



# Backpropagation and Regression: Comparative Utility for Neuropsychologists

Thomas D. Parsons<sup>1</sup>, Albert A. Rizzo<sup>2</sup>, and J. Galen Buckwalter<sup>3</sup>

<sup>1</sup>Fuller Theological Seminary, Graduate School of Psychology, Pasadena, CA, USA, <sup>2</sup>Department of Computer Science, University of Southern California, Los Angeles, CA, USA, and <sup>3</sup>Department of Research and Evaluation, Southern California Permanente Medical Group, Pasadena, CA, USA

## ABSTRACT

The aim of this research was to compare the data analytic applicability of a backpropagated neural network with that of regression analysis. Thirty individuals between the ages of 64 and 86 (Mean age = 73.6; Mean years education = 15.4; % women = 50) participated in a study designed to validate a new test of spatial ability administered in virtual reality. As part of this project a standard neuropsychological battery was administered. Results from the multiple regression model ( $R^2 = .21, p < .28$ ; Standard Error = 18.01) were compared with those of a backpropagated ANN ( $R^2 = .39, p < .02$ ; Standard Error = 13.07). This 18% increase in prediction of a common neuropsychological problem demonstrated that an ANN has the potential to outperform a regression.

Conventional methods for prediction in neuropsychological research make use of the General Linear Model's (GLM) statistical regression (Neter, Wasserman, & Kutner, 1989). Although linear regression analysis subsumes univariate analyses and can provide a robust understanding of data, studies are regularly carried out and inferences made without verifying normality and error independence (Box, 1966; Darlington, 1968; Dempster, 1973; Tukey, 1975). While linear regression analysis is fairly robust against departures from the normality assumption (Mosteller & Tukey, 1977), there are instances (correlated error, curvilinear relations, etc.) where parametric data analysis can pose a significant amount of constraint. Consequently, nonparametric models (Gallant, 1987; Gordon, 1968; Green & Silverman, 1994; Haerdle, 1990; Ross, 1990; Seber & Wild, 1989), including Artificial Neural Networks (ANNs), have become more appealing (Bishop, 1995; Ripley, 1993).

ANNs can provide several advantages over conventional regression models. They are claimed to possess the property to learn from a set of data without the need for a full specification of the decision model; they are believed automatically to provide any needed data transformations. They are also claimed to be more robust in the presence of noise and distortion (Bishop, 1995; Hertz, Krogh, & Palmer, 1991; Hinton, 1992; Pao, 1989; Ripley, 1993; Soucek, 1992; Wasserman, 1989).

In this research the aim was to demonstrate the applicability of a backpropagated ANN for use in a common neuropsychological problem. Additionally we compared its performance with that of conventional regression analysis. The goal is to make the often heuristic and ad hoc process of neural network development transparent to interested neuropsychologists and to encourage neuropsychological researchers to view ANNs as viable data analytical tools.

### General Linear Model

The GLM underlies most of the statistical analyses used in neuropsychological research. It is a conceptualization of variance between-groups (effect) and within-groups (error). It is comprised of three components: the grand mean, the predicted effect, and random error (McCullagh & Nelder, 1989; Licht, 1995). In the GLM's regression analysis, relationships among variables are expressed in a linear equation that conveys a criterion as a function of a weighted sum of predictor variables. Neuropsychological researchers use regression to assess both (a) the degree of accuracy of prediction and (b) the relative importance of different predictors' contribution to variation in the criterion (Kachigan, 1986).

Although the GLM is well known in data analysis, is reliable, and can provide robust particulars, the user must have the time and resources to perform an evaluation of the entire database. Further, managing the error independence problems found in neuropsychological research necessitates even more sophisticated proficiencies (Gallant, 1987; Haerdle, 1990). The GLM tends to ascertain the more concrete significant trends while negating individual particularities (Gordon, 1968). In cases where linear approximation is not possible due to noise (noise is not an inherent randomness or absence of causality in the world; rather, it is the effect of missing, or inaccurate, information about the world. In neuropsychology, noise may include things such as confounding variables, nonparametric data, nonlinear associations, measurement error), or when nonlinear approximations may prove more efficacious the models suffer accordingly (Green & Silverman, 1994; Ross, 1990; Seber & Wild, 1989). An example of a situation in which neuropsychologists confront conditions where noise could confound a linear association is the testing individuals with physical conditions that preclude standardized administration of tests or when testing environments face external interruptions. Nonlinear associations, not necessarily clearly understood but likely present, include age-related changes in cognition (Lineweaver & Hertzog, 1998) and differences in the qualitative characteristics of memories (Qin, Raye, Johnson, & Mitchell, 2001).

