development for cognitive science. Such interactions will lead to a deeper understanding of the interpretation and learning tasks, and may ultimately help us to address other cognitive tasks, perhaps including creative thinking and scientific discovery, as well.

ACKNOWLEDGMENTS

Support for the preparation of this article was provided by NIMH Project Program Grant MH47566-03 and by NIMH Research Scientist Award MH00385-13.

Comment

B. D. Ripley

Bing Cheng and Mike Titterington have reviewed many of the areas of neural networks; their paper overlaps the flood of books on the subject. I also recommend Weiss and Kulikowski (1991) (Segre and Gordon, 1993, provide an informative review) and Gallant (1993) for their wider perspective and Wasserman (1993) for coverage of recent topics. My own review article, Ripley (1993a), covers this and many of the cognate areas as the authors comment. The five volumes of the NIPS proceedings (*Advances in Neural Information Processing Systems*, 1989–1993, various editors) provide a very wide-ranging overview of highly-selected papers. Much of the latest work is available electronically from the ftp archive at `archive.cis.ohio-state.edu` in directory `pub/neuroprose`.

At the time I received this paper to discuss, I had recently attended a NATO Advanced Study Institute on *From Statistics to Neural Networks* (whose proceedings will appear as Cherkassky, Friedman and Wechsler, 1994), which despite the direction of the title revealed that current thoughts in neural networks are not to subsume statistics in neural networks but vice versa. Many researchers in neural networks are becoming aware of the statistical issues in what they do and of relevant work by statisticians which encourages fruitful discussions.

Cheng and Titterington concentrate on similarities between statistical and neural network methods. I feel the differences are more revealing as they indicate room for improvement on at least one side. However, I believe the most important issues to be those of practice which are almost ignored in the paper. Before I turn to those, there are two points I wish to attempt to clarify.

B. D. Ripley is Professor of Applied Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, United Kingdom. This comment was written while on leave at the Isaac Newton Institute for Mathematical Sciences, Cambridge, United Kingdom.

1. PROJECTION-PURSUIT REGRESSION

The connection between multilayer perceptrons (MLPs) and projection-pursuit regression (PPR) is much deeper than the authors appear to suggest. Other empirical comparisons (apart from my own cited in the paper) are given by Hwang et al. (1992a,b, 1993), and Barron and Barron (1988) viewed PPR from a network viewpoint. In the authors' notation PPR is

$$y_i = w_{0i} + \sum_k \gamma_i \psi_k (x^T v_k),$$

where I have allowed for multiple outputs. An MLP with linear output units is the special case of logistic ψ_k; of course both PPRs and MLPs can be given nonlinear output units. Since we can approximate any continuous ψ_k of compact support uniformly by a step function and can approximate (nonuniformly) a step function by a logistic, we can approximate ψ_k uniformly by a sum of logistics. This fact plus the (elementary) approximation result for PPR of Diaconis and Shahshahani (1984) gives the approximation results of Cybenko and others. There is a version of Barron's L_2 result for PPR by Zhao and Atkeson (1992). (This point of view, approximating ψ_k by a simple neural net of one input, corresponds to organized weight-sharing between input-to-hidden-unit weights for groups of units, a sensible procedure in its own right.)

These results suggest that the approximation capabilities of MLPs and PPR are very similar (suggesting an affirmative partial answer to the question in Section 7). However, PPR will have an advantage when there are many inputs, only a few combinations of which are relevant, in making better use of each projection and hence fewer projections and parameters. My suspicion is that this is commonly the case.