Deep Learning: Lecture 0

Keith L. Downing

The Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
keithd@idi.ntnu.no

January 12, 2020

Representations for AI Problem Solving

The Big Choice

1. Crafted by Human Engineers: Hard work for knowledge engineers, but the AI's decision making is normally straightforward to understand.
2. Formed by the AI System: Easy to setup (e.g. via SciKit Learn) but VERY difficult to interpret results, particularly with non-symbolic systems (e.g. neural nets)

Outline

1. A Brief History of Neural Networks
2. Linear Separability
3. Brief Overview of Backpropagation
4. A Few Example Applications

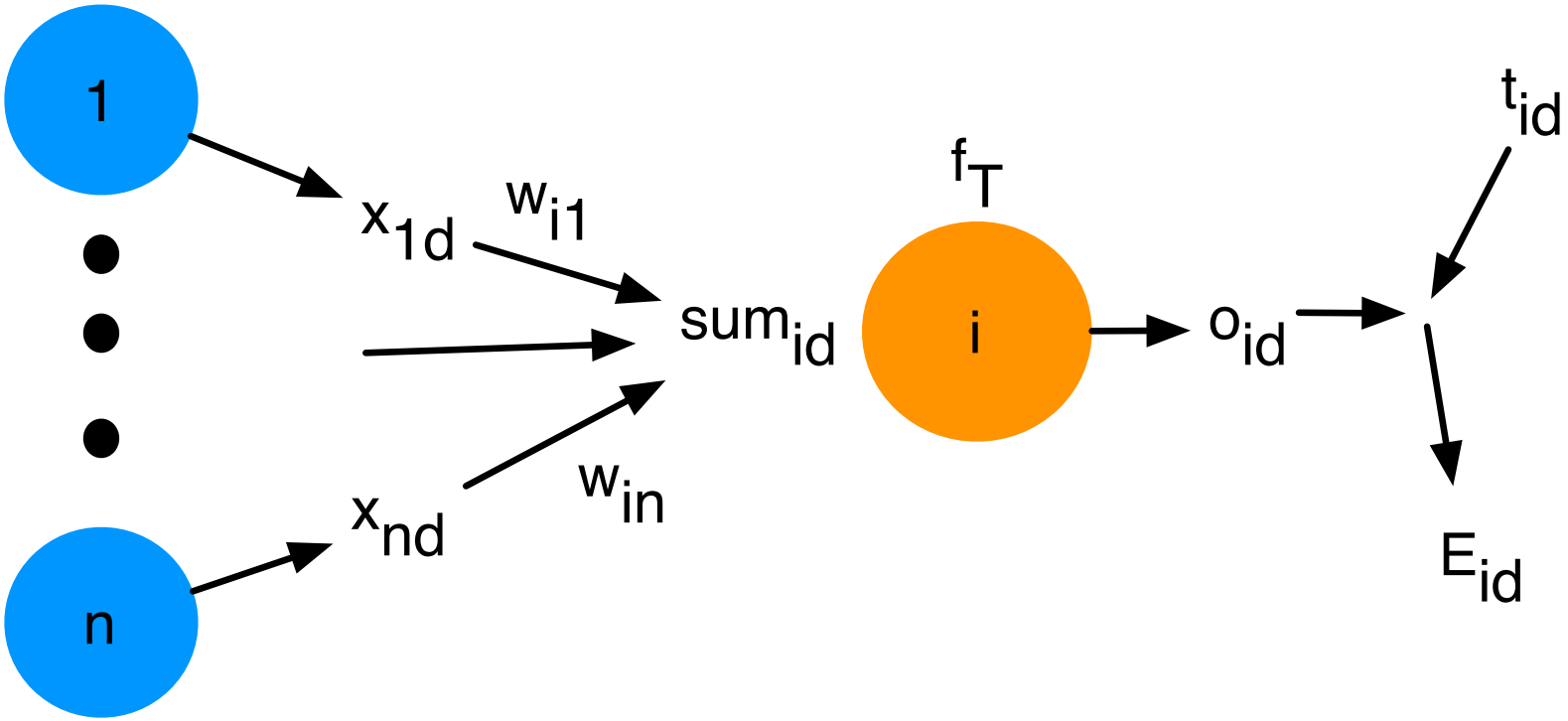
The Early History of Neural Networks

- McCulloch + Pitts (1943) - neuron model similar to logic gates: no weights and no learning, but special excitatory and inhibitory connections.
- Rosenblatt (1958) - The **perceptron**, a 3-layered network. Today, we call his output layer a perceptron, since connections between other two layers were not adaptive.
- Widrow + Hoff (1960) - **Adalines + Delta Rule** for training them, where error signal is based on the weighted sum of inputs, not the output of an activation function..
- Minsky + Papert (1969) - Proved that non-linearly-separable functions (e.g. XOR) could not be represented by a two-layered neural network (regardless of the type of neuron). Since a) Most hard data sets are not linearly separable, and b) Delta rule fails for nets with more than 2 layers $\rightarrow$ Nets for hard data sets cannot be trained!
- Near death of neural net research (1970-1985)

