Statistical Inference for Modeling Neural Network in Multivariate Time Series

Dhoriva Urwatul Wutsqa[1)], Subanar[2)], Suryo Guritno[2)] & Zanzawi Soejoeti[2)]
[1)]*Mathematics Department, Yogyakarta State University, Indonesia*
[2)]*Mathematics Department, Gadjah Mada University, Indonesia*

ABSTRACT

We present a statistical procedure based on hypothesis test to build neural networks model in multivariate time series case. The method involved strategies for specifying the number of hidden units and the input variables in the model using inference of R^2 increment. We draw on forward approach starting from empty model to gain the optimal neural networks model. The empirical study was employed relied on simulation data to examine the effectiveness of inference procedure. The result showed that the statistical inference could be applied successfully for modeling neural networks in multivariate time series analysis.

Keywords : neural networks, R^2 increment, multivariate time series

INTRODUCTION

In daily life, we frequently observe the time series concerning interdependency between different series of variables, which is called vector time series or multivariate time series. The prominent works in that topic are established by Brockwell and Davis (1997). Their results mostly concentrate on linear model and usually require a tight assumption.

Recently, there has been a growing interest in nonlinear modeling. Neural network is a relatively new approach for modeling nonlinear relationship. Numerous publications disclose that neural networks (NN) has successfully applied in data analysis, including in time series analysis (e.g. Che *et al*. 2001, Dhoriva *et al*. 2006, Suhartono 2005, Suhartono & Subanar 2006). NN model becomes popular because of its flexibility, by means that it needs not a firm prerequisite and that it can approximate any Borel-measureable function to an arbitrary degree of accuracy (e.g. Hornik *et al*. 1990, White 1990). However, this flexibility leads to a specification problem of the suitable neural network model. A main issue related to that problem is how to obtain an optimal combination between number of input variables and unit nodes in hidden layer (e.g. Haykin 1999, Subanar *et al*. 2005)

Many researchers have started developing strategy based on statistical approach to model selection for modeling neural network. The concepts of hypothesis testing have been introduced by White, Granger & Terasvirta. (Subanar 2005). The current result is from Kaashoek and Van Dijk (2002) proposing backward method, which is started from simple model and then carry out an algorithm to reduce number of parameters based on R^2 increment and principal component analysis of network residuals criteria until attain an optimal model.

Whereas, Swanson & White (1995, 1996) applied a criterion of model selection, Schwarz Information Criteria (SIC), on "bottom-up" procedure to increase number of unit nodes in hidden layer and select the input variables until finding the best Feed Forward Neural Network (FFNN) model. This procedure is also recognized as "constructive learning", one of the most popular is "cascade correlation" (e.g. Fahlman & Lebiere 1990, Prechelt 1997). It can be seen as "forward" method in statistical modeling.

Appealed to their method, Suhartono *et al*. (2006) put forward the new procedure by using the inference of R^2 increment (as suggested in Kaashoek & Van Dijk 2002) and SIC criteria, which is started from simple model and compare the result with one from Kaashoek & Van Dijk . They found that both procedures gave the same optimal FFNN model, however their new method delivered less number of running steps.

Since all their work is focused in univariate case, it is still an open problem how the procedures work in multivariate case especially in time series modeling. Based on their result, this paper presents the procedure of specifying neural network model for multivariate time series. The method developed here is restricted only for forward approach.

Neural network model

Feed forward neural network (FFNN) is the most widely used NN model in performing time series prediction. Typical FFNN with one hidden layer for univariate case generated from Autoregressive model is called Autoregressive Neural Network (ARNN). In this model, the input layer contains the preceding lags observations, while the output gives the predictive future values. The nonlinear estimating is processed in the hidden layer (layer between input layer and output layer) by a transfer function. Here, we will construct FFNN with one hidden layer for multivariate time series case. The structure of the proposed model is motivated by the generalized space-time autoregressive (GSTAR) model from Lopuhaa & Borokova (2005). The following are the steps of composing our FFNN model .

Supposed the time series process with m variables $Z_t = (Z_{1,t}, Z_{2,t} \dots , Z_{m,t})'$ is influenced by the past p lags values and let n as the number of the observations.

