Generalized Single-Hidden Layer Feedforward Networks

Ning Wang[1,⋆], Min Han[2], Guifeng Yu[1], Meng Joo Er[1,3], Fanchao Meng[1], and Shulei Sun[1]

[1] Marine Engineering College, Dalian Maritime University, Dalian 116026, China
n.wang.dmu.cn@gmail.com
[2] Faculty of EIEE, Dalian University of Technology, Dalian 116024, China
minhan@dlut.edu.cn
[3] School of EEE, Nanyang Technological University, Singapore 639798
emjer@ntu.edu.sg

Abstract. In this paper, we propose a novel generalized single-hidden layer feedforward network (GSLFN) by employing polynomial functions of inputs as output weights connecting randomly generated hidden units with corresponding output nodes. The main contributions are as follows. For arbitrary N distinct observations with n-dimensional inputs, the augmented hidden node output matrix of the GSLFN with L hidden nodes using any infinitely differentiable activation functions consists of L sub-matrix blocks where each includes $n+1$ column vectors. The rank of the augmented hidden output matrix is proved to be no less than that of the SLFN, and thereby contributing to higher approximation performance. Furthermore, under minor constraints on input observations, we rigorously prove that the GLSFN with L hidden nodes can exactly learn $L(n+1)$ arbitrary distinct observations which is $n+1$ times what the SLFN can learn. If the approximation error is allowed, by means of the optimization of output weight coefficients, the GSLFN may require less than $N/(n+1)$ random hidden nodes to estimate targets with high accuracy. Theoretical results of the GSLFN evidently perform significant superiority to that of SLFNs.

Keywords: single-hidden layer feedforward networks, polynomial functions, output weights, hidden node numbers, approximation capability.

1 Introduction

In the field of neural networks, in addition to various fuzzy neural networks [1,2,3], single-hidden layer feedforward networks (SLFNs) have been investigated thoroughly in the past two decades. In the 1990's, it has been shown that SLFNs

⋆ This work is supported by the National Natural Science Foundation of China (under Grant 51009017), Applied Basic Research Funds from Ministry of Transport of P. R. China (under Grant 2012-329-225-060), China Postdoctoral Science Foundation (under Grant 2012M520629), and Fundamental Research Funds for the Central Universities of China (under Grant 2009QN025, 2011JC002 and 3132013025).

C. Guo, Z.-G. Hou, and Z. Zeng (Eds.): ISNN 2013, Part I, LNCS 7951, pp. 91–98, 2013.

with N hidden nodes can exactly learn N distinct observations. Tamura *et al.* [4] proved that a SLFN with N hidden units using sigmoid functions can give any N input-target relations exactly. The further improvement proposed by Huang [5] revealed that if input weights and hidden biases are tunable the SLFN with at most N hidden neurons using any bounded nonlinear activation function which has a limit at one infinity can learn N distinct samples with zero error. In contrast to previous SLFNs which adjust all the parameters of hidden layers, some researchers suggested incremental SLFNs allowing only newly added hidden nodes to be tuned. In this case, parameters of hidden nodes need to be updated only once based on training data. Nevertheless, the computation burden would also be heavy. Alternatively, Huang *et al.* [6] developed an innovative learning scheme termed as extreme learning machine (ELM) for SLFNs with randomly generated hidden units using infinitely differentiable activation functions. Corresponding results [7] indicated that SLFNs with N hidden nodes using any infinitely differentiable activation functions can learn N distinct samples exactly and SLFNs may require less than N hidden nodes if learning error is allowed. Similar to [8], Ferrari *et al.* showed that SLFNs with N sigmoidal hidden nodes and with input weights randomly generated but hidden biases appropriately tuned can exactly learn N distinct observations. Besides, several interesting investigations on compact structure of SLFNs were implemented by using singular value decomposition (SVD) [9] and regularized least-squares (RLS) [10] methods, *etc.* However, all the previous works focused on the SLFNs using constant output weights whether hidden node parameters are adjusted or not. Rationally, we refer to the abovementioned SLFNs as standard SLFNs since all the output weights are confined to be constants independent on inputs. In this case, the constant output weights would impose much deficiency on the capability of approximation and generalization.

