Backpropagation a Worked Example

Hypothesis Space and Search Bias

- Hypothesis space
  - m weights -> m-dimensional space, each dimension takes a value (-1,1).
  - Continuous nature not discrete like decision trees
- Search bias
  - Consider to be a smooth interpolation between points.
- Hidden Layer
  - Holds intermediate representations.
During Training …

Convergence Issues

Gradient descent to some local minimum

- Perhaps not global minimum...
- Add momentum
- Stochastic gradient descent
- Train multiple nets with different initial weights

Nature of convergence

- Initialize weights near zero
- Therefore, initial networks near-linear
- Increasingly non-linear functions possible as training progresses
Over-fitting

- Training epoch is passing all data through the network
- How many training epochs?
  - Validation set
  - Weight decay
- Why weight decay?
  - With identical weights $\rightarrow$ smooth decision surface
  - Increased weights mean more complex surfaces which can overfit the training set.
Overfitting and Weight Decay

\[ \Delta w_{ij} = -\mu \frac{\partial E}{\partial w_{ij}} = -\mu \frac{\partial E}{\partial h_{ij}} - \mu \nu w_{ij} \]

Recurrent Networks

(a) Feedforward network

(b) Recurrent network

(c) Recurrent network unfolded in time
Dynamically Modifying Structure

• Effectively dynamically changing the model/hypothesis space.
• Cascade-correlation
  – Start with no hidden nodes, Train
  – Add an extra hidden node with fixed chosen weights
  – Retrain, Whenever a hidden node is added, its inputs are the all the input layer nodes and the other hidden layer nodes.
  – Reduces training time
• Start with overly complex network
  – Remove nodes with zero weights or perform sensitivity analysis with respect to the weights.

Reading for Next Classes

• Additional reading
  – Rumelhart work on neural networks mimicking learning verb tense
  – www.dcs.shelf.ac.uk/~yorick/ai_course/lecture_9.ppt
• Lecture 9 – 03/04/02 and onwards
  – Unsupervised Learning
  – K-Means, SOM
  – Reinforcement Learning (Mitchell, Chapter 13)