More on Neural Nets

Andrea Danyluk
April 17, 2017

Announcements

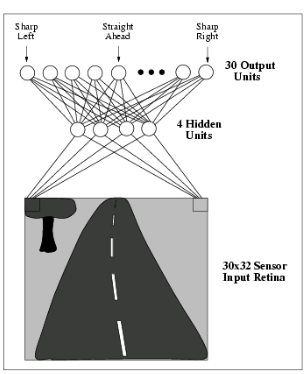
- Classifier learning assignment in progress
- Final project
  - Schedule/deliverables posted on the course website

Today's Lecture

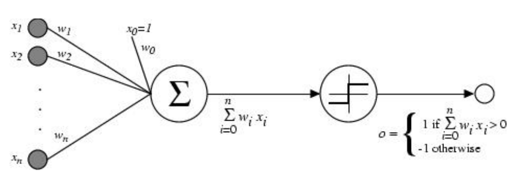
- Finishing up with neural nets and backprop
- Pomerleau papers

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)

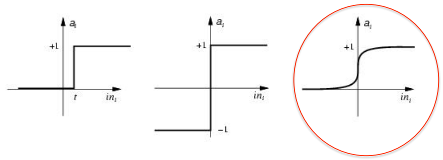


An Artificial Neuron: the Perceptron



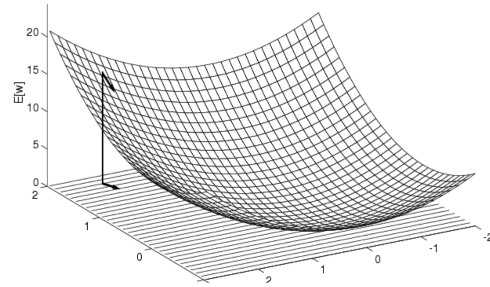
Note slight difference from the R&N formulation. In theirs, x_0 always has value -1, rather than 1.

Activation Functions

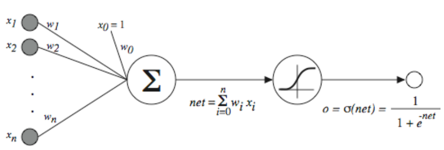


(a) Step function (b) Sign function (c) Sigmoid function

Gradient Descent

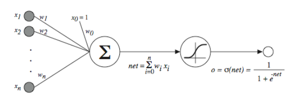


The Sigmoid Unit



- σ(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

Error Gradient for a Sigmoid Unit



$$\begin{aligned}
\frac{\partial E}{\partial w_i} &= \frac{\partial}{\partial w_i} \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2 \\
&= \frac{1}{2} \sum_d \frac{\partial}{\partial w_i} (t_d - o_d)^2 \\
&= \frac{1}{2} \sum_d 2(t_d - o_d) \frac{\partial}{\partial w_i} (t_d - o_d) \\
&= \sum_d (t_d - o_d) \left(-\frac{\partial o_d}{\partial w_i} \right)
\end{aligned}$$

Backpropagation Algorithm

Initialize all weights to small random numbers.
Until satisfied, Do

- For each training example, Do
  1. Input the training example to the network and compute the network outputs
  2. For each output unit k

     $$\delta_k \leftarrow o_k(1 - o_k)(t_k - o_k)$$

  3. For each hidden unit h

     $$\delta_h \leftarrow o_h(1 - o_h) \sum_{k \in outputs} w_{h,k} \delta_k$$

  4. Update each network weight $w_{i,j}$

     $$w_{i,j} \leftarrow w_{i,j} + \Delta w_{i,j}$$

     where

     $$\Delta w_{i,j} = \eta \delta_j x_i$$

Paper Discussion

- Pioneering work in autonomous vehicles
- Dean Pomerleau '87