Syllabus
Neural Networks and Deep Learning
CSCI 7222
Spring 2015

W 10:00-12:30
Muenzinger D430

Instructor

Professor Michael Mozer
Department of Computer Science
Engineering Center Office Tower 741
(303) 492-4103
Office Hours: W 13:00-14:00

Course Objectives

Neural networks have enjoyed several waves of popularity over the past half century. Each time they become popular, they promise to provide a general purpose artificial intelligence--a computer that can learn to do any task that you could program it to do. The first wave of popularity, in the late 1950s, was crushed by theoreticians who proved serious limitations to the techniques of the time. These limitations were overcome by advances that allowed neural networks to discover distributed representations, leading to another wave of enthusiasm in the late 1980s. The second wave died out as more elegant, mathematically principled algorithms were developed (e.g., support-vector machines, Bayesian models). Around 2010, neural nets had a third resurgence. What happened over the past 20 years? Basically, computers got much faster and data sets got much larger, and the algorithms from the 1980s--with a few critical tweaks and improvements--appear to once again be state of the art, consistently winning competitions in computer vision, speech recognition, and natural language processing. Below is a comic strip circa 1990, when neural nets reached public awareness. You might expect to see the same comic today, touting neural nets as the hot new thing, except that now the field has been rechristened *deep learning* to emphasize the architecture of neural nets that leads to discovery of task-relevant representations.

In this course, we'll examine the history of neural networks and state-of-the-art approaches to deep learning. Students will learn to design neural network architectures and training procedures via hands-on assignments. Students will read current research articles to appreciate state-of-the-art approaches as well as to question some of the hype that comes with the resurgence of popularity. We will use Geoff Hinton's Coursera lectures as background, since nobody in the field can explain ideas as well as Geoff, and class time will be devoted to discussing the lectures and delving into more detail about the methods.

Prerequisites

The course is open to any students who have some background in cognitive science or artificial intelligence and who have taken introductory probability/statistics and linear algebra.

Course Readings

We will rely primarily on current research articles, since -- following suitable introductory lectures -- the articles are pretty easy to follow. If you want additional reading, I recommend the following:

- Chris Bishop's Pattern Recognition and Machine Learning
- Deng & Yu's monograph on Deep Learning: Methods and Applications
- wikipedia is often a useful resource.

The research articles we'll cover in class are contained in links below on the class-by-class syllabus.

Course Discussions

We will use Piazza for class discussion. Rather than emailing me, I encourage you to post your questions on Piazza. The Piazza signup page is here. Once you've signed up, the class page is here.

Course Requirements

Readings

In the style of graduate seminars, I will expect you to have read required readings prior to class and to watch required videos prior to class. (At present, we'll do most of the videos in class, but that plan may change.) Come prepared to class to discuss the material (asking clarification questions, working through the math, relating papers to each other, critiquing the papers, presenting original ideas related to the paper).

Homework Assignments

We can all delude ourselves into believing we understand some math or algorithm by reading, but implementing and experimenting with the algorithm is both fun and valuable for obtaining a true understanding. Students will implement small-scale versions of as many of the models we discuss as possible. I will give about half a dozen homework assignments that involve implementation over the semester, details to be determined. My preference is for you to work in matlab, both because you can leverage existing software and because matlab has become the de facto work horse in machine learning. One or more of the assignments may involve writing a commentary on a research article or presenting the article to the class.

Semester Grades

Semester grades will be based 20% on class attendance and participation and 80% on the homework assignments. I will weight the assignments in proportion to their difficulty, in the range of 10-20% of the course grade. Students with backgrounds in the area and specific expertise may wish to do in-class presentations for extra credit.

Class-By-Class Plan and Course Readings

When you see a "<" beside a video, it's a video you should watch before class. When you see a ">", it's a video you should watch after class. Other videos we'll watch in class.

Date	Activity	Hinton Videos	Readings	Lecture Notes	Assignments
Jan 14	• history • Perceptrons (classification) • linear models (regression) • Hebbian learning • LMS	• What are neural networks? • Some simple models of neurons • A simple example of learning • Types of neural network architectures • Perceptrons • A geometrical view of perceptrons • >Why the learning works • >Learning the weights of a linear neuron • >The error surface for a linear neuron	Bengio, Learning deep architectures for AI (section 1) Chronicle of Higher Education article on Deep Learning	introduction.pptx learning1.pptx	assignment 1
Jan 21	• activation functions • error functions • back propagation • local and distributed representations	• <What perceptrons can't do • <Learning the weights of a logistic output neuron • <The backpropagation algoritm • >Learning to predict the next word • >Using the derivatives computed by backpropagation		learning2.pptx	assignment 2
Jan 28	• Diversion: Catrin Mills on modeling climate change	• <Overview of mini-batch gradient descent • >A bag of tricks for mini-batch gradient descent • >The momentum method • >Adaptive learning rates		Catrin's climate change introduction	assignment 3 handed out

