Deep Learning Papers Review

1) Deep Learning (Yann LeCun, Yoshua Bengio, Geoffrey Hinton)

The paper is an introductory paper to Deep Learning authored by 3 of the big names in AI. It descibes how depp learning is better than machine learning and what are the problems solved by Deep Learning. The paper gives intro to to the concepts of Stochastic Gradient Descent, Backpropogation algorithm, CNNs, RNNs, LSTM. It also mentions about the concept of selectivity-invariance: “shallow classifiers require a good feature extractor that solves the selectivity–invariance dilemma — one that produces representations that are selective to the aspects of the image that are important for discrimination, but that are invariant to irrelevant aspects such as the pose of the animal.” They also describe about the future expectations, quoted “We expect much of the future progress in vision to come from systems that are trained end-to-end and combine ConvNets with RNNs that use reinforcement learning to decide where to look. Systems combining deep learning and rein-forcement learning are in their infancy, but they already outperform passive vision systemsat classification tasks and produce impressive results in learning to play many different video games.”

2) A Fast Learning Algorithm for Deep Belief Nets (Geoffrey Hinton, Simon Osindero, Yee-Whye Teh)

The paper describes about Deep Belief Networks. This paper show how to use “complementary priors” to eliminate the explaining-away effects that make inference difficult in densely connected belief nets that have many hidden layers. Quoting : “Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory. The fast, greedy algorithm is used to initialize a slower learning procedure that fine-tunes the weights using a contrastive version of the wake-sleep algorithm. After fine-tuning, a network with three hidden layers forms a very good generative model of the joint distribution of handwritten digit images and their labels.”
It describes about Restricted Boltzmann Machines and Contrastive Divergence Learning.

3) Reducing the Dimensionality of Data with Neural Networks (Geoffrey Hinton, R. R. Salakhutdinov)

Quote : “High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such ‘‘autoencoder’’ networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data." They compare their network (autoencoders) with PCA.
After pretraining multiple layers of feature detectors, the model is unfolded to produce encoder and decoder networks that initially use the same weights. The global finetuning stage then replaces stochastic activities by deterministic, real-valued probabilities and uses backpropagation through the whole auto-encoder to fine-tune the weights for optimal reconstruction. They stack multiple RBMs to form the network and then use backpropogation for fine tuning. Quote “The first layer of feature detectors then become the visible units for learning the next RBM.” Also, “The pixels correspond to visible units of the RBM because their states are observed; the feature detectors correspond to hidden units.”

4) ImageNet Classification with Deep Convolutional Neural Networks - AlexNet (Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton)

“We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax.”

How they improved : ReLu activation function (non saturating neurons), Training on multiple GPUs, overlapping pooling, Local Response Normalization. Fo overcoming overfitting, they used Data augmentation techniques and dropouts. Quote : “In our implementation, the transformed images are generated in Python code on the CPU while the GPU is training on the previous batch of images. So these data augmentation schemes are, in effect, computationally free.” The data augmentation techniques they used are 1) generating image translations and horizontal reflections, 2) altering the intensities of the RGB channels in training images.