### Artificial Neural Network

To offset these deficiencies, artificial neural networks (ANNs) can be used. ANNs exhibit robust flexibility in the face of dimensionality problems that hamper attempts to model nonlinear functions with large numbers of variables (Geman, Bienenstock, & Doursat, 1992; Wasserman, 1989). Though noisy input causes refined degradation of function and can result in failure of the GLM, ANNs can still respond appropriately given their nonlinear proficiencies (Lippman, 1987). ANNs are also well adapted for problems that require the resolution of many conflicting constraints in parallel (Bishop, 1995; Pao, 1989; Soucek, 1992). Although GLMs are capable of multiple constraint satisfaction, ANNs have been found to provide more unaffected measures for dealing with such problems (Hertz et al., 1991).

Backpropagation is the most popular ANN (BN-ANN) methodology in use today (Cherkassky & Lari-Najafi, 1992; Dayhoff, 1990; Fausett, 1994; Fu, 1994; Rumelhart & McClelland, 1986; Zurada, 1992). This popularity has resulted from the ANNs' ability to provide robust nonlinear modeling and their availability in commercial ANN shells (Medsker & Liebowitz, 1994; Schocken & Ariav, 1994). The BP-ANN is based upon the multilayer perceptrons (MLPs) originally developed by Rumelhart and McClelland (1986) and is discussed at length in most neural network texts (e.g., Bishop, 1995). Like regression, the BP-ANN makes use of a weighted sum of their inputs (predictors). The configuration of a BP-ANN allows it to adjust its weights to new circumstances. The BP-ANN consists of a system of interconnected artificial neurons (nodes) made up of three groups, or layers, of units: a layer of "input" units is connected to a layer of "hidden" units, which is connected to a layer of "output" units. Input units (predictors) are "weighted," to create hidden units. Hidden unit activity is determined by weighted connections between input and hidden units. Hence, the effect each input (predictor) has on the output (criterion) is dependent upon the weight of a particular input. An input weight is a quantity which when multiplied with the input gives the weighted input. If the sum of weighted inputs exceeds a preset threshold value the "neuron" fires

$(X_1W_1 + X_2W_2 + X_3W_3 + \dots > T)$ . In any other case the neuron does not fire. The BP\_ANN differs from the GLM in that it runs multiple simulation runs in which the weights of the net are continually adjusted and updated to reflect the relative importance of different patterns of input. Eventually, the trained system generates the (unknown) function that relates input and output variables, and can subsequently be used to make predictions where the output is not known (Hinton, 1992; Ripley, 1993).

A BP\_ANN with only one input layer (single-layer perceptron) functions in a manner analogous to a simple linear regression (SLR). SLR fits a straight line through one predictor and criterion by the method of least squares. This fit is used to test the null hypothesis that the slope is 0. Likewise, each neuron in the BP\_ANN adjusts its weights according to the predicted output and the actual output using the “perceptron delta rule”:  $[\Delta w_i = x_i\delta]$  where  $[\delta]$  (delta) is the desired output minus the actual output. A single-layer BP\_ANN uses an activation function that sums the total net input and outputs “1” if this sum is above a threshold, and “0” otherwise.

A BP\_ANN, with multiple layers, functions in a manner analogous to that of a multiple linear regression (MR). MR fits a criterion as a linear combination of multiple predictors by the method of least squares. Likewise, the extension of the single-layer perceptron to a multi-layer perceptron requires delta level modifications to avoid

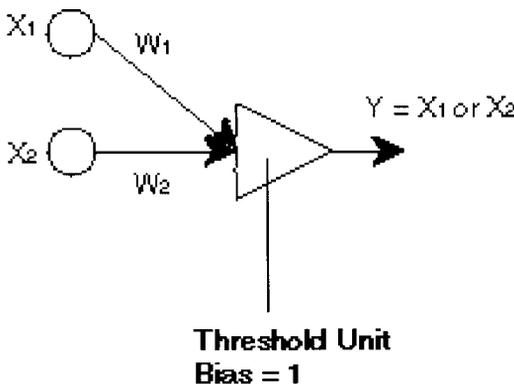


Fig. 1. Sigmoid function plateaus at 0 and 1 on the y-axis, and crosses the y-axis at 0.5.

