Issues of choosing model complexity and assessing performance and "generalization" (Section 4.3.4) are among the most important open questions. There is some evidence that methods such as cross-validation and AIC are too "local" to fully assess the variability of very flexible methods; therefore some of the assessed benefits of nonlinear methods may be illusory. [On "generalization", Haussler (1992) is a far-reaching extension of the ideas of VCdim to which statisticians, especially David Pollard and Luc Devroye, have contributed; and Anthony and Biggs (1992) is an introductory text on the seminal ideas of Blumer et al., 1989.]

One thing statisticians can contribute to the debate is experience in careful use of sophisticated nonlinear methods. Software is readily available, including in S, and I would encourage statisticians to experiment rather than quote inadequately designed propaganda studies.

To end on a positive note, some very impressive applied statistics is being done using neural networks, and the explosive growth of the subject has opened the eyes of some statisticians (including myself) to the complexity of problems that may be fruitfully attacked by nonlinear methods. I and others have been particularly impressed by some work of my Oxford Engineering Science colleague, Lionel Tarassenko, on analyzing sleep EEG data using both Kohonen nets and radial basis functions to detect structure and anomalous signals (Roberts and Tarassenko, 1993, 1994).

Comment

Robert Tibshirani

Cheng and Titterington's paper is a scholarly overview of the field of neural networks. It should raise the statisticians' awareness of this interesting and important field. One of the authors' objectives was to encourage cross-disciplinary research between neural network researchers and statisticians. Here at the University of Toronto, I have been collaborating informally with Geoffrey Hinton of the Computer Science department, and I think that this collaboration has been fruitful for both of us.

First I would like to make a general point drawing a distinction between statistics and neural networks:

Statisticians tend to work with more interpretable models, since measuring the effects of individual input variables, rather than prediction, is often the purpose of the analysis.

Having said that, there is still much that one field can learn from the other. I will briefly summarize some of the main points:

WHAT THE STATISTICIAN CAN LEARN FROM NEURAL NETWORK RESEARCHERS

1. We should worry less about statistical optimality and more about finding methods that work, especially with large data sets.
2. We should tackle difficult real data problems like some of those addressed by neural network researchers, like character and speech recognition and DNA structure prediction. As John Tukey has said, it is often better to get an approximate solution to a real problem than an exact solution to an oversimplified one.
3. Models with very large numbers of parameters can be useful for prediction, especially for large data sets and problems exhibiting high signal-to-noise ratios.
4. Modelling linear combinations of input variables can be a very effective approach because it provides both feature extraction and dimension reduction.
5. Iterative, nongreedy fitting algorithms (like steepest descent with a learning rate) can help to avoid overfitting in models with large numbers of parameters.
6. We (statisticians) should sell ourselves better.

WHAT THE NEURAL NETWORK RESEARCHER CAN LEARN FROM STATISTICIANS

1. They should worry more about statistical optimality or at least about the statistical properties of methods.
2. They should spend more effort comparing their methods to simpler statistical approaches. They will be surprised how often linear regression performs as well as a multilayered percep-

Robert Tibshirani is Associate Professor, Department of Preventive Medicine and Biostatistics, University of Toronto, 12 Queens Park, Toronto, Ontario M5S 1A8, Canada.

tron. They should not use a complicated model where a simple one will do.

The history of the projection pursuit regression (PPR) model illuminates some of the differences between the two fields. As noted by Cheng and Titterington, the PPR model has the same form as a single layer perceptron. When PPR was introduced into the statistics field by Friedman and Stuetzle in 1981, it did not have much practical impact. There are a number of possible reasons for this. Computationally, it was ahead of its time: many statisticians still do not feel comfortable using very large amounts of computation in an analysis. In addition, statisticians do not often tackle the large prediction problems that can often benefit from such an approach. Finally, the particular fitting (learning) procedure might have been too greedy to work effectively with large number of projections.

In contrast, neural network researchers have developed and applied the PPR model to some difficult problems with considerable success. In recent years, they have further improved their results by applying classical statistical techniques such as regularization, cross-validation and Bayesian modelling. This suggests that both fields should be listening and learning from each other. Cheng and Titterington's paper will help this cause.

Rejoinder

Bing Cheng and D. M. Titterington

We are very grateful to the discussants for the time and effort they have expended in commenting on our paper. When we submitted the revised version of the paper, we felt some trepidation that, in spite of our best effort at brevity, the paper still seemed very long in comparison to many other contributions to the journal, and yet we were fully aware that we had not done justice to important aspects of the field. Fortunately, some of our sins of omission have been absolved by the choice of discussants, and we are happy to regard many of their comments as complementary to our presentation. *Statistical Science* can, therefore, be said to be publishing a 10-author review of the interface between statistics and neural network research rather than a two-author review plus discussion. We are glad that the discussants include representives from what one may call (against Breiman's advice) the mainstream neural-network community (McClelland), as well as distinguished statisticians with both short and long (in terms of time) records of involvement in the area. We apologize to all discussants for not having space to respond to each of the many points they have made.

Later in our rejoinder we shall remark on some points raised by individual discussants, and we shall finish by pulling together views about the future of the interface. First, we mention three areas of research on which several discussants expressed views. These areas were implicitly identified by Amari and, in slightly different form, by Breiman.

- Mathematical modeling of real cognitive processes
- Theoretical investigations of networks and neurocomputing
- Development of useful tools for practical prediction and pattern recognition

MODELING OF REAL COGNITIVE PROCESSES

The dominant discussant here is McClelland. He emphasizes the fact that machine intelligence still has far to go to emulate many human mental processes, a view echoed by Bienenstock and Geman. McClelland sounds more hopeful than they do that concepts closely akin to artificial networks, presumably as known today, might prove to be key aspects. Furthermore, he suggests that the mechanics of statistics will be important in the development of such realistic cognitive machines: first, manipulation of probability models using Bayes' theorem could be the way to mimic the brain's approach to data analysis ("interpretation"); second, nondeterministic elements seem to be inevitable in modeling any realistic learning process. Practical realization of such models does, however, seem to be a daunting prospect. In the "interpretation" question, for instance, the equivalent of a prior distribution will have to include representation of all useful contextual and background information.

However, it seems clear from McClelland's penultimate paragraph that there are important new developments in areas such as speech processing, even in irritatingly irregular languages such as English and even then in the arguably less irregular American version. We are, nevertheless, doubtful about