In this paper, we propose a novel kind of generalized single-hidden layer feedforward networks (GSLFNs) which extend the standard SLFNs by using polynomial functions of inputs instead of constants as the output weights. To be specific, for arbitrary N distinct observations $(\mathbf{x}_k, \mathbf{t}_k) \in \mathbf{R}^n \times \mathbf{R}^m$, L hidden nodes using any infinitely differentiable activation functions are randomly generated and output weight coefficients are allowed to be adjustable for desired performance of approximation and generalization. In this case, the augmented hidden node output matrix consists of L sub-matrix blocks whereby each one includes $n + 1$ column vectors containing N entities. Each column vector in the ith sub-matrix block is defined by the Hadamard product of the input vector in the j-dimension (i.e., $\mathbf{x}^j = [x_{1j}, \cdots, x_{Nj}]^{\mathrm{T}}$) and the ith hidden node output vector with respect to the kth input observation. Accordingly, preliminary results reveal that the rank of augmented hidden node output matrix in the GSLFN would be no less than that of hidden node output matrix in SLFN, and thereby contributing to higher potentials for approximation capability. Furthermore, we rigorously prove that under minor constraints on input observations the GSLFN with any L randomly generated hidden nodes can exactly learn $L(n+1)$ arbitrary distinct observations which are $n + 1$ times what the SLFN can learn.

2 Preliminaries

Given N arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k)$ where $\mathbf{x}_k = [x_{k1}, x_{k2}, \cdots, x_{kn}]^{\mathrm{T}} \in \mathbf{R}^n$ and $\mathbf{t}_k = [t_{k1}, t_{k2}, \cdots, t_{km}]^{\mathrm{T}} \in \mathbf{R}^m$, the standard single-hidden layer feedforward networks (SLFNs) with L hidden nodes and activation function $g(\mathbf{x})$ can be mathematically modeled as,

$$\mathbf{y}_k = \sum_{i=1}^{L} \boldsymbol{\beta}_i g(\mathbf{a}_i \cdot \mathbf{x}_k + b_i),\ k = 1, 2, \cdots, N \tag{1}$$

where $\mathbf{a}_i = [a_{i1}, a_{i2}, \cdots, a_{in}]^{\mathrm{T}} \in \mathbf{R}^m$ is the weight vector connecting the ith hidden node and the input nodes, $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^{\mathrm{T}} \in \mathbf{R}^m$ is the weight vector connecting the ith hidden node and the output nodes, and b_i is the threshold of the ith hidden node. $\mathbf{a}_i \cdot \mathbf{x}_k$ denotes the inner product of $\mathbf{a}_i$ and $\mathbf{x}_k$.

If the outputs of the SLFN are equal to the targets, we have the compact formulation as follows:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \tag{2}$$

where,

$$\begin{aligned} &\mathbf{H}(\mathbf{a}_1, \ldots, \mathbf{a}_L, b_1, \ldots, b_L, \mathbf{x}_1, \ldots, \mathbf{x}_N) \\ &= \begin{bmatrix} g(\mathbf{a}_1, b_1, \mathbf{x}_1) & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ g(\mathbf{a}_1, b_1, \mathbf{x}_N) & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_N) \end{bmatrix}_{N \times L} \end{aligned} \tag{3}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \tag{4}$$

Here, $\mathbf{H}$ is called the hidden-layer output matrix of the SLFN, whereby the ith column is the ith hidden node's output vector with respect to inputs $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and the jth row is the output vector of the hidden layer with respect to input $\mathbf{x}_j$. $\boldsymbol{\beta}$ and $\mathbf{T}$ are corresponding matrices of output weights and targets, respectively.

It has been proved that standard SLFNs with a wide type of random computational hidden nodes possess the universal approximation capability as follows.

Lemma 1. [6] Given a standard SLFN with N hidden nodes and activation function $g : \mathbf{R}^n \to \mathbf{R}$ which is infinitely differentiable in any interval, for N arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k)$, where $\mathbf{x}_k \in \mathbf{R}^n$ and $\mathbf{t}_k \in \mathbf{R}^m$, for any $\mathbf{a}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, the hidden layer output matrix $\mathbf{H}$ of the SLFN is invertible and $\|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\| = 0$. □

Lemma 2. [6] Given any small positive value $\varepsilon > 0$ and activation function $g : \mathbf{R}^n \rightarrow \mathbf{R}$ which is infinitely differentiable in any interval, there exists $L \leq N$ such that for N arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k)$, where $\mathbf{x}_k \in \mathbf{R}^n$ and $\mathbf{t}_k \in \mathbf{R}^m$, for any $\mathbf{a}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, $\|\mathbf{H}_{N\times L}\boldsymbol{\beta}_{L\times m} - \mathbf{T}_{N\times m}\| < \varepsilon$. □