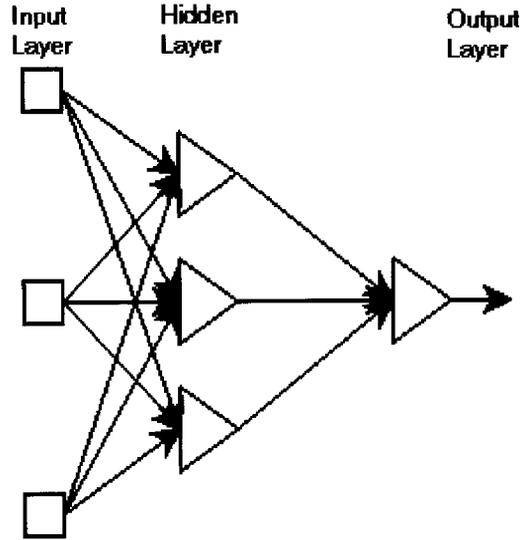


Fig. 2. Single-layer network using the “perceptron delta rule.”

nonlinearly separable problems (see Minsky & Papert, 1969; Rumelhart & McClelland, 1986). Weight adjustments anywhere in the network necessitate a deduction of the effect said adjustment will have on the overall outcome of the network. The multi-layered network makes use of the backpropagated delta rule. This is a further development of the simple delta rule, in which a hidden layer is added. Here, the input layer connects to a hidden layer (more than one hidden layer can be used if desired). The hidden layer (interconnects with other hidden layers if present) learns to provide a representation for the inputs through an alteration of the weights and then connects to the output layer.

Weight alteration depends upon an amount proportional to the error at a given unit multiplied by the yield of the unit connecting into the weight. One must look at the derivative of the error function with respect to a given weight. Weighted information is summed and presented to a pre-set “activation function” (threshold value). Alterations in weights require an existing point on the error surface to descend into a vale of the error surface.

This “gradient descent” occurs in a direction that corresponds to the steepest gradient or slope at the existing point on the error surface

(Kindermann & Linden, 1990). However, the descent of total error into a vale of the error surface may not lead to the lowest point on the entire error surface. Consequently, it may become “trapped in a local minimum.” Further, if the gradient is very steep, it approaches a hard limiter function (a sigmoid with infinite). Of special importance at this juncture is that the hard-limiter function for the perceptron is non-continuous, thus non-differentiable. To deal with this problem, a sigmoid function is used, in which the function plateaus out at 0 and 1 on the y-axis, and crosses the y-axis at 0.5, making the function relatively easy to differentiate. A sigmoid function (or squashing function) introduces nonlinearity into input mapping, in which low inputs are mapped near the minimum activation, high inputs are mapped close to the maximum activation, and intermediate inputs are mapped nonlinearly between the activation limits. The sigmoid function is not the only squashing function used in ANNs. Other functions, such as Gaussian and tanh can be used, but sigmoid is the most common and is therefore chosen here. As a result, the earlier formula for the delta rule ( $\Delta w_i = x_i \delta$ ) receives the addition of a constant  $y = 1/(1 + e^{-x})$ . This allows one to look at the derivative of the error function with respect to a given weight.

The network’s calculation of hidden layer error requires a further addendum to a definition of  $[\delta]$ . This supplement is important because the researcher needs to know the effect on the output of the neuron if a weight is to change. Therefore,

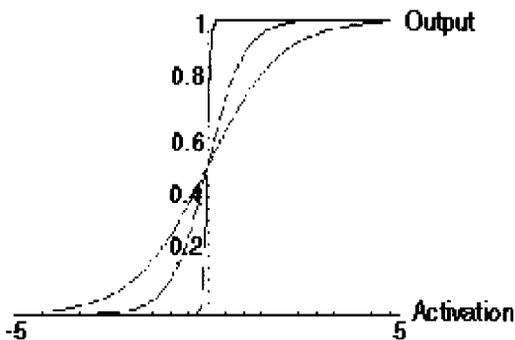


Fig. 3. Multi-layered network using the “perceptron delta rule.”

