ARTICLE IN PRESS

Applied Mathematical Modelling xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Applied Mathematical Modelling

journal homepage: www.elsevier.com/locate/apm

Multivariate Jackson-type inequality for a new type neural network approximation ☆

Shaobo Lin *, Yuanhua Rong, Zongben Xu

School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Article history:
Received 18 July 2012
Received in revised form 21 March 2014
Accepted 20 May 2014
Available online xxxx

Keywords:
Neural networks
Jackson-type inequality
Error estimate
Sigmoidal function

ABSTRACT

In this paper, we introduce a new type neural networks by superpositions of a sigmoidal function and study its approximation capability. We investigate the multivariate quantitative constructive approximation of real continuous multivariate functions on a cube by such type neural networks. This approximation is derived by establishing multivariate Jackson-type inequalities involving the multivariate modulus of smoothness of the target function. Our networks require no training in the traditional sense.

1. Introduction

Feed-forward neural networks (FNNs) can be formally described as devices producing input–output functions depending on flexible parameters. Often input–output functions have the form of linear combinations of functions computable by units specific for the given type of networks. Both coefficients of the linear combinations and parameters of the computational units are adjustable in the process of learning. Mathematically, the FNN can be represented by

$$\sum_{i=1}^{M} c_i \sigma(a_i \cdot x + b_i), \quad x \in \mathbf{R}^d, \tag{1.1}$$

where $a_i \in \mathbf{R}^d$, $c_i \in \mathbf{R}$, $b_i \in \mathbf{R}$ are the inner weight, outer weight, and threshold of the FNN, respectively.

Theoretically, any continuous functions defined on a compact set in $\mathbf{R}^d$ can be approximated by neural networks to any desired accuracy by increasing the number of hidden neurons. This result is usually called the density problem of FNNs. This problem has been tackled in [1–7] and references therein. Compared to the density problem, a related and more important problem is the complexity: to determine how many neurons are necessary to yield a prescribed degree of approximation. There have been many studies for this problem. We refer the readers to Anastassiou [8], Barron [9], Ferrari and Stengel [10], Maiorov and Meir [11], Makovoz [12], Mhaskar and Micchelli [13] for more information about the complexity problem.

On the other hand, the Jackson-type inequality which describes the relation between the smoothness of the target function and the rate of approximation has been extensively used in approximation theory and neural networks. If the activation

☆ The research was supported by the National 973 Programming (2013CB329404), the Key Program of National Natural Science Foundation of China (Grant No. 11131006).

* Corresponding author. Tel.: +86 18392137967.
E-mail address: sblin1983@gmail.com (S. Lin).

http://dx.doi.org/10.1016/j.apm.2014.05.018
0307-904X/

function σ is analytic and non-polynomial, Mhaskar [14] proved that the Jackson-type inequality held for neural networks formed as (1.1). If σ is a sigmoidal function, i.e.,

$$\lim_{t\to\infty}\sigma(t)=1,\quad \lim_{t\to-\infty}\sigma(t)=0,$$

then Chen [15] established a Jackson-type inequality for neural networks (1.1) with $d=1$. He also proposed an open question that whether the Jackson-type inequality held for neural network (1.1) with $d\geqslant 2$ if σ is sigmoidal. In the recent paper [8], Anastassiou devoted to giving an answer to this question. However, he only proved that if the activation function is a combination of sigmoidal function, then the Jackson-type inequality held. In this paper, we will give an answer to the above question in another direction. We will prove that if we give a little change for the structure of neural networks by introducing a distance based on a partition of the cube, then the multivariate Jackson-type inequality holds for the new type of neural network with sigmoidal activation function.

This paper is organized as follows. In the next section, we will introduce a partition-based distance on and give the construction of the neural networks. Our main result will be given in the third section, where a Jackson-type error estimates for approximation by neural network will be given. In Section 4, we will verify our statement by two simulation experiments. In the last section, we will draw a conclusion of this paper.

