## NEURAL NETWORK WITH LOGISTIC SIGMOIDS

| Property | Rating | Comment |
|---|---|---|
| Network output function: | | |
| Continuous | Yes | |
| Smooth | Yes | Basis functions—logistic sigmoids—are infinitely differentiable. |
| Easily visualized | No | Result of summing sigmoids is very difficult to visualize. |
| Generalizes beyond domain | Somewhat | Sigmoids asymptote at 0 or 1 as argument values approach $\pm\infty$. Thus, the network output is bounded and is constant over wedge shaped regions extending out to infinity. |
| Method for determining coefficients or weights: | | |
| Inner product of target function and basis function | No | Logistic sigmoids in infinite number are complete but not orthogonal. |
| Simultaneous equations for data points | Poor | Suffers from the limitations of high-order polynomial fits—possible oscillations between data points. This method also fails to dictate values for hidden-layer weights. |
| Gradient descent with all points on surface available for training. | Good | Can get stuck in local minimum, but despite this drawback results are usually good. Gradient descent is usually the most viable method of training. |
| Gradient descent with finite number of points on surface available for training. | Adequate | As a rule of thumb, results are satisfactory if the number of training points is at least ten times the number of weights in the network. |
| Behavior at data points: | | (Gradient descent training assumed.) |
| Reproduces data values exactly | No | Gives results similar to least squares. |
| Suited to randomly scattered data versus regular grid | Yes | Scattered data poses no problem so long as coverage of domain is thorough and fairly uniform. |
| Well-behaved between data points | Poor | Training for too long causes overfitting of data and oscillations between data points. |
| Expands for new data points | Yes | Need only resume or continue training as new points are added. |
| Matches slope at finite number of data points | Perhaps | There is no provision for matching slopes in standard gradient descent, but this could be added. |
| Complexity: | | |
| Difficulty of writing computer program | Medium | Gradient descent equations are somewhat difficult to code correctly and debug. |
| Speed of coefficient calculation | Low | Gradient descent usually requires many thousands of training pattern presentation. |
| Speed of function evaluation | Low | Every sigmoid must be evaluated, and each sigmoid calculation is time consuming. |