Statistical Aspects of Artificial Neural Networks 3.5.1

Comparison of ANNs to Statistical Analysis In traditional statistical analysis, the modeller is required to specify the precise relationship between inputs and outputs and any restrictions that may be implied by theory. ANNs differ from conventional techniques in that the analyst is not required to specify the nature of the relationships involved; the analyst simply identifies the inputs and the outputs. According to Sarle [1994], no knowledge of ANN training methods such as backpropagation is required to use ANNs. In addition, Sarle states that the MLP’s main 26 This assumes that the global minimum is not very far from the local minima. Chapter 3: The Technical and Statistical Aspects of Artificial Neural Networks CNW Tan Page 74 strength lies in its ability to model problems of different levels of complexity, ranging from a simple parametric model to a highly flexible, nonparametric model. For example, an MLP that is used to fit a nonlinear regression curve, using one input, one linear output, and one hidden layer with a logistic transfer function, can function like a polynomial regression or least squares spline. It has some advantages over the competing methods. Polynomial regression is linear in parameters and thus is fast to fit but suffers from numerical accuracy problems if there are too many wiggles. Smoothing splines are also linear in parameters and do not suffer from numerical accuracy problems but pose the problem of deciding where to locate the knots. MLPs with nonlinear transfer function, on the other hand, are genuinely nonlinear in the parameters and thus require longer computational processing time. They are more numerically stable than high-order polynomials and do not require knot location specification like splines. However, they may encounter local minima problems in the optimization process. 3.5.2 ANNs and Statistical Terminology Although there are many similarities between ANN models and statistical models, the terminology used in both fields are quite different. For example, Sarle [1994] claims that the terminology ‘back-propagation’, should refer only to the simple process of applying the chain rule to compute derivatives for the generalized delta rule algorithm, and not the training method itself. He adds that this confusion is symptomatic of the general failure in ANN literature to differentiate between models and estimation methods. Sarle [1996] gives a list of statistics terminology that has its equivalence in ANN literature. Some of the more common ones are listed in Table 3-2. Chapter 3: The Technical and Statistical Aspects of Artificial Neural Networks CNW Tan Page 75 Table 3-2 Statistical and ANN Terminology Statistical Terminology ANN Terminology variables features independent variables inputs predicted values outputs dependent variables targets or training values. residuals errors estimation training, learning, adaptation, or self-organization. an estimation criterion an error function, cost function, or Lyapunov function observations patterns or training pairs parameter estimates (synaptic) weights regression and discriminant analysis supervised learning cluster analysis or data reduction unsupervised learning, self-organization or competitive learning interpolation and extrapolation generalization intercept bias error term noise forecasting prediction 3.5.3 Similarity of ANN Models to Statistical Models In general, feedforward nets with no hidden layer are basically generalized linear models [Neural Nets FAQ, Part 1, 1996]. Sarle [1994] states that the perceptron model with different transfer functions has been shown to have equivalent statistical models. For example: Chapter 3: The Technical and Statistical Aspects of Artificial Neural Networks CNW Tan Page 76 · A perceptron model with a linear transfer function is equivalent to a possibly multiple or multivariate linear regression model [Weisberg 1985; Myers 1986]. · A perceptron model with a logistic transfer function is a logistic regression model [Hosmer and Lemeshow 1989]. · A perceptron model with a threshold transfer function is a linear discriminant function [Hand 1981; McLachlan 1992; Weiss and Kulikowski 1991]. An ADALINE is a linear two-group discriminant. MLP models have their statistical equivalent models too. For example: · An MLP with one output is a simple nonlinear regression [Sarle 1994]. · An MLP with a moderate number of hidden neurons is essentially the same as a projection pursuit regression, except that an MLP uses a predetermined transfer function while the projection pursuit regression model uses a flexible nonlinear smoother [Sarle 1994]. · An MLP becomes a nonparametric sieve if the number of hidden neurons is allowed to increase with the sample size [White 1988]. This makes it a useful alternative to methods such as kernel regression [Hardle 1990] and smoothing splines. Kuan and White [1995] states that when Cowan [1967] proposed the replacement of the Heaviside activation function with a smooth sigmoid function, specifically, the