2. Construction of neural networks

In this section, we introduce a new type of neural networks. Let $\mathbf{I}^d:=[0,1]^d$. In [15], Chen proved that the neural networks formed as

$$N_n^*(x)=c_0+\sum_{i=1}^{n-1}c_i\sigma(a_ix+b_i), \tag{2.1}$$

possessed prominent approximation capability when $d=1$ and σ is a bounded sigmoidal function. An obvious extension for this type of neural networks to multivariate case is

$$N_n^{(1)}(\mathbf{x}):=c_0+\sum_{i=1}^{n-1}c_i\sigma(\mathbf{a}_i\cdot\mathbf{x}+b_i), \tag{2.2}$$

where $\mathbf{x}\cdot\mathbf{y}$ denotes the inner product between the vectors $\mathbf{x}$ and $\mathbf{y}$. However, to the best of our knowledge, it is not very easy to establish a Jackson-type inequality for such type of network if $d\geqslant 2$ and σ is sigmoidal. The main reason is that there is not a strict order for any points in $\mathbf{I}^d$ when $d\geqslant 2$, but there are strict orders for approximation by neural networks [16]. So, we turn to another type of extension of (2.1). For $u=0$, we have

$$a_ix+b_i=a_i(x-v_i)=a_i((x-u)-(v_i-u))=a_i(d(x,u)-d(v_i,u)),$$

where $d(x,y)$ denotes the Euclidean distance between x and y. Thus, we can rewrite (2.1) as

$$N_n^*(x)=c_0+\sum_{i=1}^{n-1}c_i\sigma(a_i(d(x,u)-d(v_i,u))).$$

Noting this, we define

$$N_n(\mathbf{x}):=c_0+\sum_{i=1}^{n-1}c_i\sigma(a_i(\overline{d}(\mathbf{x},\mathbf{u})-\overline{d}(\mathbf{v}_i,\mathbf{u}))), \tag{2.3}$$

where $\overline{d}(\mathbf{x},\mathbf{y})$ is a partition-based distance between $\mathbf{x}$ and $\mathbf{y}$ which is introduced to guarantee the order for points in $\mathbf{I}^d$. Thus, as far as the strict order is concerned, the special neural networks (2.3) is a more suitable extension of the univariate counterpart (2.1). Based on this property, we can deduce a multivariate Jackson inequality for neural networks approximation, which can be regarded as an extension of Chen's [15] result. The proposed network $N_n(\cdot)$ can be interpreted as a model of feed-forward neural networks with four layers:

- The first one is the input layer with the input $\mathbf{x}$ ($\mathbf{x}\in\mathbf{I}^d$).
- The second one is the pre-handling layer, which transform an input $\mathbf{x}$ into the partition-based distance between $\mathbf{u}$ and $\mathbf{x}$, $\overline{d}(\mathbf{x},\mathbf{u})$.
- The third one is the handling layer with n neurons in it.
- The last one is the output layer.

Before giving a concrete definition for $\overline{d}(\cdot,\cdot)$, we need give a division of $\mathbf{I}^d$. For the sake of brevity, we only study it for $d=2$. Divide $\mathbf{I}^2$ into n^2 small square with length $\frac{1}{n}$, J_k, $k=1,\ldots,n^2$. Let $\mathbf{x}_1,\ldots,\mathbf{x}_{n^2}$ be the centers of $J_1,\ldots,J_{n^2}$, and $d(\mathbf{x}_k,\mathbf{x}_{k+1})=\frac{1}{n}$. For example, if n is odd, then we have (see Fig. 1).

Please cite this article in press as: S. Lin et al., Multivariate Jackson-type inequality for a new type neural network approximation, Appl. Math. Modell. (2014), http://dx.doi.org/10.1016/j.apm.2014.05.018

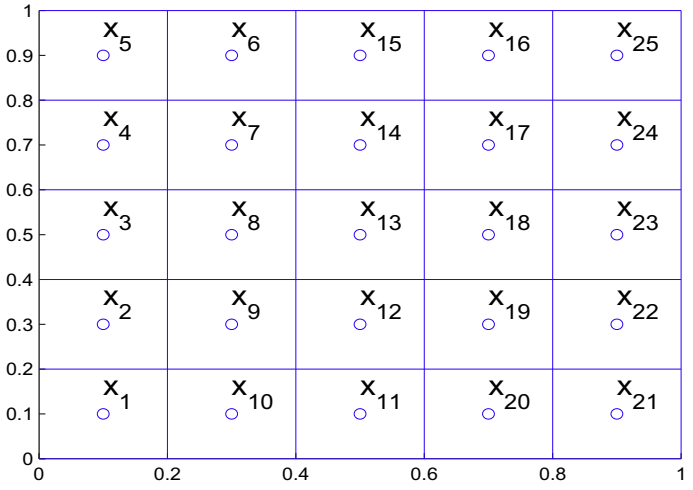

Fig. 1. rearrangement of points with $n = 5$.