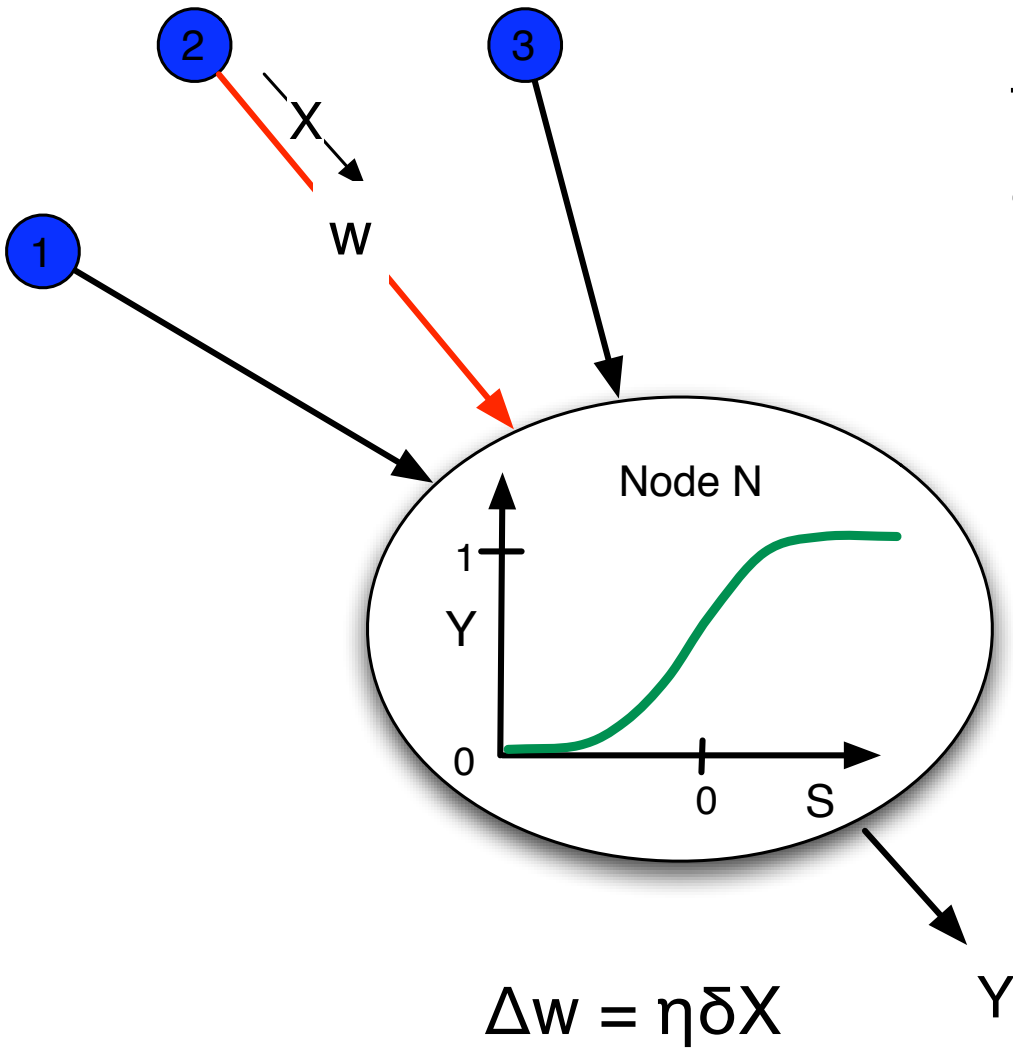
The Ressurection of Neural Networks

- Werbos (1982), Parker (1985), Rumelhart, Hinton + Williams (1986) - **Backpropagation** invented for training multi-layer networks. Replaced non-differentiable step function with the sigmoid.
- Ackley, Hinton, Sejnowski (1985) - **Boltzmann Machines** - Using probabilistic, binary neurons.
- Neural Net Explosion (1985-1995) - Diverse applications.
- Hibernation (1995-2005) - Trappings of local minima and failure of deep nets (due to attenuated backpropagation signals, i.e. gradients) became glaring weaknesses that prevented scaling up.
- The Deep Learning Revolution (2006-present) - Unsupervised pre-training (later found unecessary) + many small (but significant) changes/extensions to backpropagation + major hardware improvements + BIG DATA $\rightarrow$ Learning in nets with 100+ layers !!