logistic function, the model proposed by McCulloch and Pitts [1943] becomes similar to the binary logit probability model [Ameniya 1981, 1985 p. 268]. Ameniya states that these models have great utility in econometric applications where binary classifications or decisions are involved. Chapter 3: The Technical and Statistical Aspects of Artificial Neural Networks CNW Tan Page 77 White [1992, p. 84] states that the back-propagation algorithm is not new and is, in fact, a statistical method call stochastic approximation, first proposed by Robins and Munro in 1951. This had lead to an explosion of both theoretical and applied research of the field in the past 40 years. It has been used extensively in pattern recognition and systems identification literature. The major advantage of this, is that the considerable statistics and engineering literature on stochastic approximation can be applied to make general statements about the BP algorithm [White 1989a]. 3.5.4 ANNs vs. Statistics Sarle [1993] claims that “many ANN researchers are engineers, physicists, neurophysiologists, psychologists, or computer scientists, who know little about statistics and nonlinear optimization and they often reinvent methods that have been known in statistical or mathematical literature for decades or centuries without understanding how these methods work”. He reasons that the common implementation of ANNs is biological or engineering criteria based, such as how easy it is to fit the net on a chip, rather than on well-established statistical and optimization criteria. Ripley [1993] seems to agree with Sarle (both being statisticians by training), stating that ANNs have been developed very rapidly by workers with diverse backgrounds, most with little or no experience in data analysis. Ripley [1993] claims that although comparisons of ANNs to other methods are rare, however, when done carefully, often show that statistical methods can outperform the state-of-the-art ANNs. His paper includes a comment from Aharonian [1992] on ANNs and financial applications. Aharonian states that most ANNs papers on financial analysis either report results no more accurate than those obtained by traditional statistical Chapter 3: The Technical and Statistical Aspects of Artificial Neural Networks CNW Tan Page 78 techniques; or they fail to compare their results to traditional statistical analysis and by not doing so, invalidate any claims of a breakthrough. Before the popularity of ANNs, few financial institutions used any form of statistical methods (except for technical analysis, which some may claim to be pseudo-statistics) for financial trading, and even fewer had a dedicated quantitative analysis unit for financial analysis which is now a common sight in most major banks’ dealing rooms. As mentioned in chapter 1, financial institutions are second only to the US Department of Defense in sponsoring research into ANNs [Trippi and Turban 1996]. 3.5.5 Conclusion of ANNs and Statistics Sarle [1994] concludes it is unlikely that ANNs will supersede statistical methodology as he believes that applied statistic is highly unlikely to be reduced to an automatic process or ‘expert system’. He claims that statisticians depend on human intelligence to understand the process under study and an applied statistician may spend more time defining a problem and determining what questions to ask than on statistical computation. He does, however, concede that several ANNs models are useful for statistical applications and that better communication between the two fields would be beneficial. White [1992, p. 81] agrees that statistical methods can offer insight into the properties, advantages and disadvantages of the ANN learning paradigm, and conversely ANN learning methods have much to offer in the field of statistics. For example, statistical methods such as Bayes analysis and regression analysis have been used in generating forecasts with confidence intervals that have deeper theoretical roots in statistical inference and data generating processes. ANN is superior for pattern recognition and is able to deal with any model whereas statistical methods require randomness. Chapter 3: The Technical and Statistical Aspects of Artificial Neural Networks CNW Tan Page 79 ANNs have contributed more to statistics than statisticians would care to admit. They have enabled researchers from different disciplines and backgrounds to use modeling tools that were once only available to statisticians due to the complexities and restrictive conditions imposed by statistical models. By making modeling more accessible (and more interesting perhaps), ANNs researchers without statistical background are beginning to gain an appreciation of statistical methodologies due to the inevitable crossing of paths between ANNs and statistics. There are definitely more visible ANN commercial applications than statistical applications even though the claim that some of the ANN methodologies were already been ‘known for decades if not centuries in statistical and mathematical literature [Sarle 94]’.