3 Generalized Single-Hidden Layer Feedforward Networks

We are now in a position to extend standard SLFNs to generalized SLFNs (GSLFNs) by defining the output weights as polynomial functions of input variables (i.e., $\boldsymbol{\beta} \triangleq \boldsymbol{\beta}(\mathbf{x})$) as follows:

$$\beta_{ij}(\mathbf{x}) = w_{ij}^{(0)} + w_{ij}^{(1)}x_1 + \cdots + w_{ij}^{(n)}x_n,\ i = 1, 2, \cdots, L,\ j = 1, 2, \cdots, m \tag{5}$$

where $w_{ij}^{(0)}, w_{ij}^{(1)}, \cdots, w_{ij}^{(n)}$ are corresponding weights for input variables. Accordingly, if the outputs of the GSLFN estimate the targets with zero errors, we obtain the following compact formulation,

$$\mathbf{GW} = \mathbf{T} \tag{6}$$

where,

$$\mathbf{G}(\mathbf{a}_1, \ldots, \mathbf{a}_L, b_1, \ldots, b_L, \mathbf{x}_1, \ldots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{a}_1, b_1, \mathbf{x}_1)\bar{\mathbf{x}}_1^{\mathrm{T}} & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_1)\bar{\mathbf{x}}_1^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ g(\mathbf{a}_1, b_1, \mathbf{x}_N)\bar{\mathbf{x}}_N^{\mathrm{T}} & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_N)\bar{\mathbf{x}}_N^{\mathrm{T}} \end{bmatrix}_{N\times L(n+1)} \tag{7}$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_{11} & \mathbf{w}_{12} & \cdots & \mathbf{w}_{1m} \\ \mathbf{w}_{21} & \mathbf{w}_{22} & \cdots & \mathbf{w}_{2m} \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{w}_{L1} & \mathbf{w}_{L2} & \cdots & \mathbf{w}_{Lm} \end{bmatrix}_{L(n+1)\times m} \tag{8}$$

$$\bar{\mathbf{x}}_k = [1, x_{k1}, x_{k2}, \cdots, x_{kn}]^{\mathrm{T}},\ \mathbf{w}_{ij} = \left[w_{ij}^{(0)}, w_{ij}^{(1)}, \cdots, w_{ij}^{(n)}\right]^{\mathrm{T}} \tag{9}$$

Here, $\mathbf{G}$ is referred to be the augmented hidden-layer output matrix of the GSLFN consisting of $N \times L$ blocks, whereby the kith block $\mathbf{G}_{ki}$ is the product of the ith hidden node output with respect to the kth input vector, i.e., $g_i(\mathbf{x}_k)$, and the corresponding augmented input vector $\bar{\mathbf{x}}_k^{\mathrm{T}}$, and thereby constituting a $[N \times L(n+1)]$-dimension matrix. Accordingly, $\mathbf{W}$ is the output coefficient matrix consisting of $L \times m$ blocks, whereby the block $\mathbf{w}_{ij}$ in the ith-row-jth-column position corresponds to the coefficient vector of the output weight connecting the ith hidden node to the jth output node, and thereby contributing a $[L(n+1) \times m]$-dimension matrix.

4 Main Results

Furthermore, one can obtain the main results on the approximation capabilities of the GSLFN as follows.

Theorem 1. Given a GSLFN with N hidden nodes and activation function $g : \mathbf{R}^n \to \mathbf{R}$ which is infinitely differentiable in any interval, for N arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k)$, where $\mathbf{x}_k \in \mathbf{R}^n$ and $\mathbf{t}_k \in \mathbf{R}^m$, for any $\mathbf{a}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, the rank of the augmented hidden-layer output matrix $\mathbf{G}$ for the GSLFN satisfies $N \leq rank(\mathbf{G}) \leq N(n+1)$, and there exists at least one coefficient matrix $\mathbf{W}$ such that $\|\mathbf{GW} - \mathbf{T}\| = 0$. □

