COMS 4776: Neural Networks and Deep Learning

Lectures:	Tuesday, Thursday 2:40-3:55
Lecture Room:	Mudd 833
Instructor:	Richard Zemel
Office hours:	Tuesday 4:00-5:00 CEPSR 619, and by appointment
Teaching Assistants:	Nghi Minh Dao, Ben Eyre, Amogh Imandar, Sidharth Sharma, Saivignesh Venkatraman, and Anubha Vyasamudri

Overview

It is very hard to hand design programs to solve many real world problems, e.g., distinguishing images of cats versus dogs. Machine learning algorithms allow computers to learn from example data, and produce a program that does the job. Neural networks are a class of machine learning algorithm originally inspired by the brain, but which have recently have seen a lot of success at practical applications. They're at the heart of large language and multimodal models, and production systems at all kinds of companies. This course gives an overview of both the foundational ideas and the recent advances in neural net algorithms.

Pre-requisites

This is a second course in machine learning, so it has some substantial prerequisites. Required courses: Machine Learning; Multivariable Calculus; Linear Algebra; Probability & Statistics. These prerequisites will not be enforced, but without them the course will be extremely challenging.

Readings

There is no required textbook for the class. A few small readings may be assigned if the need arises. These required readings will all be available on the web, for free.

There are also some relevant resources which are freely available online. We will try to provide links on a lecture-by-lecture basis.

- *Deep Learning: Foundations and Concepts*, a new textbook by Chris Bishop. `http://www.bishopbook.com/`

- *Deep Learning*, a textbook by Yoshua Bengio, Ian Goodfellow, and Aaron Courville. http://www.deeplearningbook.org/
- Andrej Karpathy's lecture notes on convolutional networks. These are very readable and cover the material in roughly the first half of the course. http://cs231n.github.io/
- Richard Socher's lecture notes, focusing on RNNs. http://cs224d.stanford.edu/syllabus.html
- Metacademy, an online website which helps you construct personalized learning plans and which has links to lots of resources relevant to particular concepts. We'll post links to relevant Metacademy concepts as the course progresses. http://www.metacademy.org
- Video lectures for Hugo Larochelle's neural networks course. http://info.usherbrooke.ca/hlarochelle/neuralnetworks/content.html
- David MacKay's excellent textbook, *Information Theory, Inference, and Learning Algorithms.* This isn't focused on neural nets per se, but it has some overlap with this course, especially the lectures on Bayesian models. http://www.inference.phy.cam.ac.uk/mackay/itila/
- *Neural Networks and Deep Learning*, a book by physicist Michael Nielsen which covers the basics of neural nets and backpropagation. http://neuralnetworksanddeeplearning.com/

Course requirements and grading

The format of the class will be lecture, with some discussion. I strongly encourage interaction and questions. There will also be tutorials during some lectures.

The grading in the class will be divided up as follows:

2 Assignments (Programming & Written)	40%
2 Quizzes	40%
Project	20%

CLASS SCHEDULE

Shown below are the topics for lectures and tutorials, and quizzes, and associated dates. These are subject to change. The notes will be available on the class website the day of the class meeting.

Dates	Topic
9/2, 9/4	Introduction & Single-Layer Models
9/9, 9/11	Multilayer Perceptrons & Distributed Representations
9/16	Backpropagation
9/18, 9/23	Optimization
9/25, 9/30	Convolutional Neural Networks
10/2	Image Classification
10/7	Generalization
10/9	Recurrent Networks
10/14	Attention
10/16	Quiz 1
10/21, 10/23, 10/28	Transformers & Autoregressive Models
10/30	Generative Adversarial Networks
11/6, 11/11	Autoencoders & VAEs
11/13	Diffusion
11/18	Interpretability
11/20	Model Compression
11/25	Quiz 2
12/2	Hopfield Networks & Boltzmann Machines
12/4	AI Safety