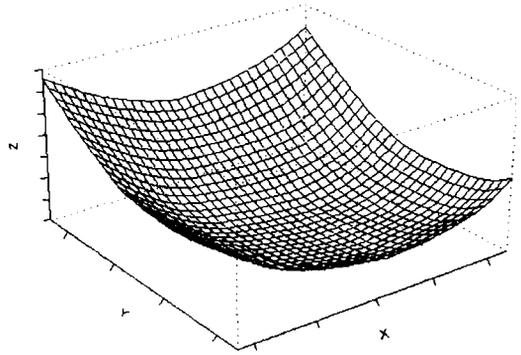


Fig. 4. Error surface.

the researcher needs to know the derivative of the error with respect to that weight. To find this, the researcher analyses the backpropagation learning  $[\delta_{p+1}]$ , in which each hidden layer’s  $[\delta]$  value requires that the  $[\delta]$  value for the layer after it be calculated. It is important that the learning rate  $[\eta]$  (eta) is kept to a minimum so that the backpropagation accurately follows the path of steepest descent on the error surface. In a multi-layered network, “backpropagation” is viewed as the error from the output layer that is slowly propagated backwards through the network through the following process: (a) first, the output-layer’s  $[\delta]$  is calculated using the first  $[\delta]$  formula shown, (b) next, this value is used to calculate the remaining hidden layers using the formula shown above.

ANNs appear to offer a promising alternative to standard regression techniques. However, their usefulness for neuropsychological research is limited if researchers present only prediction results and do not present features of the underlying process relating the inputs to the output (Barron & Barron, 1988; Geman et al., 1992; Ripley, 1993, 1996). A foundational necessity for any data analytic strategy incorporated by a neuropsychological researcher is always an empirical confirmation (Kibler & Langley, 1988). In order for the neuropsychological researcher to be certain that the portion that he or she is able to observe is representative of the whole number of events in question the procedures of statistical inference will need to be incorporated. This allows researchers to draw conclusions from the evidence provided by samples. Through the use of statistical testing,

researchers can be ensured that the observed effects on the dependent variables are caused by the varied independent variables and not by mere chance. Consequently, statistical evaluation of neural network research is fundamental.

In summary, the backpropagated algorithm includes a feed-forward transmission, in which the outputs are computed and the output unit(s) error is determined. Next, there is a backward dissemination, in which the error of the output unit is exercised to revise weights on the output units. Finally, a backpropagation of output unit error through the weights determines the hidden nodes and their weights are altered. This process is a recursive process that occurs until the error is at a low enough level. Currently, interpretability of the backpropagated ANN necessitates reincorporating it back into the parametric model.

### Comparison of ANNs and the GLM

This study aims to compare the performance of regression models with that of ANNs. In the analysis, both classes of models will be used to model data with various distributional properties. To perform this kind of research, neuropsychological researchers advocate (e.g., Hogarth, 1986) testing alternative models side by side in critical experiments. There is precedent for this kind of study using ANNs (Fisher & McKusick, 1989; Weiss & Kapouleas, 1989) and in statistics (Paarsch, 1984; Pendleton, Newman & Marshall, 1983). Thus, this experiment is a side-by-side comparison of two competing methods.

## METHOD

An exemplary analytic problem in neuropsychology is to understand what contributes to performance in a specific domain. We used both the general linear model's multiple regression and the artificial neural network model's backpropagated algorithm to compare the performance of these two analytic methods.

### Participants

Thirty community dwelling older adults (15 men and 15 women) between the ages of 64 and 86 (Mean age = 73.6) participated in the present study. Participants consisted mainly of volunteers from the Andrus

Gerontology Center at the University of Southern California and resided in the greater Los Angeles area. Participants were paid \$50.00 for their participation in the study. The average level of education was 15.4 years. None reported a history of any neurological condition. All were screened for cognitive functioning with the Telephone Interview of Cognitive Status (TICS) and all scored above 31. Welsh, Breitner, and Magruder-Habib (1993) have reported that no cases of dementia have been observed among individuals scoring above 31 on the TICS.

### Tests

The neuropsychological test battery included: Trails A; Block Design from the Wechsler Adult Intelligence Scale – Revised (Wechsler, 1981); Long Delay Free Recall from the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987), Visual Reproduction II (VR II) test from the Wechsler Memory Scale – Revised (Wechsler, 1987); and the Judgment of Line Orientation (JLO; Benton, Varney, & Hamsner, 1978).

