Modern Deep Learning

2016-2017

Instructor:	Mark Hamilton	**Time:**	F 2:30 – 3:30
Email:	mhamilton723@gmail.com	**Place:**	Microsoft NERD, Cambridge MA

Description

Deep learning is a vibrant and rich field that brings together many areas of mathematics to solve problems previously deemed impossible. Studying this material will give you several complementary perspectives on intelligence, information, and cognition through a mathematical lens. In this course we will investigate a wide breadth of topics from the fields of deep learning, statistics, machine learning and mathematics in general. We will begin at the humble linear regression, and work our way into architectures for solving problems in a variety of real world domains, speech, vision, text, etc. We will push back against the naiive notion that machine learning is all about regression or classification and actively explore other domains like unsupervised, semi-supervised, and reinforcement learning.

The course assumes a basic grounding in math and linear algebra. this course puts a strong focus on the intuition behind these tools, and I would much rather give students a deep intuition than muck through a proof on the board. The course optimized for content delivery, and the lectures will be fast paced. However, I aim to include a real application in each week's lectures to keep the material relevant to a broad group. The goal is to bring those with little to no knowledge of machine learning to a level where they can actively contribute to modern research efforts.

Course Pages:

1. https://microsoft.sharepoint.com/teams/NeuralNetworkDiscussionGroup

Syllabus with Resources:

This is brief description of the lectures given. Videos can be found on the course page and are currently Microsoft internal. If you would like to help edit and improve the quality of these videos please do not hesitate to contact.

Semester 1

Week 1: From Linear Regression to Auto-Encoders

- Gradient Descent
- RMSE loss function
- Linear regression
- Multivariable regression
- Matrix regression
- Multi-Layer Perceptrons
- Auto-encoders
  http://www.iro.umontreal.ca/~vincentp/Publications/denoising_autoencoders_tr1316.pdf
- Brief note on universality and no free lunch
  https://en.wikipedia.org/wiki/Universal_approximation_theorem
  https://en.wikipedia.org/wiki/No_free_lunch_theorem
  http://www.zabaras.com/Courses/BayesianComputing/Papers/lack_of_a_priori_distinctions_wolpert.pdf
- Brief into to recurrent nets
  Research Topic: Regularized recurrent auto-encoders

Week 2: Backprop and Deep Dream

- Forward + Reverse mode differentiation
  http://colah.github.io/posts/2015-08-Backprop/
- Neural Network Inversion
  http://www.sciencedirect.com/science/article/pii/S0893608006001730
  Sorry this isn't free
- Deep Dream for visualizing platonic forms
  https://www.robots.ox.ac.uk/~vedaldi/assets/pubs/mahendran15understanding.pdf
- Total Variation/ Naturalness priors
  https://www.robots.ox.ac.uk/~vedaldi/assets/pubs/mahendran15understanding.pdf
- Deep Dream: maximizing layer activations
  https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Week 3: Convolutional Nets

- Image Filters
- Convolutional Layers
  http://colah.github.io/posts/2014-07-Conv-Nets-Modular/
  http://cs231n.github.io/convolutional-networks/
- Max Pooling Layers
- Walkthrough of all of the steps from "a neural algorithm for artistic style"
  https://arxiv.org/abs/1508.06576
  Reconstruction Loss: Content Loss
  Texture Reconstruction: Style Loss

Week 4:Distribution Metrics and Generative Adversarial Nets

- Generative Models
- Images of random variables
- Earth Mover Distance
  https://en.wikipedia.org/wiki/Earth_mover%27s_distance
  http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/RUBNER/emd.htm
- Maximum Mean Discrepancy Distance
  http://alex.smola.org/teaching/iconip2006/iconip_3.pdf
- Generative Adversarial Nets
  https://arxiv.org/pdf/1406.2661v1.pdf
  https://github.com/Newmu/dcgan_code
- Minimax games
- Convergence of GAN
- Brief Description of Information Regularized GANs
  https://arxiv.org/abs/1606.03657
- Brief Description of Disentangled Code Spaces
- Brief Discussion of research topics
  4d convolutions, brain decoders, and multi-task learning
  Message marhamilmicrosof.com to collaborate

Week 5: Information Theory and infoGAN

- Information Theory
  https://colah.github.io/posts/2015-09-Visual-Information/
- Prob+Stats review
- Coding Theory
- Entropy + Differential Entropy

- Brief Into to information geometry and the fisher information metric
  https://en.wikipedia.org/wiki/Information_geometry
  http://www.cs.cmu.edu/~lebanon/pub/thesis/thesis-2x1.pdf
- Cross Entropy, KL-Divergence
- Conditional, joint, and Mutual Entropy
- Information Regularized GAN
  https://arxiv.org/abs/1606.03657

Semester 2

Week 6: Recurrent Nets and Language Models

- Brief review of neural nets
- Turing Completeness
- Recurrent Nets
- Latent information channels
  http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- One-Hot encoding
- Embedding layers
- Sparse Matrix lookups
- Interpreting embeddings
  http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/
- The softmax layer
  http://cs231n.github.io/linear-classify/#softmax-classifier
- Language modelling
  http://sebastianruder.com/word-embeddings-1/index.html