Proof. By Lemma 1, for any $\mathbf{a}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, the hidden-layer output matrix $\mathbf{H}$ satisfies $rank(\mathbf{H}) = N$. In addition, the augmented hidden-layer output matrix $\mathbf{G}$ can be represented by,

$$\mathbf{G} = \mathbf{P} \begin{bmatrix} g(\mathbf{a}_1, b_1, \mathbf{x}_1) & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_1) & g(\mathbf{a}_1, b_1, \mathbf{x}_1)\mathbf{x}_1^{\mathrm{T}} & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_1)\mathbf{x}_1^{\mathrm{T}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ g(\mathbf{a}_1, b_1, \mathbf{x}_N) & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_N) & g(\mathbf{a}_1, b_1, \mathbf{x}_N)\mathbf{x}_N^{\mathrm{T}} & \cdots & g(\mathbf{a}_L, b_L, \mathbf{x}_N)\mathbf{x}_N^{\mathrm{T}} \end{bmatrix} \mathbf{Q}$$

where, $\mathbf{P} \in \mathbf{R}^{N \times N}$ and $\mathbf{Q} \in \mathbf{R}^{N(n+1) \times N(n+1)}$ are elementary transformation matrices. It follows, with probability one, that $N \leq rank(\mathbf{G}) \leq N(n+1)$. In this case, eqn. (6) becomes an under-determined problem with N independent equations since the number of equations is larger than that of unknown parameters. As a consequence, there exists at least one solution for the coefficient matrix $\mathbf{W}$ in (6). This concludes the proof. □

Similar to Lemma 2, for the GSLFN, we can straightforward obtain the following result.

Theorem 2. Given any small positive value $\varepsilon > 0$ and activation function $g : \mathbf{R}^n \to \mathbf{R}$ which is infinitely differentiable in any interval, there exists $L \leq N$ such that for N arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k)$, where $\mathbf{x}_k \in \mathbf{R}^n$ and $\mathbf{t}_k \in \mathbf{R}^m$, for any $\mathbf{a}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, $\|\mathbf{G}_{N \times L(n+1)} \mathbf{W}_{L(n+1) \times m} - \mathbf{T}_{N \times m}\| < \varepsilon$. □

Proof. Following the proof of Theorem 1, the rank of $\mathbf{G}$ would be less than L with high probability. Accordingly, the columns of $\mathbf{G}$ might belong to a subspace of dimension no more than N. In other words, the independent equations of eqn. (6) would be no more than N, and thereby resulting in an under-determined equation with partial targets not be exactly estimated. Fortunately, the tuning of coefficient matrix $\mathbf{W}$ can make the approximation error infinitely small especially when $L = N$. The proof is completed. □

Furthermore, the GSLFN using polynomial functions as output weights features significant characteristics as follows.

Theorem 3. Given a GSLFN with L hidden nodes and activation function $g : \mathbf{R}^n \rightarrow \mathbf{R}$ which is infinitely differentiable in any interval, for $N \leq L(n+1)$ arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k)$, where $\mathbf{x}_k \in \mathbf{R}^n$, $\mathbf{t}_k \in \mathbf{R}^m$ and $x_{kj} \neq x_{k'j'}, \exists k \neq k', j \neq j'$, for any $\mathbf{a}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, the rank of the augmented hidden-layer output matrix $\mathbf{G}$ satisfies,

$$rank(\mathbf{G}) = N \tag{10}$$

and there exists at least one coefficient matrix $\mathbf{W}$ such that $\|\mathbf{GW} - \mathbf{T}\| = 0$. □

Proof. The augmented hidden-layer output matrix $\mathbf{G}$ consists of L sub-matrices $\mathbf{G}_i, i = 1, 2, \cdots, L$, and thereby totally contributing to $L(n+1)$ column vectors given by,

$$\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_L]_{N \times L(n+1)}$$

$$\mathbf{G}_i = \begin{bmatrix} g(\mathbf{a}_i, b_i, \mathbf{x}_1) & g(\mathbf{a}_i, b_i, \mathbf{x}_1)x_{11} & \cdots & g(\mathbf{a}_i, b_i, \mathbf{x}_1)x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ g(\mathbf{a}_i, b_i, \mathbf{x}_N) & g(\mathbf{a}_i, b_i, \mathbf{x}_N)x_{N1} & \cdots & g(\mathbf{a}_i, b_i, \mathbf{x}_N)x_{Nn} \end{bmatrix}$$

Note that $\mathbf{a}_i$ are randomly generated based on a continuous probability distribution, we can assume that $\mathbf{a}_i \cdot \mathbf{x}_k \neq \mathbf{a}_i \cdot \mathbf{x}_{k'}$ for all $k \neq k'$. Consider the jth column of the ith matrix block $\mathbf{G}_i$, i.e.,

$$\mathbf{g}(b_i, \mathbf{x}^j) = \mathbf{c}(b_i) \odot \mathbf{x}^j \tag{11}$$

where $\odot$ denotes the Hadamard product, and

$$\mathbf{c}(b_i) = [g(b_i + d_{i1}), \cdots, g(b_i + d_{iN})]^{\mathrm{T}}, \ \mathbf{x}^j = [x_{1j}, \cdots, x_{Nj}]^{\mathrm{T}}, \ j = 0, 1, \cdots, n$$

where $d_{ik} = \mathbf{a}_i \cdot \mathbf{x}_k$, $b_i \in (a, b) \subset \mathbf{R}$ and $x_{k0} = 1, k = 1, 2, \cdots, N$.