### Data Analysis

To compare results from two analytic procedures (GLM vs. ANN) – used to test the hypothesis that processing speed substantially reduces or eliminates age-related variance in memory measures – a multiple regression was first performed. Next, we trained a BP\_ANN and calculated its output-layer's delta. This value was then used to calculate the remaining hidden layers. The layered BP\_ANN's adjusted outputs were compared with the results of the multiple regression analysis.

In order to increase the probability of generalization and to avoid the overfitting of the observed sample, we considered three data sets: (a) the "training set" was used to develop estimates of the network's weights for prediction; (b) the "validation set" was used to assess the predictive ability of the network on sample units that had not been considered in the training; and (c) the "test set" was used to calculate the global predictive ability of the network for generalizations to future practical applications.

Following Kindermann and Linden (1990), we used a gradient descent technique (in our BP\_ANN) to minimize least squared error and avoid getting "trapped in a local minima." To accomplish this, we adjusted nodes in the BP\_ANN's hidden layer. To assure that the BP\_ANN got as close as possible to true (absolute) minimum error, we followed Maghami and Sparks's (2000) findings that one should build a BP\_ANN with one hidden layer and continually double the number of nodes until the error is no longer reduced.

After the development and implementation of the BP\_ANN, we compared its output and that of the

GLM's regression by performing the following tasks in hierarchical order: (a) the same predictor data was input into both systems; (b) the criterion from the BP\_ANN was recorded; (c) predictor set and the criterion output from the BP\_ANN were input into a new regression analysis; (d) standard error of the estimate and  $R^2$  were computed from the BP\_ANN and regression; (e) the results of (d) were compared with the straightforward regression analysis; (f) the variance of the standard error of the estimates was noted to determine if the difference was statistically significant – the model with the smallest standard error of the estimate was considered preferable.

In an effort to be efficient in our comparison, we also used a significance test of the difference between the independent  $B$ s of the backpropagated ANN versus those of the GLM. This test of significance was done with reference to the rationale that the differences found between the backpropagated ANN and the GLM may be found in the delta rule's adjustment of the backpropagated ANN's weights. Consequently, we tested the significance of the differences between the  $b$ s using a significance test of the difference between two proportions.

## RESULTS

### Training

In the preliminary tests to assure that the ANN achieved its optimal point, we experimented with networks containing 3, 6, 12, and 24 nodes in the single hidden layer. We found the improvement in error after 6 nodes insignificant, while the processing speed and convergence rate were significantly worse. Given these results and our small

sample size, we chose the network with 3 interior nodes to be most appropriate over all conditions. Thus, the ANN structure implemented in this exercise consisted of 5 input nodes, 1 output node and 3 nodes in a single hidden layer (5-3-1 network;  $\eta = 0.35$ ). The neural network weights were adjusted following the presentation of each  $(x, y)$  pattern. Convergence was reached in 500 training epochs.

### Generalization

Descriptive statistics for all tests are shown in Table 1. The results from the regression and neural network are represented in Table 2. The results from the significance test comparing the respective independent  $b$ s of the backpropagated ANN versus those of the GLM are presented in Table 3. When analyzing results from the multiple regression, the model (using Trails A as criterion) included five predictors: age ( $b = .69, p = .32$ ), Block Design ( $b = .24, p = -.55$ ), CVLT ( $b = 1.31, p = .23$ , VRII test ( $b = -.28, p = .20$ ) and JLO ( $b = .41, p = .68$ ). Further, results revealed an  $R^2 = .21, p < .28$ ; and a Standard Error of estimate = 18.01.

When analyzing the BP\_ANN, the model (using Trails A as criterion; and corrected for BP\_ANN training) included five predictors: age ( $b = .81, p < .11$ ), Block Design ( $b = .03, p = .89$ ), CVLT ( $b = 1.36, p = .09$ ), VRII test ( $b = -0.36, p = .02$ ); and JLO ( $b = .22, p = .76$ ). Further, results revealed an  $R^2 = .39, p < .02$ ; and a Standard Error of estimate = 13.07.