The Perceptron



The Delta Rule



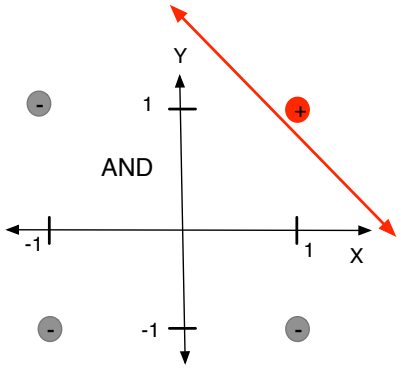
T = target output value

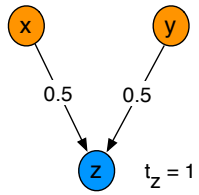
δ = error

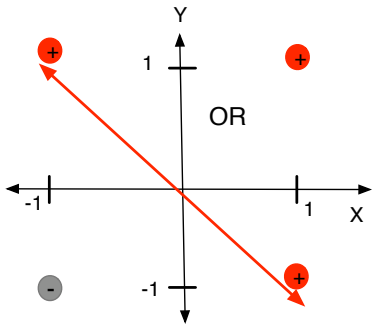
$\delta = T - Y$

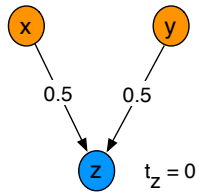
$\Delta w = \eta \delta X$

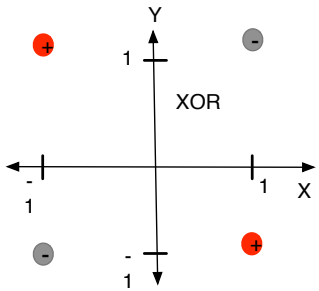
XOR: The (Near) Death of Neural Networks

