CS 4824/ECE 4424: Convolutional 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.
Large Networks

- What kind of neural networks can be used for large or variable length input vectors (e.g., time series)

- Common networks
  - Pre-training– Convolutional networks
  - Recurrent networks
Convolution

- Convolution: mathematical operation on two functions $x()$ and $w()$ that produces a third function $y()$ that can be viewed as a modified version of one of the original functions $x()$

$$y(i) = \int_{t} x(t)w(i - t)dt$$

$$y(i) = (x * w)(i)$$

- Where $*$ is an operator denoting a convolution
Effect of convolution — smoothing
Discrete convolution

- Discrete convolution
  \[ y(i) = \sum_{t=-\infty}^{\infty} x(t)w(i - t) \]

- Multidimensional convolution
  \[ y(i, j) = \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} x(t_1, t_2)w(i - t_1, j - t_2) \]
Application: edge detection

- Consider a grey scale image
- Detect vertical edges: \( y(i,j) = x(i,j) - x(i-1,j) \)

\[
\text{hence } w(i - t_1, j - t_2) = \begin{cases} 
1 & t_1 = i, t_2 = j \\
-1 & t_1 = i - 1, t_2 = j \\
0 & \text{otherwise}
\end{cases}
\]
Convolutions for feature extraction

- In neural networks
  - A convolution denotes the linear combination of a subset of units based on a specific pattern of weights.

\[
a_j = \sum_i w_{ji} z_i
\]

- Convolutions are often combined with an activation function to produce a feature

\[
z_j = h(a_j) = h\left(\sum_i w_{ji} z_i\right)
\]
Convolution Neural Network (CNN)

- A CNN refers to any network that consists of an alternation of convolution and pooling layers, where some of the convolution weights are shared

- Architecture:
Pooling

- **Pooling**: commutative mathematical operation that combines several units

- **Examples**:
  - max, sum, product, average, Euclidean norm, etc.

- **Commutative property** (order does not matter):
  - $\max(a, b) = \max(b, a)$