Table 1. Descriptive Statistics for Neuropsychological Tests.

| Test                            | Mean  | SD    | Range  |
|---------------------------------|-------|-------|--------|
| Judgment of Line Orientation    |       |       |        |
| Raw Score                       | 24.2  | 5.40  | 10–30  |
| Trails A                        | 39.9  | 18.49 | 17–102 |
| Block Design                    | 32.03 | 12.97 | 0–65   |
| Visual Reproduction II          |       |       |        |
| Raw Score                       | 46.43 | 24.15 | 0–87   |
| California Verbal Learning Test |       |       |        |
| List A – Long Delay Free Recall | 8.10  | 3.80  | 0–15   |

Note. For all analyses,  $N = 30$ .

Table 2. Processing Speed Results From the Regression and Neural Network.

| Test         | Processing speed regression | Processing speed corrected for BP_ANN |
|--------------|-----------------------------|---------------------------------------|
| Age          | .69<br><i>ns</i>            | .81<br><i>ns</i>                      |
| Block Design | .24<br><i>ns</i>            | .03<br><i>ns</i>                      |
| CVLT LD Free | 1.31<br><i>ns</i>           | 1.36<br><i>ns</i>                     |
| VR II        | -.28<br><i>ns</i>           | -.36<br>.02                           |
| JLO          | .41<br><i>ns</i>            | .22<br><i>ns</i>                      |

Note. For all analyses,  $N = 30$ .

Table 3. Significance Test Comparing Independent  $b$ s of ANN Versus GLM.

| Test         | Regression $b$ | BP_ANN $b$ | $p$ |
|--------------|----------------|------------|-----|
| Age          | .69            | .81        | .14 |
| Block Design | .24            | .03        | .01 |
| CVLT LD Free | 1.31           | 1.36       | .5  |
| VR II        | -.28           | -.36       | .25 |
| JLO          | .41            | .22        | .05 |

Note. For all analyses,  $N = 30$ .

## DISCUSSION

In conclusion, the research reported demonstrated the applicability of the BP\_ANN. With a simple multiple-layered, fully connected BP\_ANN typology (5-3-1 with systematically selected network parameters, a learning rate of 0.35, and about 500 epochs) this research illustrated that the BP\_ANN can perform better than regression in both prediction and generalization. Although the reported regression analysis provided us with an adequate understanding of our data, the regression model's normality and independence of error variance restrictions may limit its ability to predict and generalize under nonlinear conditions. Contrariwise, our backpropagated ANN possesses the property to learn from a set of data without the need for a full specification of the decision model.

When compared to the GLM's multiple regression analysis, the BP\_ANN was found to proffer an 18% increase in prediction of a common neuropsychological problem. A possible reason for this increase in predictability may be found in the BP\_ANN's ability to learn from new examples and generalize. Their ability to adjust the interconnectivity of weight coefficients between neurons results in error (between the computed output dependent vector and the known dependent vector of the trained patterns) to be minimized. The training process of the BP\_ANN transmits backward the error to the network and adjusts the weights between the units connecting the output layer and the hidden layer and the hidden layer and the input layer.

In situations where age-related changes in the cognitive system are associated with a decline in some general and fundamental mechanism, all of the age-related variance in cognitive variables may be shared by a single common factor (Verhaeghen & Salthouse, 1997). If this is the case, the age-related influences on many cognitive variables may be caused by the same factor. Although a multiple regression analysis will not work well with such non-independence of error variance, ANNs can see through noise and irrelevant data and are comparatively robust and fault tolerant. Consequently, ANNs are better able to identify patterns between predictors and criterions in a data set – they are not as affected (as the GLM) by nonlinear transformations and data discontinuities.

A possible drawback of applying the ANN approach is that the current techniques for development of high-quality neural networks are not effortless tasks. In fact, the multiple regression method is a much more straightforward method and requires less human judgment than does a backpropagation model. However, as Darlington (1968) has pointed out, the regression model tends to be one of the most abused statistical methods, in which, tests are routinely performed and inferences made without verifying whether the assumptions of regression such as normality and independence of error variance are satisfied. Hence, there are situations where regression is more appropriate than a trained system and the use of ANN could be inappropriate as well.