Set design matrix $\underset{\sim}{X}$ = diag($\mathbf{X_1}$, $\mathbf{X_2}$, …, $\mathbf{X}_m$) , output vector $Y = (Y_1', Y_2' \dots , Y_m')'$, parameter vector $\underset{\sim}{\gamma} = (\gamma_1, \gamma_2 \dots , \gamma_m)$ with $\gamma_i = (\gamma_{1,1}^{(i)}, \dots, \gamma_{1,p}^{(i)}, \dots, \gamma_{m,1}^{(i)}, \dots, \gamma_{m,p}^{(i)})$ and error vector $u = (u_1', u_2', \dots, u_m')'$, where

$$\mathbf{X}_i = \begin{pmatrix} Z_{1,p} & \cdots & Z_{1,1} & \cdots & Z_{m,p} & \cdots & Z_{m,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{1,n-1} & \cdots & Z_{1,n-p} & \cdots & Z_{m,n-1} & \cdots & Z_{m,n-p} \end{pmatrix},$$

$$Y_i = \begin{pmatrix} Z_{i,p+1} \\ Z_{i,p+2} \\ \vdots \\ Z_{i,n} \end{pmatrix}, \text{ and } \mathbf{u}_i = \begin{pmatrix} u_{i,p+1} \\ u_{i,p+2} \\ \vdots \\ u_{i,n} \end{pmatrix}$$

Then we have the FFNN model for multivariate time series, which can be expressed as

$$Y = \beta_0 + \sum_{h=1}^{q} \lambda_h \psi(X\gamma) + u \tag{1}$$

where u is an iid multivariate white noise with

$$E(uu' \mid X) = \sigma I,$$

$$E(u \mid X) = 0, \ X = (1, \underset{\sim}{X}), \text{ and}$$

$$\gamma = (\gamma_0, \underset{\sim}{\gamma})'.$$

The functions ψ represents non linear form, used as logistic sigmoid in this paper

$$\psi(X\gamma) = \{1 + \exp(-X\gamma)\}^{-1}.$$

The architecture of this model is illustrated in Figure 1, particularly for bivariate case with input one previous lag (lag 1).

The notations used in Figure 1. are defined as

$\mathbf{Z}^*_t = \begin{pmatrix} \mathbf{Z}_{1,t} \\ \mathbf{Z}_{2,t} \end{pmatrix}$, $\mathbf{Z}_{11,t-1} = \begin{pmatrix} \mathbf{Z}_{1,t-1} \\ \mathbf{0} \end{pmatrix}$, $\mathbf{Z}_{12,t-1} = \begin{pmatrix} \mathbf{Z}_{2,t-1} \\ \mathbf{0} \end{pmatrix}$, $\mathbf{Z}_{21,t-1} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{1,t-1} \end{pmatrix}$, and $\mathbf{Z}_{22,t-1} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{2,t-1} \end{pmatrix}$

Notify that from the above expression (1), we have separated model

$$\mathbf{z}_{i,t} = \sum_{h=1}^{q} \lambda_h \psi_h \left(\sum_{j=1}^{m} \sum_{k=1}^{p} \gamma_{j,k}^{(i)} \mathbf{z}_{j,t-k} + \gamma_0 \right) + \mathbf{u}_{i,t}, \tag{2}$$

for each site $i = 1, \dots, m$.

The procedure of model selection will be constructed based on model (1) instead of model (2), since it needs only one neuron in output layer; thus, it produce simpler architecture, though it can evaluate all the functional relationships simultaneously.

Forward selection procedure

The strategy to obtain the optimal neural network model correspond to the specifying a network architecture, where in multivariate time series case, it involves selecting the appropriate number of hidden units, the order (lags) of input variables included in the model, and the relevant input variables. All selection problems will be dealt with forward procedure through statistical approach.

