CS 4824/ECE 4424: Deep Neural Networks I

Acknowledgement:
Many of these slides are derived from Tom Mitchell, Pascal Poupart, Pieter Abbeel, Eric Eaton, Carlos Guestrin, William Cohen, and Andrew Moore.
Deep Neural Networks

- **DNN**: neural network with many hidden layers

- **Advantage**: highly expressive

- **Challenges**:
  - How to train a deep neural network?
  - How to avoid overfitting?
Expressiveness

- Neural networks with one hidden layer of sigmoid/tanh units can approximate arbitrarily closely neural networks with several layers of sigmoid/hyperbolic units.

- However, as we increase the number of layers, the number of units needed may decrease exponentially (with the number of layers).
Example – Parity Function

- Odd or even
  \[
  \begin{align*}
  1 & \text{ if odd} \\
  -1 & \text{ if even}
  \end{align*}
  \]

- Possible odd combinations

\[\begin{array}{cccc}
X_1 & X_2 & X_3 & X_4 \\
\hline
1 & 1 & 1 & -1 \\
1 & 1 & -1 & 1 \\
1 & 1 & -1 & -1 \\
-1 & 1 & 1 & 1 \\
-1 & 1 & 1 & -1 \\
-1 & 1 & -1 & 1 \\
-1 & 1 & -1 & -1 \\
-1 & -1 & 1 & 1 \\
-1 & -1 & 1 & -1 \\
-1 & -1 & -1 & 1 \\
-1 & -1 & -1 & -1 \\
\end{array}\]
Example – Parity Function

- Single layer of hidden nodes

\[
\begin{align*}
\text{parity}(x_1, x_2, x_3, x_4) &= \begin{cases} 
1 & \text{if odd} \\
-1 & \text{if even}
\end{cases} \\
\text{for } n=4, \quad 2^{n-1} = 2^3 = 8 \text{ odd subsets}
\end{align*}
\]
Example – Parity Function

- $2n - 2$ layers of hidden nodes

The diagram shows a parity function with $2n - 2$ layers of hidden nodes. The function output is given by:

$$\begin{cases} 
1 & \text{if odd} \\
-1 & \text{if even}
\end{cases}$$
The power of depth (practice)

Deep neural networks learn hierarchical feature representations

Challenge: how to train deep NNs?
Speech

- **2006** (Hinton and coworkers): first effective algorithm for deep NN
  - layer-wise training of Stacked Restricted Boltzmann Machines (SRBM)s

- **2009**: Breakthrough in acoustic modeling
  - replace Gaussian Mixture Models by SRBMs
  - Improved speech recognition at Google, Microsoft, IBM

- **2013**: recurrent neural nets (LSTM)
  - Google error rate: 23% (2013) to 8% (2015)
  - Microsoft error rate: 5.9% (Oct 17, 2016) same as human performance

- ...
Image Classification

- ImageNet Large Scale Visual Recognition Challenge
Challenges in Deep Neural Networks

- How to train a deep neural network?
- How to avoid overfitting?