It can be proved by contradiction that vectors $\mathbf{g}$ does not belong to any subspace whose dimension is less than N. Suppose that $\mathbf{g}$ belongs to a subspace of dimension $N-1$. Then there exists a vector $\boldsymbol{\alpha} \neq \mathbf{0}$ which is orthogonal to this subspace, i.e.,

$$\langle \boldsymbol{\alpha}, \mathbf{g}(b_i, \mathbf{x}^j) - \mathbf{g}(a, \mathbf{x}^{j'}) \rangle = 0 \tag{12}$$

Note that, for $N \leq L(n+1)$ arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k) \in \mathbf{R}^n \times \mathbf{R}^m$, $x_{kj} \neq x_{k'j}, \exists k \neq k', j \neq j'$. For the cases $j = j' = 0$ and $j \neq j'$ in (12), we can simply set $b_i + d_{ij} \neq a + d_{ij'}$ and $b_i + d_{ij} = a + d_{ij'}$, respectively. As a consequence, it holds that

$$\mathbf{g}(b_i, \mathbf{x}^j) - \mathbf{g}(a, \mathbf{x}^{j'}) = \begin{cases} (\mathbf{c}(b_i) - \mathbf{c}(a)) \odot \mathbf{x}^j, & j = j' = 0 \\ \mathbf{c}(b_i) \odot \left(\mathbf{x}^j - \mathbf{x}^{j'}\right), & j \neq j' \end{cases}$$

It follows that $\mathbf{g}(b_i, \mathbf{x}^j) - \mathbf{g}(a, \mathbf{x}^{j'}) \neq \mathbf{0}$ is guaranteed since $\mathbf{x}^j \neq \mathbf{x}^{j'}, \forall j, j' \neq 0$ and $\mathbf{x}^0 = \mathbf{1}$.

Using (11), eqn. (12) can be further written as,

$$\alpha_1 g(b_i + d_{i1})x_{1j} + \alpha_2 g(b_i + d_{i2})x_{2j} + \cdots + \alpha_N g(b_i + d_{iN})x_{Nj} - \boldsymbol{\alpha} \cdot \mathbf{g}(a, \mathbf{x}^{j'}) = 0$$

Without loss of generality, we assume $\alpha_N \neq 0$ and obtain,

$$g(b_i + d_{iN})x_{Nj} = -\sum_{k=1}^{N-1} \gamma_k g(b_i + d_{ik})x_{kj} + \boldsymbol{\alpha}' \cdot \left(\mathbf{c}(a) \odot \mathbf{x}^{j'}\right)$$

where $\gamma_k = \alpha_k/\alpha_N$, $k = 1, 2, \cdots, N-1$ and $\boldsymbol{\alpha}' = \boldsymbol{\alpha}/\alpha_N$. Since the activation function $g(.)$ is infinitely differentiable in any interval, we have

$$g^{(l)}(b_i + d_{iN})x_{Nj} = -\sum_{k=1}^{N-1} \gamma_k g^{(l)}(b_i + d_{ik})x_{kj},\ l = 1, 2, \cdots, N, N+1, \cdots$$

where $g^{(l)}$ is the lth derivative of function g of b_i. However, there are only $N-1$ free coefficients: $\gamma_1, \cdots, \gamma_{N-1}$ for the derived more than $N-1$ linear equations, this is contradictory. Thus, vector g does not belong to any subspace whose dimension is less than N.

As a consequence, from any interval (a, b) it is possible to randomly choose $L' = ceil(N/(n+1))$ bias values $b_1, \cdots, b_{L'}$ for the L' hidden nodes such that N vectors $\mathbf{g}(b_i, \mathbf{x}^j), i = 1, 2, \cdots, L', j = 0, 1, \cdots, n$ span $\mathbf{R}^N$. This means that for any input weight vectors $\mathbf{a}_i$ and bias values b_i chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, the column vectors of $\mathbf{G}$ can be made row full-rank, i.e., $rank(\mathbf{G}) = N$ if $N \leq L(n+1)$.