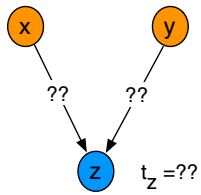




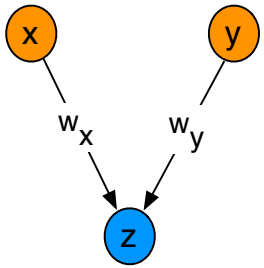


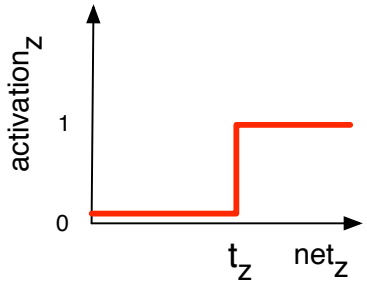


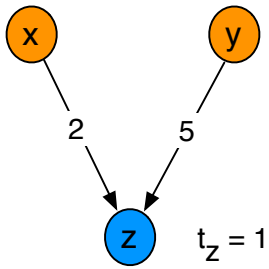


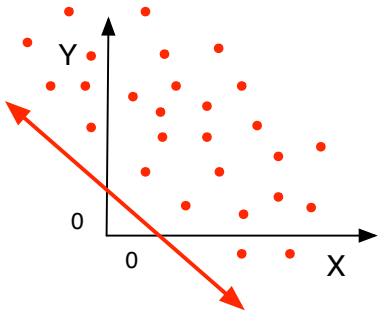


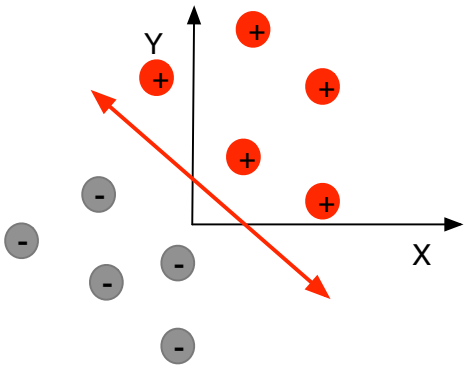
Linear Separability of Data





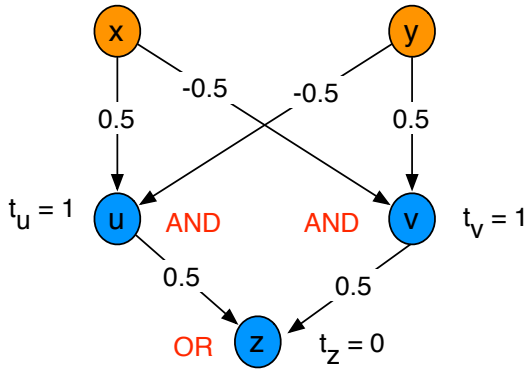


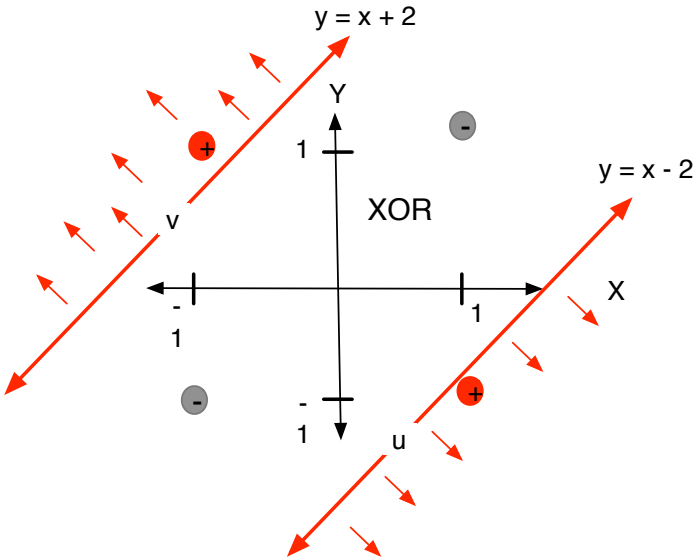




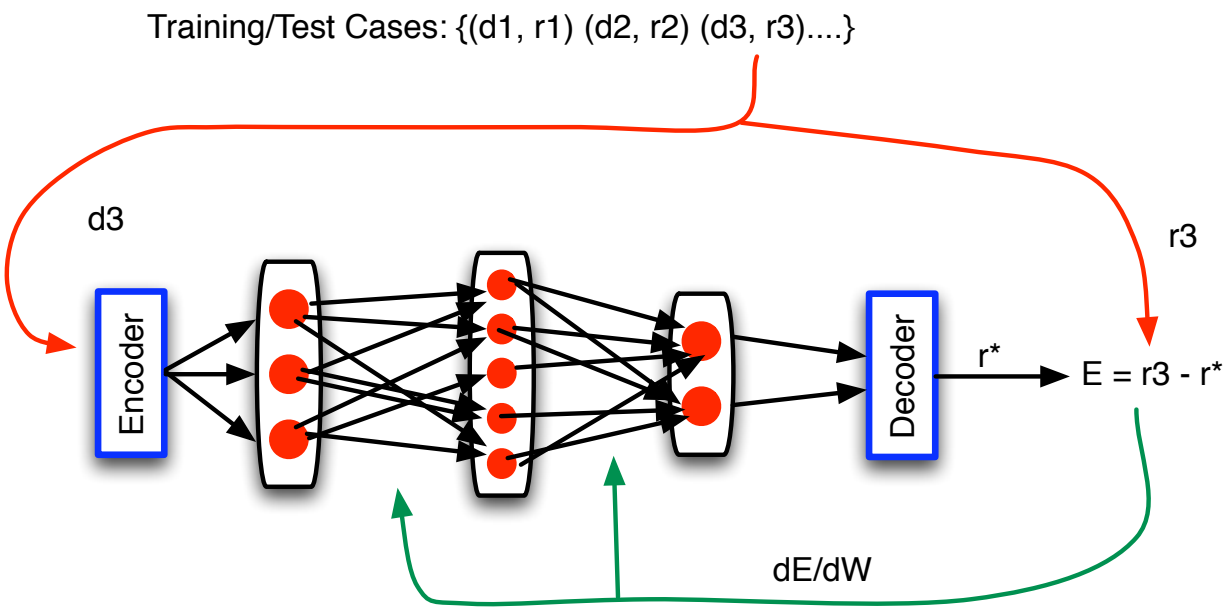
Adding a Hidden Layer

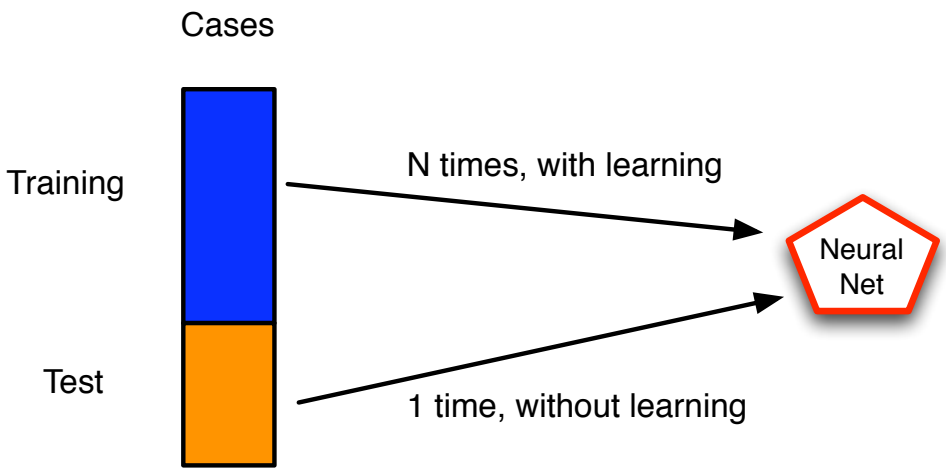
Not linearly separable $\rightarrow$ Need hidden layer with non-linear act func.



