06/073239	DeepTox: Toxicity P http://journal.frontie
11/14	Pharmacogenomics	Anshul	Molecular Graph Convolutions: Moving Beyond Fingerprints AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery Written presentation only: Massively Multitask Networks for Drug Discovery	https://arxiv.org/abs/1603.00856 **https://arxiv.org/abs/1510.02855 (https://arxiv.org/abs/1510.02855)** **https://arxiv.org/pdf/1502.02072v1.pdf (https://arxiv.org/pdf/1502.02072v1.pdf)**	
11/16	Med Records/Clinical data	Anshul, Serafim	Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records; Deep Kalman Filters	**http://www.nature.com/articles/srep26094 (http://www.nature.com/articles/srep26094)** **https://arxiv.org/abs/1511.05121 (https://arxiv.org/abs/1511.05121)**	
11/28	Med Records/Clinical data	Serafim, James	DeepCare: A Deep Dynamic Memory Model for Predictive Medicine	**http://link.springer.com/chapter/10.1007%2F978-3-319-31750-2_3 (http://link.springer.com/chapter/10.1007%2F978-3-319-31750-2_3)** **http://arxiv.org/pdf/1608.02158v1.pdf (http://arxiv.org/pdf/1608.02158v1.pdf)**	

			Deep Survival analysis		
11/30	Medical Imaging	Anshul, James	DeepCyTOF: Automated Cell Classification of Mass Cytometry Data by Deep Learning and Domain Adaptation Microscopy cell counting and detection with fully convolutional regression networks	http://biorxiv.org/content/early/2016/06/14/054411, (http://biorxiv.org/content/early/2016/06/14/054411,) http://www.tandfonline.com/doi/full/10.1080/21681163.2016.1149104	
12/5	Medical Imaging	Anshul	Deep Learning for Identifying Metastatic Breast Cancer Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation	https://arxiv.org/abs/1606.05718 (https://arxiv.org/abs/1606.05718) http://arxiv.org/abs/1603.05959 (http://arxiv.org/abs/1603.05959)	
12/7	Wrap up	TAs			
Exam period	Project presentations	Anshul, James			

Course Summary:

Date	Details	
Mon Nov 14, 2016	CS273B quiz 1 (https://canvas.stanford.edu/courses/51037/assignments/47343)	due by 4:30pm
Mon Nov 28, 2016	CS273B Quiz 2 (Proteins, pharmacogenomics) (https://canvas.stanford.edu/courses/51037/assignments/49295)	due by 11:59pm
Wed Dec 7, 2016	CS273B Quiz 3 (Clinical Data and Medical Imaging) (https://canvas.stanford.edu/courses/51037/assignments/49297)	due by 3:15pm
Mon Dec 12, 2016	Final presentation slides (https://canvas.stanford.edu/courses/51037/assignments/47444)	due by 1pm
Fri Dec 16, 2016	Final report (https://canvas.stanford.edu/courses/51037/assignments/47446)	due by 12pm
	Oral presentation (journal club) (https://canvas.stanford.edu/courses/51037/assignments/49294)	
	Project Lightning Talks (https://canvas.stanford.edu/courses/51037/assignments/49302)	
	Project Proposal (https://canvas.stanford.edu/courses/51037/assignments/49301)	
	Written Review 1 (https://canvas.stanford.edu/courses/51037/assignments/49299)	
	Written Review 2 (https://canvas.stanford.edu/courses/51037/assignments/49300)	