$$\mathbf{x}_1 = \left(\frac{1}{2n}, \frac{1}{2n}\right),\ \mathbf{x}_2 = \left(\frac{1}{2n}, \frac{1}{2n} + \frac{1}{n}\right), \ldots,\ \mathbf{x}_n = \left(\frac{1}{2n}, \frac{1}{2n} + \frac{n-1}{n}\right),$$
$$\mathbf{x}_{n+1} = \left(\frac{3}{2n}, \frac{1}{2n} + \frac{n-1}{n}\right),\ \mathbf{x}_{n+2} = \left(\frac{3}{2n}, \frac{1}{2n} + \frac{n-2}{n}\right), \ldots,\ \mathbf{x}_{2n} = \left(\frac{3}{2n}, \frac{1}{2n}\right),$$
$$\mathbf{x}_{2n+1} = \left(\frac{5}{2n}, \frac{1}{2n}\right),\ \mathbf{x}_{2n+2} = \left(\frac{5}{2n}, \frac{3}{2n}\right), \ldots,\ \mathbf{x}_{3n} = \left(\frac{5}{2n}, \frac{1}{2n} + \frac{n-1}{n}\right),$$
$$\ldots,$$
$$\mathbf{x}_{(n-1)n+1} = \left(\frac{2n-1}{2n}, \frac{1}{2n}\right),\ \mathbf{x}_{(n-1)n+2} = \left(\frac{2n-1}{2n}, \frac{3}{2n}\right), \ldots,\ \mathbf{x}_{n^2} = \left(\frac{2n-1}{2n}, \frac{2n-1}{2n}\right).$$

Then, for arbitrary $\mathbf{x} \in \mathbf{I}^2$, there exists a unique k_0 such that $\mathbf{x} \in J_{k_0}$ (sometimes, there are two or more (at most four) indices $k_1, \ldots, k_4$ such that $\mathbf{x} \in J_{k_i}, i = 1, \ldots, 4$. We only choose $k_0 = \min\{k_1, \ldots, k_4\}$.) Now, we are in a position to give the definition of the partition-based distance between $\mathbf{x}_1$ and $\mathbf{x}$ (see Fig. 2),

$$\overline{d}(\mathbf{x}_1, \mathbf{x}) := \sum_{k=1}^{k_0-1} d(\mathbf{x}_{k+1}, \mathbf{x}_k) + d(\mathbf{x}_{k_0}, \mathbf{x}). \tag{2.4}$$

From the above definition, we know that for arbitrary $\mathbf{x} \in \mathbf{I}^2$, $\overline{d}(\mathbf{x}_1, \mathbf{x})$ is a function with $\mathbf{x}$ being its variable. The following Fig. 3 illustrates the relation been $\mathbf{x}$ and $d(\mathbf{x}_1, \mathbf{x})$ with $n = 50$.

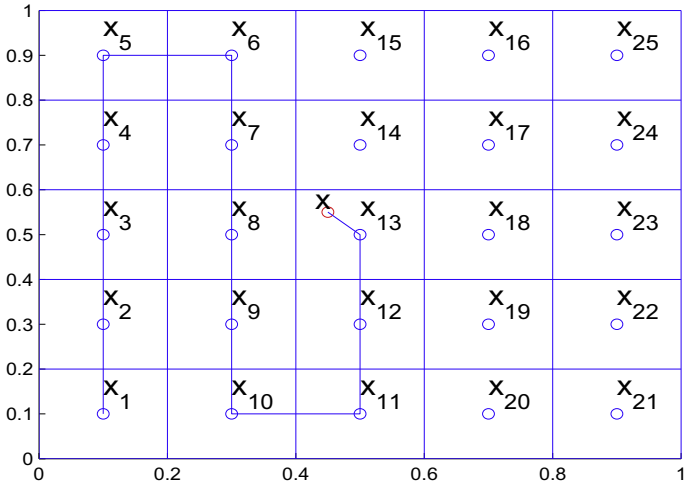

Fig. 2. Partition-based distance between $\mathbf{x}$ and $\mathbf{x}_1$ with $n = 5$.

