Multilayer Networks

- **Capabilities**
  - Depend directly on total number of weights and threshold values.
  - A one hidden layer network with sufficient number of hidden units can arbitrarily approximate any boolean function, pattern recognition problems, and well-behaved function approximation problems.
  - Sigmoidal units more powerful than linear threshold units.
Error backpropagation


- Two passes: forward pass (computational pass), backward pass (weight correction pass).
- Analog computations based on MSE criterion.
- Hidden units usually sigmoidal units.
- Initialization: weights take on small random values.
- Algorithm may not converge to global minimum.
- Algorithm converges slower than for linear networks.
- Representation is distributed.
BP Algorithm Comments

- $\delta$s are error terms computed from output layer back to first layer in dual network.
- Training is usually done on-line.
- Examples presented in random or sequential order.
- Update rule is local as weight changes only involve connections to weight.
- Computational complexity depends on number of computational units.
- Initial weights randomized to avoid converging to local minima.
BP Algorithm Comment continued

- Threshold weights updated in similar manner to other weights (input = 1).
- Momentum term added to speed up convergence.
- Step size set to small value.
- Sigmoidal activation derivatives simple to compute.
BP Architecture

Forward network

Output of computational values calculated

Output of error terms calculated

Sensitivity network
Modifications to BP Algorithm

- Batch procedure
- Variable step size
- Better approximation of gradient method (momentum term, conjugate gradient)
- Newton methods (Hessian)
- Alternate cost functions
- Regularization
- Network construction algorithms
- Incorporating time
Using Learning Theory

- For SVMs and feedforward networks there are several parameters to preset.
  - Regularization terms: C
  - Type of kernel, kernel parameters, number of hidden nodes and weights
- To perform SRM can use
  - Trial and error
  - Crossvalidation
  - Bayesian learning methods
- Occam’s razor: The simplest explanation is the best.