The design of the forward procedure is adopted from general linear test approach that can be used for nonlinear model as stated by Kutner *et al*. (2004). The procedure entails three basic steps. First, we begin with specification of the simple model from the data, which is also called the reduced or restricted model. In this study, the reduced model is a FFNN model with one hidden unit, i.e.

$$Y = \beta_0 + \lambda_1 \psi(X\gamma) + u \tag{3}$$

To decide whether the model needs hidden unit extension we use the criteria suggested by Kaashoek & Van Dijk (2002), i.e. square of the correlation coefficient. Therefore, we need to compute this value of reduced model, which is formulated as

$$R_R^2 = \frac{(\hat{y}_R' y_R)^2}{(y_R' y_R)(\hat{y}_R' \hat{y}_R)} \tag{4}$$

where $\hat{y}_R$ is the vector of network output points of reduced model. Next, we successively add the hidden unit. The extension model is considered as the complex or full model, i.e. FFNN model (1), starting from hidden units $q = 2$. Then fit the full model and obtain the square of the correlation coefficient R_F^2, with the same formula as (4). The final step is calculating the test statistic:

$$F^* = \frac{R_{(F)}^2 - R_{(R)}^2}{df_R - df_F} \div \frac{(1 - R_F^2)}{df_F} \tag{5}$$

or

$$F^* = \frac{R_{(Incremental)}^2}{df_R - df_F} \div \frac{(1 - R_F^2)}{df_F} .$$

Gujarati (2002) showed that equation (5) is equal to the following expression

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \tag{6}$$

For large *n*, this test statistic (6) and consequently the test statistic (5) are distributed approximately as $F(v1 = df_R - df_F, v2 = df_F)$ when H_0 holds, i.e. additional parameters in full model all equal to 0. This test (5) is applied to decide on the significance of the additional parameter.

Thus the model selection strategy is performed by following the entire steps. As starting point, we determine all the candidate input variables, and then we carry on all the steps sequentially until we find the additional hidden unit does not leads to be significant. Once the optimal number of hidden units is found out, we continue to find the relevant lags that influenced the time series process, begin from the lag that gives the largest R^2. Finally we employ the forward procedure to decide on the significance of the single input, again starting from the input which has the largest R^2. If the process of including the input in the model yields significant *p*-value, then it is included to the model, otherwise it is removed from the model.

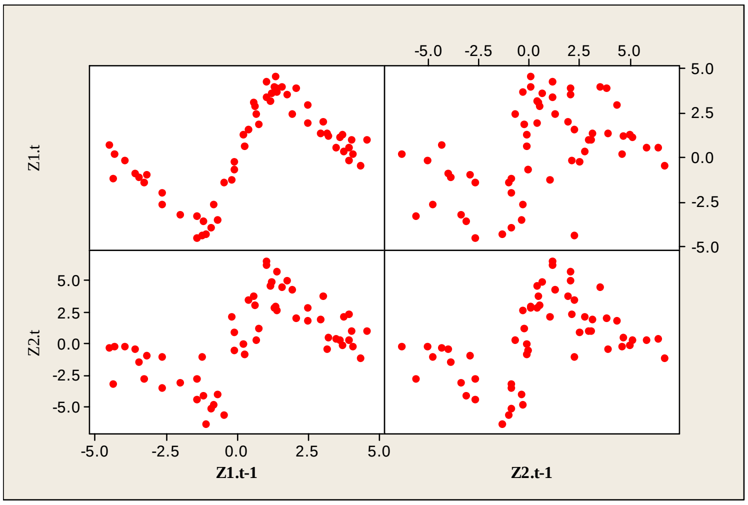
SIMULATION MODEL

In this paper, we study simulation experiment to show how the proposed strategies work to build nonlinear FFNN model for multivariate time series. We consider FFNN model only for bivariate case. The model is derived based on the Exponential Smooth Transition Autoregressive (ESTAR) model for bivariate case, which can be formulated as follows:

$$\begin{aligned} Z_{1,t} &= 4.5\, Z_{1,t-1} . \exp(-0.25\, Z_{1,t-1}^2) + u \\ Z_{2,t} &= 4.7\, Z_{1,t-1} .\exp(-0.35\, Z_{1,t-1}^2) + \\ &\quad 3.7\, Z_{2,t-1} \exp.(-0.25\, Z_{2,t-1}^2) + u \end{aligned} \tag{7}$$

The time process in model (7) depends on one previous lag (lag 1). For convenience, we called them Multivariate Exponential smooth transition autoregressive (MESTAR). The data plots with their previous lags from simulation result of MESTAR are depicted in Figure 1a and 1b.