Please cite this article in press as: S. Lin et al., Multivariate Jackson-type inequality for a new type neural network approximation, Appl. Math. Modell. (2014), http://dx.doi.org/10.1016/j.apm.2014.05.018

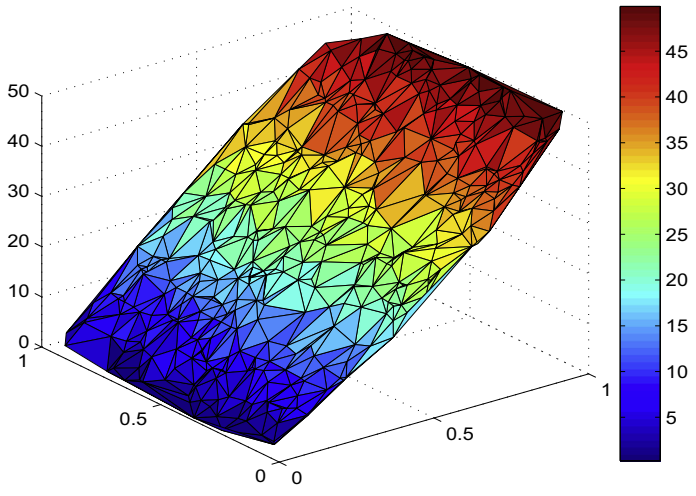

Fig. 3. Distance $d(x_1, x)$ with $n = 50$.

For $d \geqslant 3$, we can divide the cube $\mathbf{I}^d$ into n^d small cube $J_k^d,\ k = 1, \ldots, n^d$. Then, we use the same method as above to choose n^d points $\mathbf{x}_1,\ \mathbf{x}_2, \ldots,\ \mathbf{x}_{n^d}$ satisfying $\mathbf{x}_1 = (\frac{1}{2n}, \frac{1}{2n}, \ldots, \frac{1}{2n})$, and $d(x_{i-1}, x_i) = \frac{1}{n},\ i = 2, \ldots, n^d$. Therefore, for every $\mathbf{x} \in \mathbf{I}^d$, there exists a unique $k_0 \leqslant n^d$ such that $\mathbf{x} \in J_{k_0}^d$. Thus, we define a general distance between $\mathbf{x}_1$ and $\mathbf{x}$ and a general distance between $\mathbf{x}_1$ and $\mathbf{x}_j$ as $\overline{d}(\mathbf{x}_1, \mathbf{x}) := \sum_{k=1}^{k_0-1} d(\mathbf{x}_{k+1}, \mathbf{x}_k) + d(\mathbf{x}_{k_0}, \mathbf{x})$ and $\overline{d}(\mathbf{x}_1, \mathbf{x}_j) := \sum_{k=1}^{j-1} (d(\mathbf{x}_{k+1}, \mathbf{x}_k))$, respectively.

3. Main results

In this section, we give the main result of this paper, where a Jackson-type error estimate for approximation by neural networks with bounded sigmoidal function will be established.

At first, we need introduce a modulus of smoothness in $\mathbf{I}^d$ [17] as

$$\omega(f, t) := \sup_{0 \leqslant d(\mathbf{x}, \mathbf{y}) \leqslant t, \mathbf{x}, \mathbf{y} \in \mathbf{I}^d} |f(\mathbf{x}) - f(\mathbf{y})|.$$

The modulus of smoothness is usually considered as the measure of the smoothness of function and the approximation error in approximation theory. The function f is called Lipschitz α $(0 < \alpha \leqslant 1)$ continuous and is written as $f \in Lip_{C_l}(\alpha)$, if there exists a constant C_l such that $\omega(f, t) \leqslant C_l t^{\alpha}$. Furthermore, if we denote by Φ_n^{σ} the set of functions formed as (2.3), then for a given function $f \in C(\mathbf{I}^d)$, we define the best approximation error of Φ_n^{σ} by

$$E_n(f) := \inf_{g \in \Phi_n^{\sigma}} \|f - g\|,$$

where $\|f - g\| := \sup_{\mathbf{x} \in \mathbf{I}^d} |f(\mathbf{x}) - g(\mathbf{x})|$. By the help of the modulus of smoothness and the best approximation error, we obtain the following Jackson-type inequality for FNN approximation.

