Name of the subject: Deep Learning

Prerequisites: Machine Learning

Content:

Introduction (4 lectures)
Feedforward Neural networks. Gradient descent and the backpropagation algorithm. Unit saturation, aka the vanishing gradient problem, and ways to mitigate it. RelU Heuristics for avoiding bad local minima. Heuristics for faster training. Nestors accelerated gradient descent. Regularization. Dropout.

Convolutional Neural Networks (4 lectures)
Architectures, convolution / pooling layers

Recurrent Neural Networks (4 lectures)
LSTM, GRU, Encoder Decoder architectures

Deep Unsupervised Learning (4 lectures)
Autoencoders (standard, sparse, denoising, contractive, etc), Variational Autoencoders, Adversarial Generative Networks, Autoencoder and DBM

Attention and memory models, Dynamic memory networks (2 lectures)

Applications of Deep Learning to Computer Vision (4 lectures)
Image segmentation, object detection, automatic image captioning, Image generation with Generative adversarial networks, video to text with LSTM models. Attention models for computer vision tasks.

Applications of Deep Learning to NLP:
Introduction to NLP and Vector Space Model of Semantics **(2 lectures)**
Word Vector Representations: Continuous Skip-Gram Model, Continuous Bag-of-Words model (CBOW)**,** Glove, Evaluations and Applications in word similarity, analogy reasoning **(4 lectures)**
Named Entity Recognition, Opinion Mining using Recurrent Neural Networks **(3 lectures)**
Parsing and Sentiment Analysis using Recursive Neural Networks **(3 lectures)**
Sentence Classification using Convolutional Neural Networks **(1 lecture)**
Dialogue Generation with LSTMs **(1 lecture)**
Applications of Dynamic Memory Networks in NLP **(1 lecture)**
Recent Reseearch in NLP using Deep Learning: Factoid Question Asnwering, similar question detection, Dialogue topic tracking, Neural Summarization, Smart Reply **(4 lectures)**

Reference Books and Papers:

Bengio, Yoshua, Ian J. Goodfellow, and Aaron Courville. "Deep learning." An MIT Press book in preparation. (2015).

Bengio, Yoshua. "Learning deep architectures for AI." Foundations and trends in Machine Learning 2.1 (2009): 1127.
Hochreiter, Sepp, and Jargen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 17351780.
Oquab, Maxime, et al. "Learning and transferring midlevel image representations using convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
Bengio, Yoshua, et al. "A neural probabilistic language model." journal of machine learning research 3.Feb (2003).
Collobert, Ronan, et al. "Natural language processing (almost) from scratch." Journal of Machine Learning Research 12.Aug (2011): 2493-2537.
Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space." arXiv preprint arXiv:1301.3781 (2013).
Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global Vectors for Word Representation." EMNLP. Vol. 14. 2014.
Kim, Yoon. "Convolutional neural networks for sentence classification." EMNLP (2014).
Oquab, Maxime, et al. "Learning and transferring mid-level image representations using convolutional neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.
Kumar, Ankit, et al. "Ask me anything: Dynamic memory networks for natural language processing." arXiv preprint arXiv:1506.07285 (2015).
Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems. 2014.
Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom. "A convolutional neural network for modelling sentences." ACL (2014).
Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank." Proceedings of the conference on empirical methods in natural language processing (EMNLP). Vol. 1631. 2013.
Socher, Richard, et al. "Parsing with Compositional Vector Grammars." ACL. 2013.
Abadi, Martın, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." arXiv preprint arXiv:1603.04467 (2016)