Deep Learning Networks

- Deep Learning techniques address a number of these issues
- Discovery of re-used features, potential for more modular learning
- Learning from unlabelled data (followed by supervised learning)
- The ability to train deeper networks, and capture intermediate features

Machine Learning theory says we can learn any function with accuracy as close as we want with a single layer, so why bother? 2-layer MLPs and SVMs are “universal”

The right representation can be much more efficient for particular tasks

There is significant modularity in the brain- deep networks of re-used features are seen in vision, and are useful for audio and natural language tasks

A more promising approach for more general AI

http://www.technologyreview.com/featuredstory/513696/deep-learning
Deep Learning Networks

- Main approaches:
  - Autoencoder networks (unsupervised pre-training)
  - Restricted Boltzmann Machine networks (unsupervised pre-training)
  - Convolutional Neural Networks (sparse, deep topology)

Layer-wise Pre-Training

Autoencoder networks
Autoencoder networks

- Data is provided as input, and the output of the network tries to reconstruct the input
- Learning is performed using backpropagation or related methods
- The target output of the network is set to the input
- The aim of training is to minimise the error of reconstruction
- A reduced set of hidden units is used, creating an information bottleneck

Restricted Boltzmann Machine Networks

- Stable states of the network have low energy values
- Gibbs sampling is used to find a low energy state
- The energy values of configurations are defined by the weights
- Probabilistic: weights $\rightarrow$ energy values $\rightarrow$ probabilities

- RBMs are another technique for pre-training, to capture features of the input. They are recurrent networks, with a number of stable states
- Given an input, the network can be sampled. Activations are passed from the input to the hidden layer, then from hidden to the input layer, repeating until stability is reached.
- The visible layer provides a reconstruction of the input. Training the network allows capturing features of the input.
Alternating Gibbs Sampling

- The input is presented at the visible units.
- Hidden units are updated based on probabilistic activations. Visible units are updated subsequently. This repeats until stability is reached.

Pre-trained Deep Networks

Putting it all together:
- This approach can be used to pre-train a network, before performing supervised learning.
- A classification layer can be added to the top layer, representing classes for supervised learning. A fine-tuning stage adjusts weights using an error function defined at the output nodes, by backpropagation.

Deep Learning Networks

Convolutional Neural Networks

![Convolutional Neural Network Diagram]
Convolutional Neural Networks

- CNNs are a form of deep neural network with a specific topology, based on structure seen in the visual system.
- Each unit has a limited receptive field.
- Units are convolutional, the same set of weights are used to find a response in multiple positions.
- Convolutional and sub-sampling layers perform specific functions, acting in a manner similar to simple and complex cells in the visual system.

Application - Speech Recognition

- “Context-Dependent Deep Neural Network Hidden Markov Model”
- Hidden Markov Model - learn relationships of transitions between hidden states, and between hidden states and observations.

Summary

- Deep Learning approaches introduce a number of new techniques that allow an increase in depth and modularity of neural networks.
- Unsupervised learning allows capturing structure from observations, without relying on feedback from classifications.
- Unsupervised pre-training improves the reliability and accuracy of supervised learning.
- These techniques offer many new opportunities for machine learning and more general artificial intelligence.

Application - Speech Recognition

- The HMM system can be used to infer the phoneme symbols from audio samples, based on the relationships between acoustic patterns and symbols, and probabilities of symbol sequences.
- Hidden Markov Model - learn relationships of transitions between hidden states, and between hidden states and observations.
- The Deep Neural Network is used to learn the probability distribution of symbolic states from audio.
**Application - Speech Recognition**

- 7 layers, 2048 hidden units at each layer
- Trained on 309 hours of training data
- Each layer pre-trained as a Restricted Boltzmann Machine
- Fine-tuned using 9304 triphone states (output layer)
- Improvement from 27.4% to 18.5% error (30% improvement)
- Demonstration: [http://www.youtube.com/watch?v=Nu-nlQqFCKg](http://www.youtube.com/watch?v=Nu-nlQqFCKg)

**References**

- “ImageNet Classification with Deep Convolutional Neural Networks”, Krizhevsky et al 2012, [http://nips.cc/Conferences/2012/Program/event/66](http://nips.cc/Conferences/2012/Program/event/66)
References

*figure by Yoshua Bengio, Montreal. “Learning Deep Architectures for AI”
*figure by Andrew Ng, Stanford. “Sparse Autoencoder”
*figure from Zeiler & Fergus 2013