ESAT - SISTA REPORT I 39, 1992
+ C 39 IFAC Australia 1993

Nonlinear system identification using multilayer neural networks: some ideas for initial weights, number of hidden neurons and error criteria

Johan Suykens, Bart De Moor

K.U. Leuven - Department Electrical Engineering - ESAT
Kardinaal Mercierlaan 94, B-3001 Leuven, BELGIUM
tel: 32/16/22 09 31 - fax: 32/16/22 18 55
e-mail: suykens@esat.kuleuven.ac.be

Abstract

New ideas are introduced in this paper related to some hard problems in the field of nonlinear system identification by multilayer neural networks: the choice of initial weights and the number of hidden neurons to be selected. For the initial weights results of linear identification techniques like linear least squares (LLS) and total linear least squares (TLLS) are given here. A good choice for the number of hidden neurons is obtained by means of singular value decomposition techniques (SVD). Finally, 'simulation error' criteria are proposed instead of commonly used 'fitting errors'.

Keywords: nonlinear system identification, multilayer neural networks, singular value decomposition (SVD), linear least squares (LLS), total least squares (TLLS)

1 Introduction

It is well known that multilayer neural networks with one or more hidden layers can approximate any continuous nonlinear function arbitrarily well [1]. For that reason they have become a popular solution to nonlinear identification problems. The interconnection weights of the neurons are the solution to a nonlinear least squares problem that is usually solved by means of backpropagation, recursive prediction error algorithms [2] or quasi-Newton methods [3]. Comparative studies show that learning of the weights based on information up to the second order Hessian matrix (or its approximation by first order gradient information) converges a lot faster than algorithms only based on first order gradient information.
However many open problems exist about the choice of initial weights and the number of hidden neurons of the neural network. In this paper we will propose new ideas and algorithms to cope with these problems. In linear estimation theory several solutions to linear identification problems are formulated like linear least squares (LLS), and total linear least squares (TLLS) [4]. Retaining only the first order terms of a Taylor expansion of the activation function, the nonlinear multilayer neural network model is approximated by a linear model. Linear identification techniques provide then a solid foundation for the choice of the initial weights.
For the selection of the number of hidden neurons we propose an algorithm based on SVD (Singular Value Decomposition). Finally, instead of the classical nonlinear least squares criteria we propose so called 'simulation error' criteria, where we optimize a simulation result of interest instead of the 'fitting error'. The presented theory is illustrated on *real* measurements of a ball and beam system.

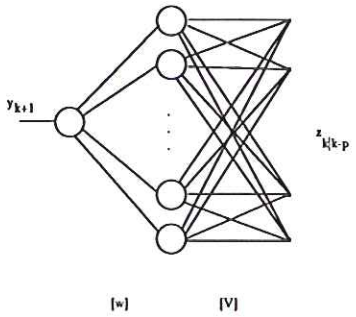

Figure 1: *Neural network model*

2 Neural network model

Consider a SISO system represented by the following model

$$y_{k+1} = f(z_{k|k-p}) \, , z_{k|k-p} = [y_k \dots y_{k-p} \, u_k \dots u_{k-p}]^t$$

where p is the maximum lag in the input and output. In the case of one hidden layer we take the model (see Fig.1)

$$y_{k+1} = w^t.tanh(V.z_{k|k-p}) \tag{1}$$

The interconnection weights $w(n_1 \times 1)$ and $V(n_1 \times n_2)$ of the neurons are usually determined as the solution to a nonlinear least squares problem

$$min_{w,V} \;\; J_N(w,V) \;=\; \frac{1}{2N} \; || \, T_l - w^t.tanh(V.T_r) \, ||_2^2 \tag{2}$$

with N the number of samples. For the case $p = 2$ the matrices T_l, T_r are given by

$$T_l = [\; y_4 \quad y_5 \quad y_6 \quad \dots \quad y_{N+3} \;]$$

$$T_r = \begin{bmatrix} y_3 & y_4 & y_5 & \dots & y_{N+2} \\ y_2 & y_3 & y_4 & \dots & y_{N+1} \\ y_1 & y_2 & y_3 & \dots & y_N \\ u_3 & u_4 & u_5 & \dots & u_{N+2} \\ u_2 & u_3 & u_4 & \dots & u_{N+1} \\ u_1 & u_2 & u_3 & \dots & u_N \end{bmatrix}$$

Suykens J.A.K., De Moor B., ``Nonlinear system identification using multilayer neural networks: some ideas for initial weights, number of hidden neurons and error criteria", in *Proc. of the 12th IFAC World Congress*, Sydney, Australia, Jul. 1993, pp. 49-52., Lirias number: 180767.