a



b

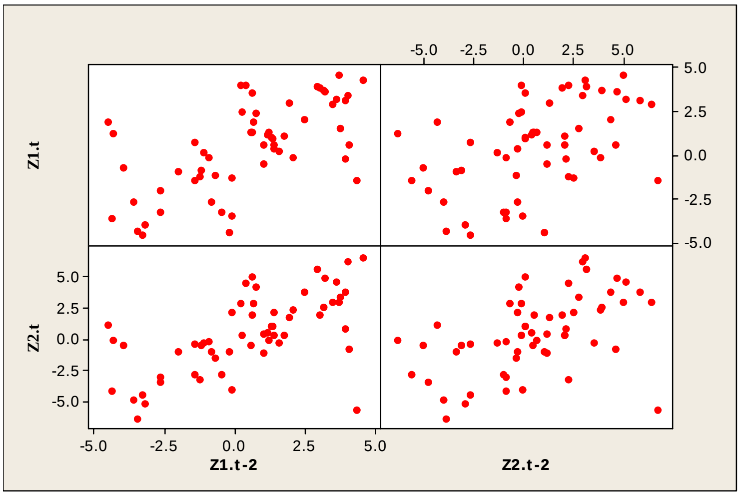


Figure 1. Plots of simulated data corresponding to a) lag 1 and b) to lag 2.

We can perceive from Figure 1a and 1b that the nonlinear autoregressive pattern arise only at lag 1, since the plots tends to spread randomly at lag 2.

We also can observe that the first variable ($Z_{1,t}$) have a strong nonlinear relationship just with $Z_{1,t-1}$, other than the second variable ($Z_{2,t}$) exhibits a nonlinear relationship with both $Z_{1,t-1}$ and $Z_{2,t-1}$.

RESULTS AND DISCUSSION

Considering our simulation model (7), we denote the previous two lags as candidate inputs. So the input layer of NN model (1) has eight units, i.e.

$$\mathbf{Z}_{11,t-1} = \begin{pmatrix} \mathbf{Z}_{1,t-1} \\ \mathbf{0} \end{pmatrix}, \mathbf{Z}_{12,t-1} = \begin{pmatrix} \mathbf{Z}_{2,t-1} \\ \mathbf{0} \end{pmatrix},$$

$$\mathbf{Z}_{11,t-2} = \begin{pmatrix} \mathbf{Z}_{2,t-2} \\ \mathbf{0} \end{pmatrix}, \mathbf{Z}_{12,t-2} = \begin{pmatrix} \mathbf{Z}_{1,t-2} \\ \mathbf{0} \end{pmatrix},$$

$$\mathbf{Z}_{21,t-} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{1,t-1} \end{pmatrix}, \mathbf{Z}_{22,t-1} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{2,t-1} \end{pmatrix},$$

$$\mathbf{Z}_{21,t-2} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{1,t-2} \end{pmatrix}, \text{ and } \mathbf{Z}_{22,t-2} = \begin{pmatrix} \mathbf{0} \\ \mathbf{Z}_{2,t-2} \end{pmatrix}.$$

We continue the forward procedure starting with a FFNN with variable inputs:

$$\left(\mathbf{Z}_{11,t-1},\mathbf{Z}_{12,t-1},\mathbf{Z}_{11,t-2},\mathbf{Z}_{12,t-2},\mathbf{Z}_{21,t-1},\mathbf{Z}_{22,t-1},\mathbf{Z}_{21,t-2},\mathbf{Z}_{22,t-2}\right)$$

and one constant input to find the optimal unit hidden layer cells. The results of an optimization steps are provided in Table 1.

We can see from Table 1 that after three hidden units, the optimization procedure shows the insignificant p-value; thus, we stop the process and determine three unit cells be the optimal result. Next, we proceed the optimization to determine the lags of input. The result find that only one preceding lag (lag 1) includes in the model (Table 2), assigning from the insignificant value of lag 2.

