Journal of the Nigerian Statistical Association, Vol. **29**, 2017, 110– 116

Investigation of effect of number of hidden neurons in Statistical Neural Network

T. O. James[a], S. U. Gulumbe[b] and A. Danbaba[c]

[a]*Department of Mathematics, Kebbi State University of Science and Techonogy, Aliero, Nigeria*; [b,c]*Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria*

One of the most important characteristics of a perception network is the number of neurons in the hidden layer(s) and researchers have had problem in determining the number of hidden units to obtain optimal network performance. This study employed Statistical Neural Network (SNN) using hidden neuron in predicting mother to child transmission of HIV. The data obtained from ANC of a certain hospital and simulated data were used for the study. Mean Square Error (MSE), Akaike Information Criteria (AIC) and Neural Information Criteria (NIC) were computed for determining the adequacy of the hidden neuron that determines the optimality of the model. The best result shows that the hidden neuron that determines the optimality of the model in prediction of HIV with CD4 as an input variable with 7 neurons is SNN (1-2-1).

Keywords: Statistical Neural Networks (SNN); Hidden Neuron (HN); hidden layers; child HIV status; CD4.

1. Introduction

Optimizing the number of hidden neurons to use without a pre-set target for accuracy is one of the major challenges for neural networks, usually referred to as the bias/variance dilemma (Geman,1992). Artificial neural networks (ANN) were inspired by biological findings relating to the behavior of the brain as a network of units called neurons (Rumelhart. et.al., 1986). An ANN is a network consisting of neurons and paths joining the neurons (Bishop, 1995). ANNs can model, especially when underlying data relationship is unknown, it identify and learn correlated patterns between data sets and corresponding target values. After training, ANNs can be used to predict the outcome of new independent input data. Figure 1.1 depicts an individual neuron (in the hidden and output layers). Each neuron, j, has a number of input arcs, x_i to x_n. Associated with each arc,i, is a weight, w_{ij} which represents a factor by which any values passing into the neuron are multiplied. A neuron,j, sums the values of all inputs according to equation (1.3).

Hidden neurons (HNs) are data transmitting channels within the summing junction (that is, between the input neurons and transfer functions, and between transfer functions and output neurons, HNs are located in Hidden units). See Figure 1. While hidden layers are intermediate layers between the input and the output layer. Many researchers have had problem in determining the number of hidden units to obtain optimal network performance. A test can be used in solving this problem. Empirical results have shown that with higher neurons, the network error is reduced. However, if care is not taken, one may be tempted to increase the hidden neurons indiscriminately and when this happens, over fitting occurs. Also, too few hidden neurons leads to error bias, which sometimes can be very embarrassing, thus making neural network statistically unfit. Putting a redundant hidden neuron into the network can over fit the coefficient of determination (Udomboso, et.al 2012). Different researcher assert that hidden nodes is chosen arbitrarily with different result and that a large number of HN will ensure correct learning, and the network would be able to correctly predict the data it has been trained on, while with too few HNs, the network may be unable to learn the relationships amongst the data and error will fail to fall below an acceptable level. That is,when an adequate number of neurons are used, the network will be able to model complex data, the training time will not become excessively long and the network