Week 7: Modern Word Embedding

- Brief review of recurrent nets
- Sequence to sequence translation
  https://www.tensorflow.org/tutorials/seq2seq/
  https://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf
- Brief mention of bidirectional RNNS
  https://en.wikipedia.org/wiki/Bidirectional_recurrent_neural_networks
- Feed Forward Language Models
- Bilinear Language Models (Word2vec)
- Skipgram
- CBOW
  http://sebastianruder.com/word-embeddings-1/index.html
  https://arxiv.org/abs/1301.3781
- Gradient of the softmax
  http://sebastianruder.com/word-embeddings-softmax/index.html
- Monte-Carlo Integration
  https://en.wikipedia.org/wiki/Monte_Carlo_integration
- Importance sampling
  http://sebastianruder.com/word-embeddings-softmax/index.html

Week 8: Word2Vec and Stochastic Matrix Factorization

- Review of Skipgram
  http://sebastianruder.com/word-embeddings-1/index.html
  https://arxiv.org/abs/1301.3781
- Negative Sampling
  http://sebastianruder.com/word-embeddings-softmax/index.html
  http://demo.clab.cs.cmu.edu/cdyer/nce_notes.pdf

- Negative Sampling as PMI factorization
  http://sebastianruder.com/secret-word2vec/
  https://papers.nips.cc/paper/5477-neural-word-embedding-as-implicit-matrix-factorization.pdf
- Matrix Factorization for Recommender Systems
- Netflix Challenge
  https://datajobs.com/data-science-repo/Recommender-Systems-%5BNetflix%5D.pdf
  https://www.youtube.com/watch?v=gCaOa3W9kM0&t=2s
- Matrix Factorization for Auto-ML
- Matrix Factorization for Clustering
- Matrix Factorization Grammars
  https://www.cs.toronto.edu/~rgrosse/uai2012-matrix.pdf

Week 9: Intro to Reinforcement Learning

- Review of Supervised Learning
- The Reinforcement Learning Paradigm
- Markov Decision Processes
  https://en.wikipedia.org/wiki/Markov_decision_process
- Policy Gradient Methods:
- Playing Atari Games
  http://karpathy.github.io/2016/05/31/rl/
- Experience Replay
- Advantage/Reward Functions
- Epsilon Greedy Exploration and Annealing
- Derivation using score expectations
- Policy Gradient as a Nondifferentiable Component
  https://arxiv.org/pdf/1506.05254v1.pdf

Week 10: Deep Q-Learning

- Review of Supervised learning
- Review of Policy Gradient Learning
- Markov Decision Processes
  https://en.wikipedia.org/wiki/Markov_decision_process
- Optimal Policies and Q-Functions
  https://www.nervanasys.com/demystifying-deep-reinforcement-learning/
- Table based Q-Learning:
  https://medium.com/emergent-future/simple-reinforcement-learning-with-tensorflow-part-0-.qimruxaea
- Brief note on convergence:
  http://users.isr.ist.utl.pt/~mtjspaan/readingGroup/ProofQlearning.pdf
- Deep-Q Learning
  https://www.nervanasys.com/demystifying-deep-reinforcement-learning/
  http://www.nature.com/nature/journal/v518/n7540/full/nature14236.html
- Brief note on Experience replay
- Brief note on epsilon greedy exploration
- Actor Critic Method for Continuous Control:
- Unifying Policy Gradient + Deep Q learning
  https://arxiv.org/pdf/1509.02971.pdf
- Briefly Mentioned AlphaGo:
  https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf

Extra Material for Week 10

- Great Video About Most of Deep Reinforcement Learning:
  http://videolectures.net/rldm2015_silver_reinforcement_learning/
- Double Q learning:
  https://arxiv.org/pdf/1509.06461v3.pdf
- Prioritized Experience Replay:
  https://arxiv.org/pdf/1511.05952v4.pdf

Week 11: AlphaGo and LSTMS

- Sketching Out The Alpha Go Architecture:
  https://storage.googleapis.com/deepmind-media/alphago/AlphaGoNaturePaper.pdf
- Deep Supervised Policy
- Rollout Supervised Policy
- Deep Policy Gradient Policy
- Value Network Trained with Policy Gradient Games
- Monte Carlo Tree Search
- Brief review of RNNs
- Vanishing and Exploding Gradients
  https://en.wikipedia.org/wiki/Vanishing_gradient_problem
  http://jmlr.csail.mit.edu/proceedings/papers/v28/pascanu13.pdf
- LSTMs:
  http://colah.github.io/posts/2015-08-Understanding-LSTMs/
- Forget, Add and Output Gates
- Applications to Text
  http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- Applications to Handwriting
  https://www.cs.toronto.edu/~graves/handwriting.html
  https://arxiv.org/abs/1308.0850

Week 12: Neural Turing Machines

- https://arxiv.org/abs/1410.5401
- Overview or LSTM and RNN memory capabilities
- Discussion of Memory Addressing
- Location Based Addressing
- Content Based Addressing
- Differentiable Memory Reads
- Differentiable Writes
- Differentiable Addressing
- Controller Networks
- Experimental Exploration
- Examining the Activations and weights

Week 13: Optimization Methods

Week 14: Image Processing with Danil Kirsanov