More on Neural Nets

Andrea Danyluk
April 17, 2017

Announcements

- Classifier learning assignment in progress
- Final project
  - Schedule/deliverables posted on the course website

Today’s Lecture

- Finishing up with neural nets and backprop
- Pomerleau papers

Artificial Neural Networks

- Characterized by:
  - A large number of (simple) neuronlike processing elements
  - A large number of weighted connections between the elements
  - Highly parallel, distributed control
  - An emphasis on learning internal representations automatically
- Theoretically principled training algorithms that aim to minimize an objective function (error)

An Artificial Neuron: the Perceptron

Note slight difference from the R&N formulation. In theirs, \( x_0 \) always has value -1, rather than 1.
Instead, we’ll consider networks of artificial neurons that use the sigmoid function as the activation function [from Mitchell].

Activation Functions

- \( \sigma(x) \) is the sigmoid function: \( 1 / (1+e^{-x}) \)
- \( d\sigma(x)/dx = \sigma(x)(1-\sigma(x)) \)
- Can derive a gradient descent rule to train
  - One sigmoid unit
  - Multilayer networks of sigmoid units

The Sigmoid Unit

\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]

\[
\sigma'(x) = \sigma(x)(1-\sigma(x))
\]

Requirements: A Smooth Threshold Function

Error Gradient for a Sigmoid Unit

\[
\frac{\partial E}{\partial w_i} = \frac{8}{2} \sum_j (t_j - o_j) \frac{\partial \sigma'(o_j)}{\partial w_i}
\]

\[
= \frac{1}{2} \sum_j (t_j - o_j) \frac{\partial (t_j - o_j)^2}{\partial w_i}
\]

\[
= \sum_j (t_j - o_j) \left(-\frac{\partial o_j}{\partial w_i}\right)
\]

Backpropagation Algorithm

- Initialize all weights to small random numbers.
- For each training example, Do
  1. Input the training example to the network and compute the network outputs
  2. For each output unit \( k \)
     \[ d_k = o_k(1-o_k)(t_k-o_k) \]
  3. For each hidden unit \( b \)
     \[ d_b = o_b(1-o_b) \sum_{k \text{output}} w_{bk} d_k \]
  4. Update each network weight \( w_{ij} \)
     \[ w_{ij} \leftarrow w_{ij} + \Delta w_{ij} \]
     where
     \[ \Delta w_{ij} = \eta d_j x_i \]

Paper Discussion

- Pioneering work in autonomous vehicles
- Dean Pomerleau ’87