Despite familiarity with regression there appear to be compelling reasons why neuropsychologists should consider incorporating ANNs into their analytic repertoire. Linear regression imposes a linear form on the mapping function that can limit its accuracy. However, cognition is clearly not limited to linear associations. Consequently, utilitarian linear regression models typically necessitates a transformation of the variables to make the relationship between independent and dependent variables linear – commonly through “dummy coding.” The transformation of variables to make the data linear can theoretically enable linear regression to be as accurate as any statistical model. However, the achievement of this goal in problems of any complexity is an arduous task that may result in violations of the linear model’s assumptions. If the neuropsychological researcher is unable to locate and resolve nonlinearities, the linear regression model will not aid the data analytic process. Further, since all the variables must be understood as an inter-related group, the use of linear regression on complex problems can lead to correlated error and erroneous results. The ANN automates the process of deciding the shape that the mapping function should have. Further, the ANN offers a statistical modeling technique that uses the data set to model the shape of a complex and flexible mapping function. Although some researchers may desire to move from standard linear regression (straight lines) to polynomial and logistic regression (simple curves), or to the arduous task of spline regression, we argue that a preferable solution is the ANN methodology because it can take on any form the data requires. On a theoretical level it can be argued that linear regression imposes a linear form on the mapping function that can limit its accuracy. This methodology, while novel, has concrete applications to frequent neuropsychological associations likely containing nonlinearities, for example, aging and cognition.

Any discussion of the adoption of ANNs for use in neuropsychological research leads to questions related to the ways in which researchers can develop an architectonic methodology for ANN training and analysis that does not require the biomedical researcher to be a computer

specialist. Other issues that arise for the application of ANNs to neuropsychological research include: network weight testing, network optimization, and determination of the neuropsychological significance of network weights relative to the backpropagated ANNs hidden layers. Again, it seems possible that the development and evolution of ANNs will result in architectonic procedures that will allow neuropsychological researchers reasonably to evaluate data given differing conditions. Further comparisons of ANNs and conventional methods from the general linear model, should aid in researchers’ understandings of the ways in which topology and parameters may be automated and selected. The resulting work that would need to be done, then, includes the development of a methodology that allows the ANN to “learn” incrementally without major network re-training when new neuropsychological information becomes available.

The research found in this paper presented a discussion of the developmental process of training, recall, and generalization of ANNs for a neuropsychological application. Further, this paper had as its goal the elucidation of specific details about backpropagation, neuropsychological experimentation, and resulting hazards of which the researcher needs to be aware. A further goal of this research was to explicate the potential use of ANNs as data analytical tools for the increasingly complex endeavors of neuropsychological research.

## REFERENCES

- Barron, A.R., & Barron, R.L. (1988). Statistical learning networks: A unifying view. In E. Wegman, (Ed.), *Proceedings of the 20th Symposium on the Interface of Statistics and Computing Science*, American Statistical Association, Washington, DC, 192–203.
- Benton, A.L., Varney, N.R., & Hamsher, K.D. (1978). Visuospatial judgment: A clinical test. *Archives of Neurology*, 35, 364–367.
- Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford: University Press.
- Box, G.E.P. (1966). The use and abuse of regression. *Technometrics*, 8, 625–629.
- Cherkassky, V., & Lari-Najafi, H. (1992). Data Representation for Diagnostic Neural Networks. *IEEE Expert*, 7, 43–53.

