Learning Capability of a Simple Neural Network

Dranreb Earl Juanico, Christopher Monterola, and Caesar Saloma*
Instrumentation Physics Laboratory, National Institute of Physics
University of the Philippines Diliman, 1101 Quezon City, Philippines
E-mail: csaloma@nip.upd.edu.ph

ABSTRACT

We demonstrate that a neural network composed of only three nodes and three connections arranged in a 2-inputs, 1-middle, 1-output architecture is able to perform differentiation of univariate functions (**Ð**). Using a proposed empirical technique, we assess the network's generalization capability by approximating a functional form for the growth function $\Delta_F(N)$. We calculated the probability of error to be $\approx 10^{-7}$ allowing us to justify the effectiveness of the simplistic approach in modeling a non-trivial task such as **Ð**.

Keywords: architecture, differentiation, generalization, growth function

INTRODUCTION

The inherent complexity of nature channels efforts to propose theories of brain function into the following directions: (1) The complexity of the brain tells us that it is important to first understand some basic principles of information processing that it uses for its interaction with the environment. The principles can then guide the quest towards a detailed one-to-one understanding of the brain. (2) The complexity of the environment, which by far exceeds even the complexity of the brain tells us that whatever analysis the brain is doing, it must have found algorithms that achieve this task very efficiently (Stetter & Obermayer, 1999).

The fundamental problem of relating structure and function in biological neural systems provides impetus to simulating solutions by means of artificial neural networks. A significant research interest for artificial neural networks is the search for fast, approximate solutions to non-trivial problems by a network of simple processors providing local and partial solutions in parallel. In this spirit, attempts should be made to avoid fully connected feedforward networks (Bose & Liang, 1996).

The rationale of this study is mainly to address the issue of modeling specialized functions of the brain in terms of simple neural networks characterized by small numbers of nodes and connections. One particular task, namely, differentiation of univariate functions (**Ð**) is tackled in this study. The task **Ð** is crucial in the brain's recognition and detection of rates of change, for instance, motion, which is the rate of change of position of a body under observation.

METHODOLOGY

The feedforward architecture shown in Fig. 1 is implemented to perform **Ð**. The network was trained using a set consisting of $N = 79$ random values of x uniformly distributed over the interval (-1,1). The shaded circles denote the input nodes; f is the sample function, and δ acts as a numerical tolerance. The parameter p_1 and p_2 are the weights of the connection of these input nodes to the middle node.

*corresponding author

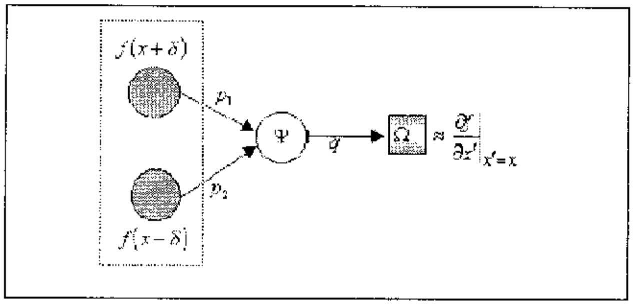

Fig. 1. Schematic diagram of the neural network

The middle node is characterized by a *hyperbolic-tangent* activation: $\Psi = \alpha\ tanh(x)$, wherein α denotes amplitude. The weight parameter q links the middle node to the output node Ω. The neural network is trained iteratively (in epochs) over the N exemplars until the total error is no greater than a preferred stopping criterion $|\varepsilon_{se}| < 1$.

The goal after training phase is to assess the generalization performance of the neural network by virtue of testing it with novel sets of data not included in the training set, i.e., with a new function $f^{\dagger}$ over a wider interval $(-R, R)$.

RESULTS AND DISCUSSION

After a network has successfully learned a set of N training exemplars, the usefulness of the learned network depends on the accuracy of the network's predictions of the output for future exemplars (Bose & Liang, 1996).

The neural network was trained to differentiate $cos(x)$ over a unit interval for 500 epochs (Fig. 2). The hope after training is for the network to be able to learn and generalize a non-linear mapping that transforms any function into its derivative over a bounded two-dimensional window. This transformation is not trivial at all. To demonstrate the generalization capability of the network after training, we tested it by using new functions and extending the domain set (Fig. 3).

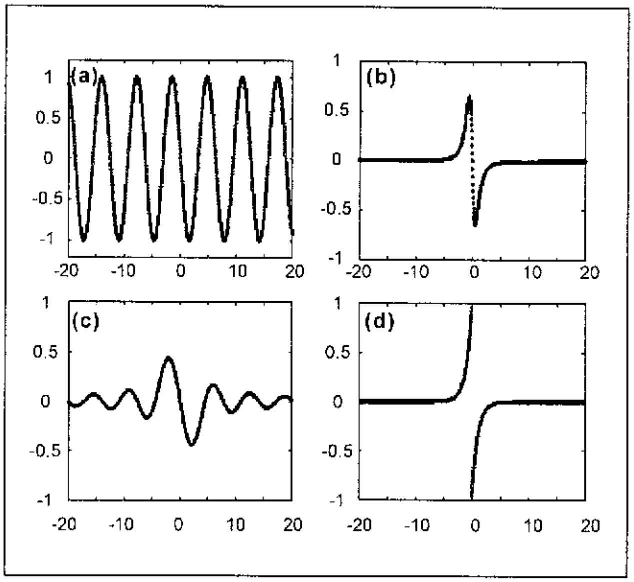

Fig. 3. Derivatives of some important functions in physics: 3a. Cos (x); 3b. Lorentzian function; 3c. Sine function; 3d. Cauchy function over $x \in (-20, 20)$

Assessment of the generalization capability is theoretically anchored in Vapnik-Chervonenkis theory (Bose & Liang, 1996). We use the result of Baum & Haussler (1989), to estimate a measure, which tells us how confident one can be about the performance of the network on future exemplars.