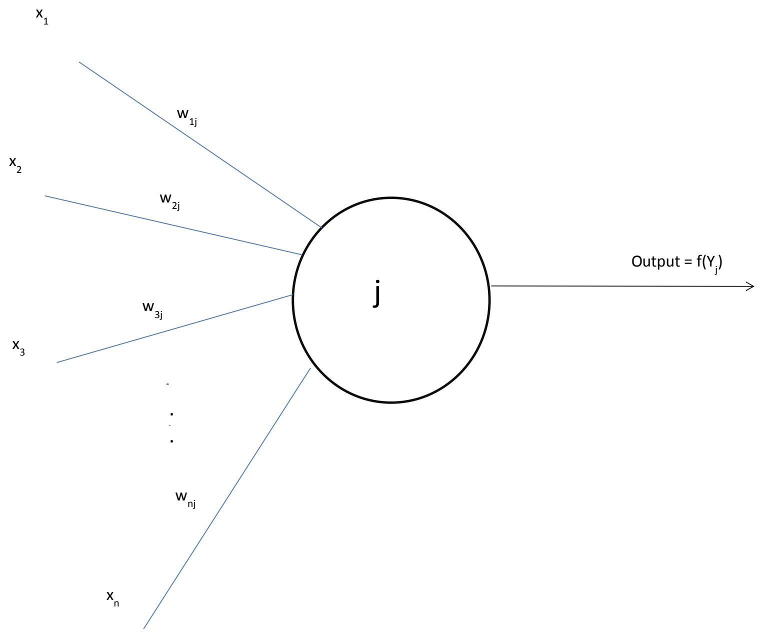

Figure 1. An artificial neuron

will not model random noise and over-fit the data, resulting in a good fit. (Lee and Park, 2001, Abraham, 2005, Sibanda and Pretorius, 2011). Also, Ganesen. et al.,(2000) opined that a common approach is to start with large number of nodes and then employ network pruning algorithms for optimization. While Jeff (2009) gives some useful rules-of-thumb in determining the correct number of neurons to be included in the hidden layer of the network. Deciding the number of neurons in the hidden layers is a very important part of deciding the overall neural network architecture, which is a crucial decision. Therefore, this study investigates the effect of hidden neurons in prediction of HIV status of children using neural network.

2. Related literature

Several researchers have studied and investigated on the approach of determining optimum number of hidden neurons in data mining and the effect of varying the number of neurons in feed forward back propagation NN Architecture and prediction applications like prediction of cancer cells. (Onoda, 1995, Keeni, et al. 1999, Wu and Hong, 2007, Xu and Chen, 2008, Gnana and Deepa, 2013, Imran, et al., 2017).

2.1 *Human Immune Deficiency Virus (HIV)*

AIDS is one of the deadliest epidemics in human history. It was first identified in 1981 in New York and California (USA). Shortly after its detection in the United States, AIDS quickly developed into a worldwide epidemic, affecting virtually every nation (Abdalla, 2011). HIV infects cells in the immune system and the central nervous system. The T-helper lymphocytes are the main type of cell that HIV infects. The role of the cells in the immune system is to coordinate the actions of other immune system cells. HIV infects the T-helper cells because it has the protein called CD+ lymphocyte. CD4+ T cells are essential in B cell antibody class switching, in the activation and growth of CD8+ cytotoxic T cells, and in maximizing bactericidal activity of phagocytes such as macrophages.Once it attaches itself into a cell, HIV produces new copies which are capable of infecting other cells. When the age of infections increases, HIV infection leads to a severe reduction in the number of T-helper cells which are responsible to help fight diseases. When the CD4 T-cell counts decline below a critical level (i.e with the loss of cell mediated immunity), the overall immune system fails to hinder the growth of HIV and the body becomes progressively more susceptible to opportunistic.

2.2 *Performance analysis*

The performance of the back-propagation neural network is evaluated by means of three statistics: the mean squared errors (MSE), Akaike Information Criterion (AIC) and Network Information Criterion (NIC). They are defined as follows:

$$MSE = \frac{\sum(y_i - \bar{y}_i)^2}{n} \tag{1}$$

$$AIC = e^{\frac{2k}{n}} \frac{SSE}{n} \tag{2}$$

Taking natural logarithm,

$$InAIC = \frac{2k}{n} + In\frac{SSE}{n} \tag{3}$$

where$k$ is the number of independent variables (including the intercept) and n is the number of observations

$$NIC \equiv [q, p(\hat{w})] \tag{4}$$

Where q is the input distribution and $p(\hat{w})$ is the conditional distribution

3. Methodology

3.1 *The statistical neural network*

The Statistical Neural Network (SNN) model structurally is composed of two parts: the predictive and the residual, as is in classical regression, given as

$$y = f(X, w) + e_i \tag{5}$$

But, $f(X, w) = \alpha X + \sum_{h=1}^{H} \beta_h g(\sum_{i=0}^{I} (\gamma_{hi} x_i)$.

where y is the dependent variable, $X = (x_0 \equiv 1, x_1, , x_I)$ is a vector of independent variables, $w = (\alpha, \beta, \gamma)$ is the network weight: α is the weight of the input unit, β is the weight of the hidden unit, and γ is the weight of the output unit, and e_i is the stochastic term that is normally distributed (that is, $e_i \sim N(0, \sigma^2 I)$. The assumptions of the statistical neural network are the same as the usual regression models.