- Darlington, R.B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69, 161–182.
- Dayhoff, J. (1990). *Neural network architectures: An introduction*. New York: Van Nostrand Reinhold.
- Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). *The California verbal learning test*. San Antonio, Texas: Psychological Corporation.
- Dempster, A.P. (1973). Alternatives to least squares in multiple regression. *Multivariate statistical inference* (pp. 25–40). In D.G. Kabe & R.P. Gupta (Eds.), Amsterdam: North-Holland Publishing Company.
- Fausett, L. (1994). *Fundamentals of neural networks: Architectures, algorithms, and applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Fisher, D., & McKusick, K. (1989). An empirical comparison of ID3 and back-propagation. *Proceedings of the International Joint Conference on Artificial Intelligence*, 788–793.
- Fu, L. (1994). *Neural networks in computer intelligence*. New York: McGraw-Hill.
- Gallant, A.R. (1987). *Nonlinear statistical models*. New York: Wiley.
- Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4, 1–58.
- Gordon, R.A. (1968). Issues in multiple regression. *American Journal of Sociology*, 73, 592–616.
- Green, P.J., & Silverman, B.W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. London: Chapman-Hall.
- Haerdle, W. (1990). *Applied nonparametric regression*. Cambridge: University Press.
- Hertz, J., Krogh, A., & Palmer, R.G. (1991). *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley.
- Hinton, G.E. (1992). How neural networks learn from experience. *Scientific American*, 267, 144–151.
- Hogarth, R.M. (1986). Generalization in decision research: The role of formal models. *IEEE Transactions on Systems, Man, and Cybernetics*, 16, 439–449.
- Kachigan, S.K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods*. New York: Reditus Press.
- Kibler, D., & Langley, P. (1988). Machine learning as an experimental science. *Machine Learning*, 3, 5–8.
- Kindermann, J., & Linden, A. (1990). Inversion of neural networks by gradient descent. *Parallel Computing*, 14, 277–286.
- Licht, M.H. (1995). Multiple regression and correlation. In L.G. Grimm & P.R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 19–64). Washington, DC: American Psychological Association.
- Lineveaver, T.T., & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging Neuropsychology and Cognition*, 5, 264–296.
- Lippmann, R.P. (1987). An introduction to computing with neural nets. *IEEE Transactions ASSP*, 4, 4–22.
- Maghami, P., & Sparks, D. (2000). Design of neural networks for fast convergence and accuracy: Dynamics and control. *IEEE Transactions Neural Networks*, 11, 113–123.
- McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London: Chapman-Hall.
- Medsker, L., & Liebowitz, J. (1994). *Design and development of expert systems and neural networks*. New York: Macmillan.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Mosteller, F., & Tukey, J.W. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.
- Neter, J., Wasserman, W., & Kutner, M.H. (1989). *Applied linear regression models* (2nd ed.). Homewood, IL: Irwin.
- Paarsch, H.J. (1984). A Monte Carlo comparison of estimators for censored regression models. *Journal of Econometrics*, 24, 197–213.
- Pao, Y. (1989). *Adaptive pattern recognition and neural networks*. Reading, MA: Addison-Wesley.
- Pendleton, B.F., Newman, I., & Marshall, R.S. (1983). A Monte Carlo approach to correlational spuriousness and ratio variables. *Journal of Statistical Computation and Simulation*, 18, 93–124.
- Qin, J., Raye, C.L., Johnson, M.K., & Mitchell, K.J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1110–1115.
- Ripley, B.D. (1993). Statistical Aspects of Neural Networks. In O.E. Barndorff-Nielsen, J.L. Jensen, & W.S. Kendall (Eds.), *Networks and chaos: Statistical and probabilistic aspects* (pp. 170–185). London: Chapman-Hall.
- Ripley, B.D. (1996). *Pattern recognition and neural networks*. New York: Cambridge University Press.
- Ross, G.J.S. (1990). *Nonlinear estimation*. New York: Springer-Verlag.
- Rumelhart, D.E., & McClelland, J. (Eds.). (1986). *Parallel distributed processing* (Vol. 1). Cambridge, MA: Massachusetts Institute of Technology Press.
- Schocken, S., & Ariav, G. (1994). Neural networks for decision support: Problems and opportunities. *Decision Support Systems*, 11, 393–414.
- Seber, G.A.F., & Wild, C.J. (1989). *Nonlinear regression*. New York: Wiley.

- Soucek, B. (1992). *Fast learning and in-variant object recognition: The sixth-generation breakthrough*. New York: Wiley.
- Tukey, J.W. (1975). Instead of Applied statistics Gauss-Markov Least Squares; What? In R.P. Gupta (Ed.), Amsterdam-New York: North Holland Publishing Company.
- Verhaeghen, P., & Salthouse, T.A. (1997). Meta-analysis of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, 122, 231–249.
- Wasserman, P.D. (1989). *Neural Computing: Theory and Practice*. New York: Van Nostrand Reinhold.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised*. New York: The Psychological Corporation.
- Wechsler, D. (1987). *Wechsler Memory Scale – Revised. Manual*. San Antonio: The Psychological Corporation.
- Weiss, S.M., & Kapouleas, I. (1989). An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. *Proceedings of the International Joint Conference on Artificial Intelligence*, 781–787.
- Welsh, K.A., Breitner, J.C.S., & Magruder-Habib, K.M. (1993). Detection of dementia in the elderly using telephone screening of cognitive status. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 6, 103–110.
- Zurada, J. (1992). *Introduction to artificial neural systems*. St. Paul, MN: West Publishing Company.