The final step is selecting the appropriate inputs among inputs of lag 1 whose result is presented in Table 3. In this step, we optimize the FFNN models through each input. Their ordered coefficients R2 are given in the first part of Table 3. It is shown that the FFNN model with input Z11,t-1 has the highest square of the correlation coefficient, so it is chosen as the restricted model. Then we employ the forward procedure based on the ordered R2 values. Subsequently the inputs are entered the model. Here the input Z12,t-1 produce insignificant p–value, thus it is not included in the model. Hence, we get the optimal FFNN model with three hidden units and input variables from lag 1:

$$\left(\mathbf{Z}_{11,t-1},\mathbf{Z}_{21,t-1},\mathbf{Z}_{22,t-1}\right)$$

which is appropriate to the simulation model MESTAR (7).

Table 1. The results of forward procedure to get the optimal number of hidden units.

Number of hidden unit	R^2	R^2incremental	F test	p-value
1	0.7194478	-	-	-
2	0.8490907	0.1296429	6.922922	4.709664e-008
3	0.9245464	0.0754557	7.833773	8.791926e-009
4	0.9386959	0.0141495	1.663691	0.105388

Table 2. The results of forward procedure to get the lag of input variables

lags	R^2	$R^2_{\text{incremental}}$	**F test**	**p-value**
1	0.9099576	–	–	–
2	0.6580684	–	–	–
1,2	0.9245464	0.0145888	1.305469	0.2306338

Table 3. The results of forward procedure to get the relevant combination of input variables.

Input	$\boldsymbol{R^2}$	$\boldsymbol{R^2_{\text{incremental}}}$	**F test**	**p-value**
$\mathbf{Z}_{11,t-1}$	0.6131744	–	–	–
$\mathbf{Z}_{12,t-1}$	0.3815016	–	–	–
$\mathbf{Z}_{21,t-1}$	0.5824822	–	–	–
$\mathbf{Z}_{22,t-1}$	0.5488144	–	–	–
$\mathbf{Z}_{11,t-1}$, $\mathbf{Z}_{12,t-1}$	0.6176742	0.0044998	0.307482	0.8199281
$\mathbf{Z}_{11,t-1}$, $\mathbf{Z}_{21,t-1}$	0.8313187	0.2181443	35.02233	1.110223e-15
$\mathbf{Z}_{11,t-1}$, $\mathbf{Z}_{22,t-1}$	0.8101274	0.196953	28.00289	0.002998011
$\mathbf{Z}_{11,t-1}$, $\mathbf{Z}_{21,t-1}$, $\mathbf{Z}_{12,t-1}$	0.834497	0.0031783	0.581279	0.6286504
$\mathbf{Z}_{11,t-1}$, $\mathbf{Z}_{21,t-1}$, $\mathbf{Z}_{22,t-1}$	0.9054434	0.0741247	23.81741	1.032274e-11
$\mathbf{Z}_{11,t-1}$, $\mathbf{Z}_{21,t-1}$,$\mathbf{Z}_{22,t-1}$,$\mathbf{Z}_{12,t-1}$	0.9099576	0.0045142	1.540758	0.2088723

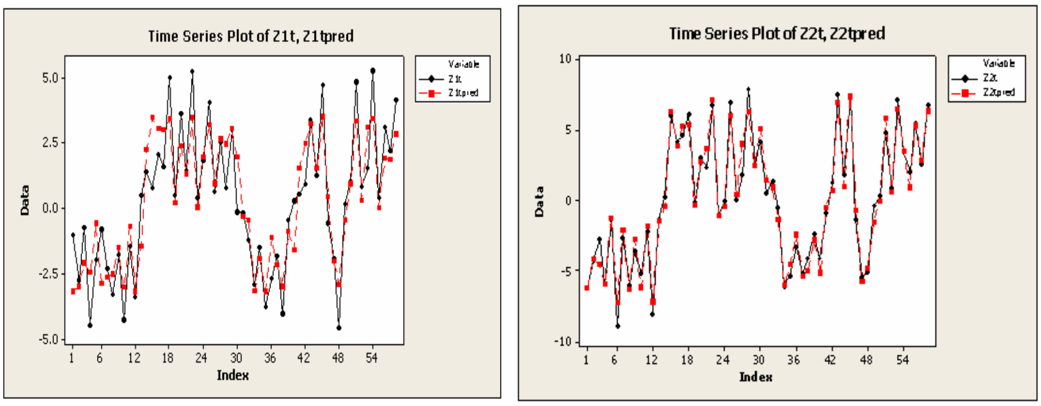


Figure 2. Time series plots of neural network output compared with actual data for each variable.