$f(X, w)$ is the artificial neural network function, thus equation (5) can expressed as

$$y = \alpha X + \sum_{h=1}^{H} \beta_h g \left(\sum_{i=0}^{I} \gamma_{hi} x_i \right) + u \tag{6}$$

$X = (x_0, x_1, x_2, , x_n)$, is a vector of input variables,$g(.)$ is the transfer function of a neuron. In this paper hyperbolic tangent function was used as the neuron transfer function.
Let $g2(.)$ = Hyperbolic Tangent function (TANH),
Then TANH be defined as

$$\tanh = f_2(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}} \tag{7}$$

3.2 *Data used*

The data used in this study are from the Antenatal Clinic of an hospital in Sokoto state of Nigeria from 2008 to 2011, the input variable was CD4 count.

3.3 *Model design*

In the model design, only one input was used in this study, it was standardized, that is, converting it to the range $(0, 1)$ before feeding it into the network. This is to avoid the application of extremely small weighting factors in the case of large input values. Also, the output values are 'destandardized' to provide meaningful results since all values leaving the network are automatically output in a standardized format.

The input layer was the CD4 count which distributes the input to the seven hidden layers. The outputs from the hidden layer then become the inputs to the output layer, which provides the network output. Training MLP is an iterative adaptive work. It was trained for 1000 epochs. The training starts from 1 hidden neuron till optimality is obtained at the $7HN$. In each epoch the entire training set is presented to the network. Errors are calculated and used to adjust the weights in the network and after reaching a satisfactory level of performance, it learns the relationship between the input variable (CD4 count) and output variable (Child Status). The trained Back Propagation (BP) is then adopted as a model to predict children as either HIV positive or negative. The weights multiply the input information. Let input be denoted byX_i, and each weight w_i, then the activation is equal to

$$\Sigma x_i w_{ij} \tag{8}$$

For choosing the number and the best set of HNs, a comparison is made for many cases by first starting the training in SNN from 1 hidden neuron till optimality (local minimum) is obtained at the least MSE that is when any additional HN increases the MSE and the $7HNs$ indicated the HN at which optimality is obtained. Thus the best one is selected from the following results of SNN model. It was implemented using MATLAB neural network tool (nntool).

In Table 1.1, the best Hidden neuron (HN) that determine the optimality of the model prediction is the one with the least MSE as well as least AIC and NIC which is 2 hidden neurons (HN) and hidden neurons HN 6 gave local minimum in the optimality trend.

Table 1.2 shows that ANN (1-2-1) is selected as the best hidden neuron for predicting MTCT of HIV, since from the MSE criterion, the model with the least or minimum MSE

Table 1.1: HNs based on MSE, AIC and SIC

SNN			
HN	MSE	AIC	NIC
1	2.0566	2.1388	2.2217
2	0.0910	0.0946	0.1645
3	1.1168	1.1614	1.1741
4	0.8064	0.8386	1.2685
5	0.2794	0.2906	0.4152
6	0.1129	0.1129	0.1871
7	0.5016	0.5216	1.3557

Table 1.2: Model selection based on MSE

Variable (Hidden Neuron)	Architectures	MSE
1	1-1-1	2.0566
2	1-2-1	0.0910
3	1-3-1	1.1168
4	1-4-1	0.8064
5	1-5-1	0.2794
6	1-6-1	0.1129
7	1-7-1	0.5016

should be selected.