	practical advice	for each connection >Rmsprop: Divide the gradient by a running average of its recent magnitude		practical_advice.pptx	
Feb 4	tricks of the trade	The softmax output function	bias-variance trade off pruning algorithms	tricks1.pptx Homa's slides on cyberbullying	
Feb 11	deep learning	<Recent developments in deep neural networks (esp. min 26-40) <Overview of ways to improve generalization Limiting the size of the weights Using noise as a regularizer The ups and downs of back propagation Dropout >Introduction to the full Bayesian approach >The Bayesian interpretation of weight decay >MacKay's quick and dirty method of setting weight costs	Understanding the difficulty of training deep feedforward neural networks (2010) Bengio, Learning deep architectures for AI (sections 2-4) Dropout Dropout2	tricks2.pptx	assignment 3 due; assignment 4 handed out
Feb 18	recurrent networks	<Modeling sequences: A brief overview <Training RNs with back propagation <A toy example of training an RNN Why is it difficult to train an RNN? Long short-term memory Echo state networks >Hessian free optimization >Learning to predict the next character		recurrent_nets.pptx	
Feb 25	Probabilistic neural nets Boltzmann machines RBMs sigmoid belief nets	<Hopfield nets <Using stochastic units to improve search How a Boltzmann machine models data >Boltzmann machine learning Restricted Boltzmann Machines An example of	Bengio, Learning deep architectures for AI (sections 5-8)	stochastic_nets.pptx	

	▪ Generative models	RBM learning ▪ >Stacking RBMs ▪ [RBMs for collaborative filtering] ▪ [Belief nets, Learning sigmoid belief nets, The wake-sleep algorithm]			
Mar 4	▪ Gregory Petropoulos: renormalization groups and deep learning ▪ unsupervised learning ▪ autoencoders	▪ <From PCA to autoencoders ▪ <Deep autoencoders ▪ <Deep autoencoders for document retrieval ▪ Semantic hashing ▪ Learning binary codes for image retrieval ▪ >Shallow autoencoders for pretraining	▪ Why does unsupervised pretraining help deep learning? ▪ Building high-level features using large scale unsupervised learning (2012) ▪ An exact mapping between the variational renormalization group and deep learning	Gregory's slides on renormalization groups unsupervised.pptx	assignment 4 due assignment 5
Mar 11	Application domains: object recognition	▪ <Why object recognition is difficult ▪ <Achieving viewpoint invariance ▪ Convolutional nets for digit recognition ▪ Convolutional nets for object recognition ▪ Le Cun Demo circa 1993	▪ Imagenet classification with deep conv. NN (2012) ▪ Manjunath: Visualizing and understanding convolutional neural networks (2013) ▪ fully convolutional nets for semantic segmentation	▪ object_recognition.pptx ▪ convolutional net demos	
Mar 18	Application domains: language Eliana Colunga on concept/word learning	▪ Neuro-probabilistic language models ▪ Dealing with the large number of possible outputs	▪ Neural probabilistic language model ▪ Domain adaptation for large-scale sentiment classification ▪ Sequence to sequence learning with NN	language.pptx	
Apr 1	Application domains: speech recognition		▪ Ridgeway: DNNs for acoustic modeling in speech recognition (2012) ▪ Improving DNN acoustic models using generalized maxout networks ▪ Maxout networks ▪ DeepSpeech	▪ speech.pptx ▪ Ridgeway.pdf	
			▪ Beckage: Deep visual-semantic alignments for generating image		

			descriptions (2014) ▪ Sukumar: Show and tell: A neural image caption generator (2014) ▪ Deep captioning with multimodal recurrent neural nets ▪ arXiv:1411.5654 ▪ arXiv:1411.4952 ▪ arXiv:1411.2539 ▪ arXiv:1411.4389		
Apr 8	Captioning images	▪ Learning images and captions		captions.pptx	
Apr 15	Odds and ends		▪ Khajah/Larsen: Practical Bayesian optimization of machine learning algorithms ▪ Hughes: Human level control through deep reinforcement learning ▪ Poursabzi: Curriculum learning ▪ Bao: Do deep nets need to be deep?		
Apr 22	Rich Caruana visit	▪ Hinton "dark knowledge" slides			
Apr 29	Limitations of deep learning	▪ The fog of progress ▪ Hinton reddit	▪ Kim+Milroy: Intriguing properties of neural networks ▪ Coy+Israelsen: Deep neural nets are easily fooled: High confidence predictions for unrecognizable images ▪ Explaining and harnessing adversarial examples ▪ Fejes: Measuring invariances in deep networks (2009)		

Other Interesting Papers

Alternative activation functions

- Discriminative learning of sum-product networks (2012)
- Delving deep into rectifiers: Surpassing human-level performance on ImageNet

Image processing

- Deep convolutional inverse graphics net
- Generative models of natural videos

Alternative training procedures

- Deeply supervised networks
- Understanding the difficulty of training deep feedforward neural networks

Relevant Links

Papers

See list at http://ufldl.stanford.edu/wiki/index.php/UFLDL_Recommended_Readings
See list at http://www.cs.toronto.edu/~hinton/deeprefs.html
See list at http://deeplearning.net/reading-list/

Popular Press

The Godfather of Artificial Intelligence 4/2015

Tutorials

See list at deeplearning.net
Mostly unsupervised methods
Bengio tutorial at AAAI 2013

Modeling tools

See list at http://deeplearning.net/software_links/
Torch7 -- looks to be pretty solid; requires learning matlab-like language (documentation)
Caffe -- rapidly evolving, but not terribly well documented; requires GPU
Theano -- general purpose but learning curve may be steep (documentation)
deep learning exercises -- code for Stanford deep learning tutorial, includes convolutional nets
convnet.js -- not the fastest, but may be the easiest
Matlab toolboxes for convolutional nets: matconvnet cnn cuda-cnn
Mocha -- deep learning framework for Julia

Additional information for students (click to read)