Additionally, we present the graphs of the resulted network output compared with actual data (Figure 2). It can be seen that the FFNN output fit with actual data. All those evidence demonstrate how the proposed strategy can be applied for selecting the optimal FFNN model.

CONCLUSION

In this study, we recommend model selection procedure for FFNN applied to multivariate time series. The building blocks of FFNN is not in black box, but developed on the basis of statistical concept, i.e. hypothesis testing. We devise forward procedure to determine the optimal number of hidden units and the input variables included in the model by utilizing R^2 incremental inference. The simulation result reveals that the inference procedure can be implemented to obtain the optimal FFNN model for multivariate time series data.

REFERENCES

Brockwell PJ & Davis RA. 1997. *Introducion to Time Series & Forecasting..* New York: Springer Verlag,

Chen X, Racine J, & Swanson NR. 2001. Semiparametric ARX Neural-Network Models with an application to Forecasting Inflation. *IEEE Transaction on Neural Networks* **12**:674-683.

Dhoriva UW, Subanar, Suryo G, & Zanzawi S. 2006. Forecasting Performance of VAR-NN & VARMA Models. Proceeding of Regional Conference on Mathematics, Statistics & Their Applications, Penang, Malaysia.

Fahlman SE & Lebiere C. 1990. The Cascade-Correlation Learning Architecture, in Touretzky, D. S. (ed.) *Advances in Neural Information Processing Systems 2*, Los Altos, CA: Morgan Kaufmann Publishers. pp. 524-532.

Gujarati DN. 2002. *Basic Econometrics*, 6th edition, McGraw Hill International, New York.

Haykin H. 1999. *Neural Networks: A Comprehensive Foundation*, Second edition, Prentice-Hall, Oxford.

Hornik KM. Stinchombe & White H. 1990. Universal approximation of an unknown mapping & its derivatives using multilayer feedforward networks, *Neural Networks*. **3**:551–560.

Kaashoek JF & Van Dijk HK. 2002. Neural Network Pruning Applied to Real Exchange Rate Analysis. *Journal of Forecasting*. **21**:559–577.

Kutner MH, Nachtsheim CJ & Neter J. 2004. *Applied Linear Regression Models*, McGraw Hill International. New York.

Lopuhaa HP & Borovkova S. 2005. Asymptotic properties of least squares estimators in generalized STAR models. *Technical report*. Delft University of Technology.

Prechelt L. 1997. Investigation of the CasCor Family of Learning Algorithms. *Journal of Neural Networks*. **10:**885-896.

Suhartono. 2005. Neural Networks, ARIMA & ARIMAX Models for Forecasting Indonesian Inflation. *Jurnal Widya Manajemen & Akuntansi*. **5**(3):45-65.

Suhartono & Subanar. 2006. The Effect of Decomposition Method as Data Preprocessing on Neural Networks Model For Forecasting Trend & Seasonal Time Series. *JURNAL TEKNIK INDUSTRI: Jurnal Keilmuan dan Aplikasi Teknik Industri*. **9**(2):27-41.

Suhartono, Subanar & Guritno S. 2006. Model Selection in Neural Networks by Using Inference of $R^2_{Incremental}$, PCA, & SIC Criteria for Time Series Forecasting, *JOURNAL OF QUANTITATIVE METHODS: Journal Devoted to The Mathematical & Statistical Application in Various Fields.* **2**(1):41-57.

Subanar, Guritno S & Hartati S. 2005. Neural Network, Pemodelan Statistik dan Peramalan Data Finansial. *Laporan Penelitian HPTP Tahun I*, UGM, Yogyakarta.

Swanson NR & White H. 1995. A model-selection approach to assessing the information in the term structure using linear models & artificial neural networks. *Journal of Business & Economic Statistics.* **13**:265–275.

Swanson NR & White H. 1997. A model-selection approach to real-time macroeconomic forecasting using linear models & artificial neural networks, *Review of Economic & Statistics.* **79:**540–550.

White H. 1990. Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mapping. *Neural Networks.* **3:**535-550.