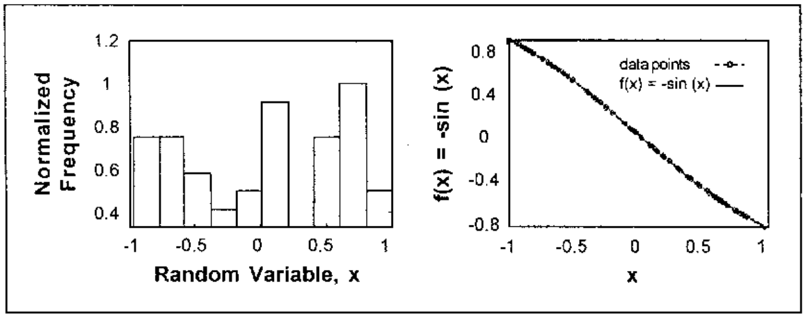

Fig. 2a. Distribution of the training variable; Fig. 2b. Training set consisting of input and desired output. Training function is cos(x) over the unit interval (-1,1).

A useful bound is the probability of error $e(\phi)$ for the mapping ϕ realized by the trained network. Let **F** be the class of functions on $\Re^2$. If N exemplars $(\mathbf{x}, \phi(\mathbf{x}))$, $\mathbf{x} \in \Re^2$ were randomly chosen according to uniform probability distribution (Fig. 2a), then the probability that there exists a mapping $\phi \in \mathbf{F}$ such that $e(\phi) > \varepsilon$ but ϕ disagrees with only a fraction $(1-\gamma)\varepsilon$ of the training exemplars is bounded by the inequality (Vapnik, 1982)

$$\Pr[e(\phi) > e | e(\phi) \le (1-\gamma)e \le 8\Delta_F(N)\exp(-\gamma^2 eN/4) \quad (1)$$

The growth function $D_F(N)$ is not easy to determine analytically especially for networks that utilize non-linear activation functions and output real instead of discrete (binary) values. In this spirit, we propose to find an approximate analytic expression for the growth function in order to analyze the network's generalization ability. Our motivation stems from the property that for simple networks with a small number of free parameters, the training error is a "good" predictor for the generalization error (Cortes, 1995; Vapnik, 1982). Thus, through an empirical technique it is possible to deduce the functional form of $D_F(N)$ by plotting the average training error $<e_t>$ over t trials versus different sizes of the training set N (Fig. 4).

The simple technique that we employed yields this important relation,

$$\Delta_F(N) \approx 0.05N \log N \quad (2)$$

By considering the worst-case wherein the probability of error yields,

$$\Pr[e(\phi) > e] \le 10^{-7}$$

Such a small magnitude on the upper bound for the probability of error guarantees that the network generalizes sufficiently for any new exemplars. This implies the property that although the network was only trained to differentiate $cos(x)$ over (-1,1), it is capable of performing the same task on any univariate function over $(-R,R)$ where $R > 1$. This is illustrated in Fig. 3.

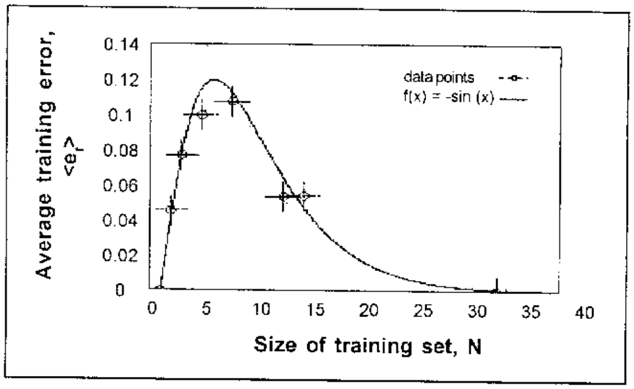

Fig. 4. The growth function approximately varies as 0.05 N log_N

CONCLUSION

We have shown that a simple neural network is able to perform as essential a task as univariate differentiation. The growth function $D_F(N)$ was determined to be proportional to $N \log N$. This enabled the calculation of the probability of error, which measures how well the network performs on new data it has not been trained with. Although the model that we have presented here is oversimplified, the result of this study undoubtedly tells us that it is possible to address the problem of structure and function — the formidable challenge presented by neuronal heterogeneity to any theoretical approach to network analysis or to the construction of relevant wiring diagrams of even the most simple circuits in the brain.

ACKNOWLEDGMENT

This project is partially supported by CHED.

REFERENCES

Baum, E. & D. Haussler, 1989. What size net gives valid generalization? *Neural Computation.* 1:151-160.

Bose, N. & P. Liang, 1996. Neural Network fundamentals with graphs, algorithms, and applications. Singapore, McGraw-Hill, Inc.

Cortes, C., 1995. Prediction of generalization ability in learning machines. Doctoral Thesis, Dept. of Computer Science, University of Rochester.

Husken, M., C.Goecker, & A. Voegel, 2000. Fast adaptation of the solution of differential equations to changing constraints. Report: Institut fur Neuroninformatik.

Stetter, M. & K. Obermayer, 1999. Biology and theory of early vision in mammals. *Informatik* 2(1): 1.

Vapnik, V., 1982. Estimation of dependence based on empirical data. Berlin, Springer-Verlag.