Accordingly, the number of independent equations in eqn. (6) is N, and thereby resulting in a well- or under-determined problem. Hence, there exists at least one solution for $\mathbf{W}$ in (6). This concludes the proof. □

Theorem 4. Given any small positive value $\varepsilon > 0$ and activation function $g : \mathbf{R}^n \to \mathbf{R}$ which is infinitely differentiable in any interval, there exists $L \leq N/(n+1)$ such that for N arbitrary distinct samples $(\mathbf{x}_k, \mathbf{t}_k) \in \mathbf{R}^n \times \mathbf{R}^m$ where $x_{kj} \neq x_{k'j'}, \exists k \neq k', j \neq j'$, for any $\mathbf{a}_i$ and b_i randomly chosen from any intervals of $\mathbf{R}^n$ and $\mathbf{R}$, respectively, according to any continuous probability distribution, then with probability one, $\|\mathbf{G}_{N\times L(n+1)}\mathbf{W}_{L(n+1)\times m} - \mathbf{T}_{N\times m}\| < \varepsilon$. □

Proof. Following the proof of Theorem 3, the rank of $\mathbf{G}$ would no more than N with high probability. In this case, (6) would be an over-determined equation. It means there might not exist exact solutions for $\mathbf{W}$ in (6). Alternatively, given any small error $\varepsilon > 0$ and the GSLFN with $L \leq N(n+1)$ hidden nodes, fine tuning of coefficient matrix $\mathbf{W}$ can make the estimation error less than ε. □

5 Conclusions

This paper extends the standard single-hidden layer feedforward networks (SLFNs) to generalized SLFNs (GSLFNs) by employing the polynomial functions of inputs as output weights. Accordingly, we have rigorously proved the significant characteristics of the GSLFN as follows. On the one hand, similar to the SLFN, the GSLFN with at most N hidden nodes using any infinitely differentiable activation functions can exactly learn N distinct observations. On the other hand, for distinct n-input m-output observations with different data in each input dimension, the GSLFN features much higher approximation capability such that the GSLFN with only $N/(n+1)$ hidden nodes using any infinitely differentiable activation functions can exactly learn N distinct observations. The number of hidden nodes in the GSLFN can be dramatically reduced by using polynomials as output weights, especially for high-dimension regressions and classifications.

References

1. Wang, N., Er, M.J., Meng, X.: A Fast and Accurate Online Self-Organizing Scheme for Parsimonious Fuzzy Neural Networks. Neurocomput. 72, 3818–3829 (2009)
2. Wang, N., Er, M.J., Meng, X.Y., Li, X.: An Online Self-Organizing Scheme for Parsimonious and Accurate Fuzzy Neural Networks. Int. J. Neural Syst. 20, 389–403 (2010)
3. Wang, N.: A Generalized Ellipsoidal Basis Function Based Online Self-Constructing Fuzzy Neural Network. Neural Process. Lett. 34, 13–37 (2011)
4. Tamura, S., Tateishi, M.: Capabilities of a Four-Layered Feedforward Neural Network: Four Layers Versus Three. IEEE Trans. Neural Netw. 8, 251–255 (1997)
5. Huang, G.-B., Babri, H.A.: Upper Bounds on the Number of Hidden Neurons in Feedforward Networks with Arbitrary Bounded Nonlinear Activation Functions. IEEE Trans. Neural Netw. 9, 224–229 (1998)
6. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme Learning Machine: Theory and Applications. Neurocomput. 70, 489–501 (2006)
7. Huang, G.-B., Chen, L., Siew, C.K.: Universal Approximation Using Incremental Constructive Feedforward Networks with Random Hidden Nodes. IEEE Trans. Neural Netw. 17, 879–892 (2006)
8. Ferrari, S., Stengel, R.F.: Smooth Function Approximation Using Neural Networks. IEEE Trans. Neural Netw. 16, 24–38 (2005)
9. Teoh, E.J., Tan, K.C., Xiang, C.: Estimating the Number of Hidden Neurons in a Feedforward Network Using the Singular Value Decomposition. IEEE Trans. Neural Netw. 17, 1623–1629 (2006)
10. Huynh, H.T., Won, Y., Kim, J.J.: An Improvement of Extreme Learning Machine for Compact Single-Hidden-Layer Feedforward Neural Networks. Int. J. Neural Syst. 18, 433–441 (2008)