4. Discussion

The study determined the optimal statistical neural network (SNN) model using hidden neurons in predicting mother to child transmission of HIV. The model was examined by one (I) variable (CD4 count). Table 1.1 shows that the values of AIC (as well as NIC) for SNN is 2HN and 6HN gave the least. Also Table 1.2, shows that SNN (1-2-1) is selected as the best hidden neuron for predicting MTCT of HIV, since from the MSE criterion, the model with the least or minimum MSE should be selected.

5. Conclusion

We investigated the effect of number of Hidden Neurons in SNN. It was seen that in Estimation (training) in ANN the number of HN to be used starts from 1 hidden neuron till optimality (local minimum) is obtained. Optimality is obtained at the least MSE (when any additional HN increases the MSE). The best HNs that determine the optimality of the model selected is based on the least MSE, AIC and NIC value. Therefore, it was concluded that no HN is assumed.

References

Abdalla, S. W. (2011). Mathematical modelling of HIV/AIDS dynamics with treatment and vertical transmission. *M.Sc. Dissertation*, University of Dar es Salaam.

Abraham, T. (2005). Application of data mining technology to identify determinant risk factors of HIV infection and to find their association rules: the case of Center for Disease Control and Prevention (CDC). *Master's Thesis*, Addis Ababa University, Addis Ababa.

Bishop C. M. (1995). *Neural Networks for Pattern Recognition.* Oxford: Oxford University Press

Ganesan, N., Venkatesh, K. and Palan, A. M. (2010). Application of neural networks in diagnosing cancer disease using demographic data. *International Journal of Computer Application* :1-14.

Geman, S., Bienenstock, E. and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1), 1-58.

Gnana, S. K. and Deepa, S. N., (2013). Review on methods to fix number of hidden neurons in neural networks. *Journal of Mathematical Problems in Engineering.* doi.org/10.1155/425740

Shafi, I., Ahmad, J., Shah, S. I., and Kashif, F. M. (2017). Impact of varying neurons and hidden layers in neural network architecture for a time frequency application. www.researchgate.net/publication/224703002

Jeff, H. (2009). *Introduction to Neural Networks.* MIT Press, Cambridge, MA.

Keeni, K., Nakayama, K. and Shimodaira, H. (1999). Estimation of initial weights and hidden units for fast learning of multi-layer neural networks for pattern classification. *Proceedings of the International Joint Conference on Neural Networks*, 3: 1652–1656.

Kwak, N. K. and Lee, C.(1997). A neural network application to classification of health status of HIV/AIDS patient. *Journal of Medical Systems*, 21(2).

Lee C.W. and Park J.A. (2000). Assessment of HIV/AIDS-related health performance using an artificial neural network. *Information & Management*, 38(4) 231–238.

Onoda, T. (1995). Neural network information criterion for the optimal number of hidden units. *Proceedings of the 1995 IEEE International Conference on Neural Networks*, 1: 275-280.

Rumelhart, D. E, Hinton, G.E. and McClelland, J.L. (eds) (1986). Learning internal representations by error propagation. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, pp: 318-362.

Sibanda, W. and Pretorius, P., (2011). Novel application of multi-layer perceptrons neural networks to model HIV in South Africa using seroprevalence data from antenatal clinics. *International Journal of Computer Applications*, 35 (5), 26-31.

Udomboso, C. G., James, T. O. and Mba, O. O. (2012). On R2 contribution and statistical inference of the change in the hidden and input units of the statistical neural networks. *Journal of the Nigeria Association of Mathematical Physics*, 22.

UNAIDS (2010). Report on the global AIDS epidemic 2010. Geneva: Joint United Nations Programme on HIV/AIDS.

WHO (2009). AIDS epidemic update.

Wu, Y. K. and Hong, J. S. (2007). A literature review of wind forecasting technology in the world. *Proceedings of the IEEE Lausanne Power Tech*, pp. 504509.

Xu, S. and Chen, L. (2008). A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining. *Proceedings of the 5th International Conference on Information Technology and Applications*, pp. 683-686.