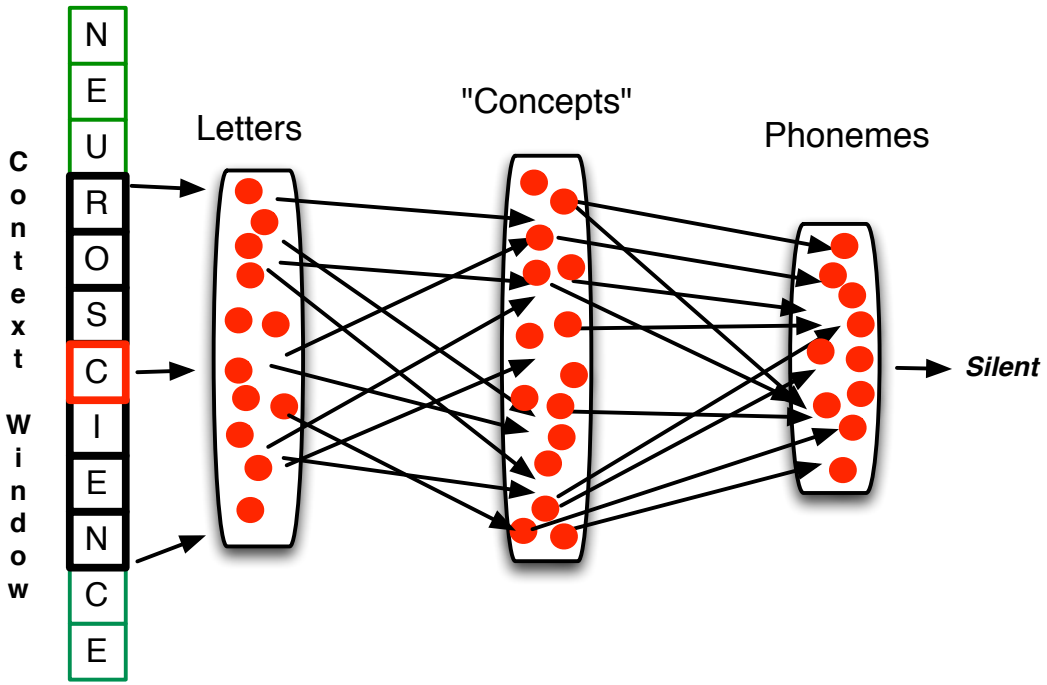


Backpropagation





NETtalk (Sejnowski + Rosenberg, 1986)



- IBM's DECtalk: **several man years** of work $\rightarrow$ Reading Machine.
- NETtalk: **10 hours** of backprop training on 1000-word text (T1000).
- 95% accuracy on T1000; 78% accuracy on novel text.
- Improvement during training sounds like a child learning to read.
- Concept layer is key: 79 different (overlapping) clouds of active neurons gradually form, with each mapping to one of the 79 phonemes.