3 Simulation error criteria

Instead of the fitting error (2) we propose here a so called 'simulation error' as criterion to identify the interconnection weights. The reason for this is that (2) does not take into account the dynamical structure of the model (1), revealed by the special structure of T_r, but fits the data in a static way. If we want to impose that dynamical structure we can state the problem as follows

$$min_{w,V,\hat{y}_k} \; J_N^{sim}(w, V, \hat{y}_k) = \frac{1}{2N} \sum_{k=1}^{N} \epsilon_k^2 \qquad (3)$$

with

$$\epsilon_k = y_k - \hat{y}_k$$

and constraints

$$\hat{y}_{k+1} = w^t.tanh(V.\hat{z}_{k|k-p}) \; , \forall k$$

A practical algorithm in order to solve this problem is to simulate the model (1) over a predefined time interval and to compute the performance index J_N^{sim} (called the simulation error) based on the simulation result $\{\hat{y}_k\}_{k=1}^N$. The input to the simulation algorithm consists of $w(0), V(0), y_1, y_2, \ldots, y_p, \{u_k\}_{k=1}^N$.
This nonlinear optimization problem can then e.g. be solved by a quasi-Newton method [3]. The choice of the initial weights $w(0)$, $V(0)$ will be discussed in section 4.

4 Number of hidden neurons and initial weights

Our discussion on the selection of the number of hidden neurons n_1 will be related to the fitting error J_N (2) and based on SVD reduction techniques. The following trivial but important remark can be made.

Remark: *If $n_1 = N$ is chosen in (2) and $w^t = T_l.H^{-1}$ with $H = tanh(V.T_r)$ and V is chosen such that H is non-singular, then a fitting error $J_N = 0$ can be achieved.*

The w, V choice here leads to very high generalization errors and is of no practical interest.
With respect to the optimization problem (2) we consider now two special cases. In a first case we fix the matrix V and estimate how many hidden neurons are needed in order to represent the nonlinear mapping. In the second case we linearize the model in (2) and derive the required number of hidden neurons if we start from linear identification results $w(0)$ and $V(0)$ in (3).

4.1 Randomly chosen and fixed V

If we fix V in (2) (choose e.g. a random matrix) we can make the following observation

Observation : *For a given fixed full rank matrix V, the optimal interconnection weights w of the output neuron (in least squares sense) are given by $w_{LS} = \sum_{i=1}^{r} s_i^{-1} . u_i^t T_l^t . v_i$ where (u_i, s_i, v_i) is the i-th singular triplet of the matrix H^t when H is defined as $H = tanh(V.T_r)$ and $r = rank(H)$.*

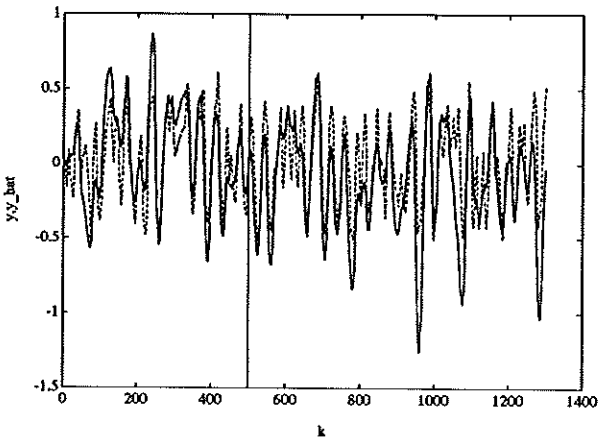

Figure 2: *Optimal simulation result of the nonlinear optimization problem (3) with LLS as starting point. The performance index J_N^{sim} decreased from 0.0395 to 0.0243 in 3000 iterations. (original data = full line, simulation result = dashed line, $N = 500$).*

Hence we can conclude that the proposed number n_1 hidden neurons may be reduced to r in order to represent the mapping.

4.2 Linearization of the model

In order to linearize the model (1) we modify it by a constant α $(0 < \alpha < 1)$ as $H = \frac{1}{\alpha}.tanh(\alpha.V.T_r)$. The following theorem can then be proven

Theorem : *When $\alpha \to 0$ and the data set is scaled as $\mathcal{O}(E\{y_k\}) = 1$, $\mathcal{O}(E\{u_k\}) = 1$ the relation $r = rank(H) \leq n_2$ holds and furthermore $V = I_{n_2}$ may be chosen without loss of generality.*

One can take now e.g. $V(0) = I_{n_2}$, $w(0) = w_{LS}$ or $w(0) = w_{TLS}$ in the optimization problem (3). A concrete example on identification of a ball and beam system is shown in Fig.2, based on *real* measurements.

Acknowledgement

This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the framework of a Concerted Action Project of the Flemish Community, entitled *Applicable Neural Networks*. The scientific responsibility is assumed by its authors.

References

[1] S. Chen, S. Billings, P.M. Grant, *Non-linear system identification using neural networks*, Int. J. Control, Vol.51, No.6, pp.1191-1214, 1990.

[2] K.-I. Funahashi, *On the approximate realization of continuous mappings by neural networks*, Neural networks, Vol.2, pp.183-192, 1989.

[3] P. Gill, W. Murray, M. Wright, *Practical Optimization*, Academic Press, London, 1981.

[4] S. Vanhuffel, J. Vandewalle, *The total least squares problem*, SIAM Philadelphia, 1991.