Reading lists for new MILA students

Research in General

How to write a great research paper

Basics of machine learning

- DL book chapter on linear algebra: http://www.iro.umontreal.ca/~bengioy/DLbook/linear_algebra.html
- DL book chapter on probability: http://www.iro.umontreal.ca/~bengioy/dlbook/prob.html
- DL book chapter on numerical computation: http://www.iro.umontreal.ca/~bengioy/dlbook/numerical.html
- DL book chapter on machine learning: http://www.iro.umontreal.ca/~bengioy/DLbook/ml.html

Basics of deep learning

- Intro to deep learning: http://www.iro.umontreal.ca/~bengioy/DLbook/intro.html
- Feedforward multi-layer nets: http://www.iro.umontreal.ca/~bengioy/DLbook/mlp.html
-
- Learning deep architectures for AI
- Practical recommendations for gradient-based training of deep architectures
- Quick'n'dirty introduction to deep learning: Advances in Deep Learning
- A fast learning algorithm for deep belief nets
- Greedy Layer-Wise Training of Deep Networks
- Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion
- Contractive auto-encoders: Explicit invariance during feature extraction
- Why does unsupervised pre-training help deep learning?
- An Analysis of Single Layer Networks in Unsupervised Feature Learning
- The importance of Encoding Versus Training With Sparse Coding and Vector Quantization
- Representation Learning: A Review and New Perspectives

- Deep Learning of Representations: Looking Forward
- Measuring Invariances in Deep Networks
- Neural networks course at USherbrooke [youtube]

Feedforward nets

- http://www.iro.umontreal.ca/~bengioy/DLbook/mlp.html
- "Improving Neural Nets with Dropout" by Nitish Srivastava
- Batch Normalization
- "Fast Drop Out"
- "Deep Sparse Rectifier Neural Networks"
- "What is the best multi-stage architecture for object recognition?"
- "Maxout Networks"

MCMC

- Iain Murray's MLSS slides
- Radford Neal's Review Paper (old but still very comprehensive)
- Better Mixing via Deep Representations
- Bayesian Learning via Stochastic Gradient Langevin Dynamics

Restricted Boltzmann Machines

- Unsupervised learning of distributions of binary vectors using 2-layer networks
- A practical guide to training restricted Boltzmann machines
- Training restricted Boltzmann machines using approximations to the likelihood gradient
- Tempered Markov Chain Monte Carlo for training of Restricted Boltzmann Machine
- How to Center Binary Restricted Boltzmann Machines
- Enhanced Gradient for Training Restricted Boltzmann Machines
- Using fast weights to improve persistent contrastive divergence
- Training Products of Experts by Minimizing Contrastive Divergence

Boltzmann Machines

- Deep Boltzmann Machines (Salakhutdinov & Hinton)
- Multimodal Learning with Deep Boltzmann Machines
- Multi-Prediction Deep Boltzmann Machines
- A Two-stage Pretraining Algorithm for Deep Boltzmann Machines

Regularized Auto-Encoders

- The Manifold Tangent Classifier

- DL book chapter on unsupervised learning: http://www.iro.umontreal.ca/~bengioy/dlbook/unsupervised.html
- DL book chapter on manifolds: http://www.iro.umontreal.ca/~bengioy/dlbook/manifolds.html
- Representation Learning: A Review and New Perspectives, in particular section 7.

Regularization

Stochastic Nets & GSNs

- Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation
- Learning Stochastic Feedforward Neural Networks
- Generalized Denoising Auto-Encoders as Generative Models
- Deep Generative Stochastic Networks Trainable by Backprop

Others

- Slow, Decorrelated Features for Pretraining Complex Cell-like Networks
- What Regularized Auto-Encoders Learn from the Data Generating Distribution
- Generalized Denoising Auto-Encoders as Generative Models
- Why the logistic function?

Recurrent Nets

- DL book chapter on recurrent nets
- Learning long-term dependencies with gradient descent is difficult
- Advances in Optimizing Recurrent Networks
- Learning recurrent neural networks with Hessian-free optimization
- On the importance of momentum and initialization in deep learning,
- Long short-term memory (Hochreiter & Schmidhuber)
- Generating Sequences With Recurrent Neural Networks
- Long Short-Term Memory in Echo State Networks: Details of a Simulation Study
- The "echo state" approach to analysing and training recurrent neural networks
- Backpropagation-Decorrelation: online recurrent learning with O(N) complexity
- New results on recurrent network training:Unifying the algorithms and accelerating convergence
- Audio Chord Recognition with Recurrent Neural Networks
- Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription

Convolutional Nets

- DL book chapter on convolutional nts: http://www.iro.umontreal.ca/~bengioy/DLbook/convnets.html
- Generalization and Network Design Strategies (LeCun)
- ImageNet Classification with Deep Convolutional Neural Networks, Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton, NIPS 2012.
- On Random Weights and Unsupervised Feature Learning

Optimization issues with DL

- Curriculum Learning
- Evolving Culture vs Local Minima
- Knowledge Matters: Importance of Prior Information for Optimization
- Efficient Backprop
- Practical recommendations for gradient-based training of deep architectures
- Batch Normalization
- Natural Gradient Works Efficiently (Amari 1998)
- Hessian Free
- Natural Gradient (TONGA)
- Revisiting Natural Gradient

NLP + DL

- The first journal paper on neural language models (there was a NIPS'2000 paper before): A Neural Probabilistic Language Model
- Natural Language Processing (Almost) from Scratch
- DeViSE: A Deep Visual-Semantic Embedding Model
- Distributed Representations of Words and Phrases and their Compositionality
- Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection

CV+RBM

- Fields of Experts
- What makes a good model of natural images?
- Phone Recognition with the mean-covariance restricted Boltzmann machine
- Unsupervised Models of Images by Spike-and-Slab RBMs

CV + DL

- Imagenet classification with deep convolutional neural networks

- Learning to relate images

Scaling Up

- Large Scale Distributed Deep Networks
- Random search for hyper-parameter optimization
- Practical Bayesian Optimization of Machine Learning Algorithms

DL + Reinforcement learning

- Playing Atari with Deep Reinforcement Learning
- True Online TD Lambda

Graphical Models Background

- An Introduction to Graphical Models (Mike Jordan, brief course notes)
- A View of the EM Algorithm that Justifies Incremental, Sparse and Other Variants (Neal & Hinton, important paper to the modern understanding of Expectation-Maximization)
- A Unifying Review of Linear Gaussian Models (Roweis & Ghahramani, ties together PCA, factor analysis, hidden Markov models, Gaussian mixtures, k-means, linear dynamical systems)
- An Introduction to Variational Methods for Graphical Models (Jordan et al, mean-field, etc.)

Writing

- Writing a great research paper (video of the presentation)

Software documentation

- Python, Theano, Pylearn2, Linux (bash) (at least the 5 first sections), git (5 first sections), github/contributing to it (Theano doc), vim tutorial or emacs tutorial

Software lists of built-in commands/functions

- Bash commands
- List of Built-in Python Functions
- vim commands

Other Software stuff to know about:

- screen/tmux
- ssh
- ipython & ipython notebook (now Jupyter)
- matplotlib
- Caffe - caffe.berkeleyvision.org
- DIGITS - https://developer.nvidia.com/digits