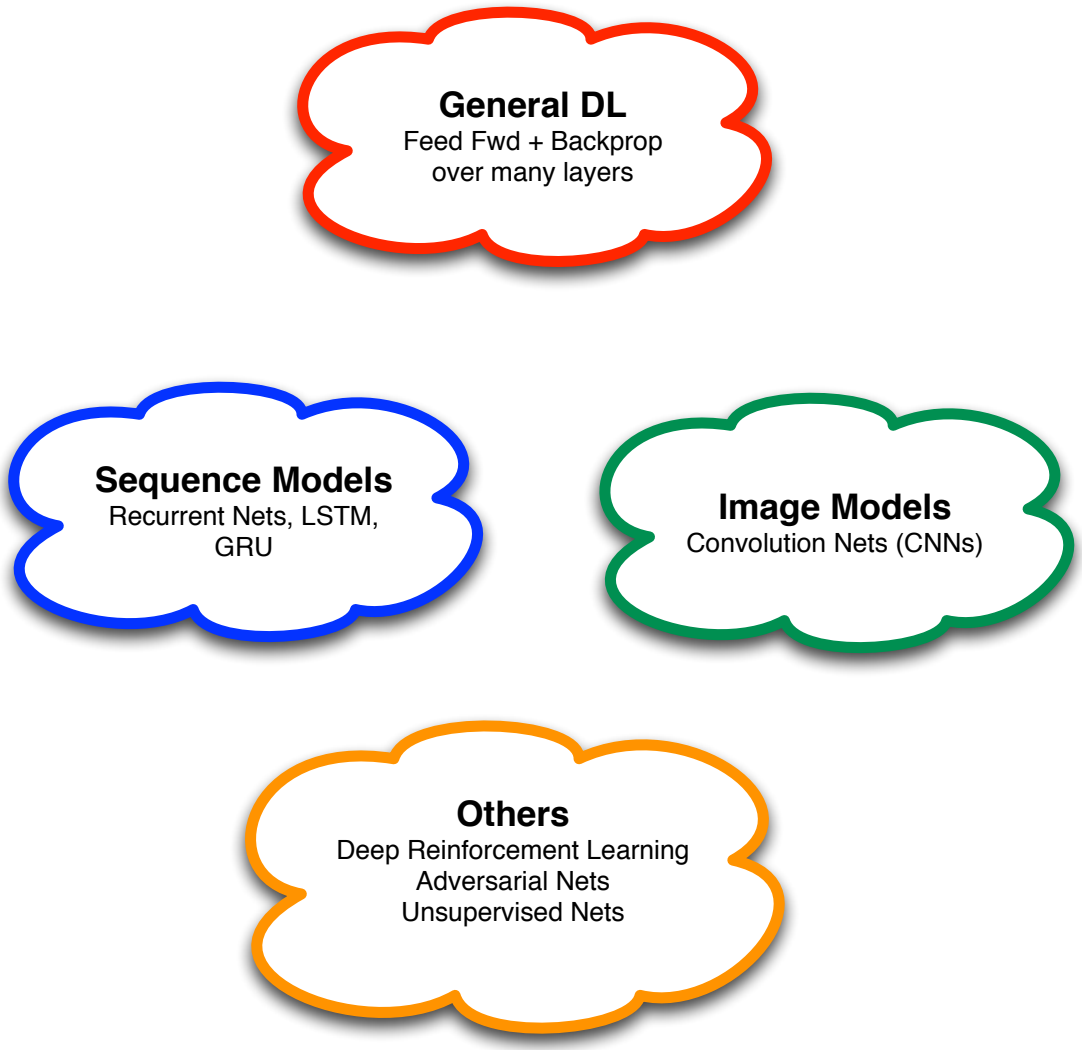
Endless Applications of Neural Networks

- Stock and commodity price predictions
- Electrical load predictions for the power industry
- Detection of disease from MRI images
- Facial recognition
- Colorization of old black-white movies.
- Natural language interpretation, generation and translation.
- Production of captions for images.
- Control of self-driving vehicles.
- Generation of art, poetry and music
- Automated journalism: given data, write article.

Secrets to Deep Learning Success

- Big Data - *Data is the new oil*
- GPUs - greatly speed up the complex calculations of backpropagation.
- Convolution nets - based on mammalian visual processing.
- LSTMs - slick implementation of recurrence adds critical memory of varying durations.
- Dropout - deactivation of random subsets of neurons improves generalization.
- Rectified Linear Units (ReLU) - very simple activation function reduces the vanishing-gradient problem $\rightarrow$ backprop works in very deep networks.

The Universe of Deep Learning



** Nuts and Bolts of Applying Deep Learning (Andrew Ng, 2016, YouTube)*