Theorem 1. *Let σ be a bounded sigmoidal function. If $f \in C(\mathbf{I}^d)$, then there exists a constant C depending only on d and σ such that*

$$E_n(f) \leqslant C\omega\left(f, n^{-\frac{1}{d}}\right). \tag{3.1}$$

Proof. Let $n \in \mathbf{N}$. Define

$$B := \frac{1}{n^d}.$$

It follows from the definition of the sigmoidal function that there exists an $A > 0$ such that

$$|\sigma(t) - 1| < B \quad \text{if} \quad t \geqslant A \quad \text{and} \quad |\sigma(t)| < B \quad \text{if} \quad t \leqslant -A.$$

If we define $\sigma^*(t) := \sigma(Ant)$, then

$$|\sigma^*(t) - 1| < B \quad \text{if} \quad t \geqslant \frac{1}{n} \quad \text{and} \quad |\sigma^*(t)| < B \quad \text{if} \quad t \leqslant -\frac{1}{n}. \tag{3.2}$$

Please cite this article in press as: S. Lin et al., Multivariate Jackson-type inequality for a new type neural network approximation, Appl. Math. Modell. (2014), http://dx.doi.org/10.1016/j.apm.2014.05.018

Without loss of generality, we assume $k_0 \geqslant 2$, then by the definition of $\overline{d}(\mathbf{x}_1, \mathbf{x})$, we have

$$\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k) \geqslant \frac{1}{n}, \quad 1 \leqslant k \leqslant k_0 - 1,$$

$$\left|\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_{k_0})\right| \leqslant \frac{\sqrt{2}}{2n},$$

$$\left|\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_{k_0+1})\right| \leqslant \frac{1}{n},$$

and

$$\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k) \leqslant -\frac{1}{n}, \quad k \geqslant k_0 + 2.$$

Hence, it follows from (3.2) that

$$\left|\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k)\right) - 1\right| < B, \quad k \leqslant k_0 - 1,$$

and

$$\left|\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k)\right)\right| < B, \quad k \geqslant k_0 + 2.$$

Define

$$N_n^\sigma(\mathbf{x}) := f(\mathbf{x}_1) + \sum_{k=1}^{n^d-1} (f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k))\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k)\right). \tag{3.3}$$

It is obvious that $N_n^\sigma \in \Phi_{n^d}^\sigma$. Then we obtain

$$\begin{aligned} f(\mathbf{x}) - N_n^\sigma(\mathbf{x}) = {} & f(\mathbf{x}) - f(\mathbf{x}_1) - \sum_{k=1}^{n^d-1} f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k)\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k)\right) = f(\mathbf{x}) - f(\mathbf{x}_1) \\ & - \sum_{k=1}^{k_0-1} (f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k))\left(\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k)\right) - 1\right) + f(\mathbf{x}_1) - f(\mathbf{x}_{k_0}) - (f(\mathbf{x}_{k_0+1}) \\ & - f(\mathbf{x}_{k_0}))\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_{k_0})\right) - (f(\mathbf{x}_{k_0+2}) - f(\mathbf{x}_{k_0+1}))\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_{k_0+1})\right) \\ & - \sum_{k=k_0+2}^{n^d-1} (f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k))\sigma^*\left(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k)\right). \end{aligned}$$

Thus, from the definition of modulus of smoothness, it follows that

$$\begin{aligned} |f(\mathbf{x}) - N_n^\sigma(\mathbf{x})| & \leqslant (k_0 - 1)\omega\left(f, \frac{1}{n}\right)B + \omega\left(f, \frac{\sqrt{2}}{2n}\right) + \|\sigma\|\omega\left(f, \frac{1}{n}\right) + \|\sigma\|\omega\left(f, \frac{1}{n}\right) + (n^d - k_0 - 1)\omega\left(f, \frac{1}{n}\right)B \\ & \leqslant \frac{k_0 - 1}{n^d}\omega\left(f, \frac{1}{n}\right) + \omega\left(f, \frac{1}{n}\right) + 2\|\sigma\|\omega\left(f, \frac{1}{n}\right) + \frac{n^d - k_0 - 1}{n^d}\omega\left(f, \frac{1}{n}\right) \leqslant C\omega\left(f, \frac{1}{n}\right), \end{aligned}$$

where $\|\sigma\| := \sup_{t\in\mathbf{R}}|\sigma(t)|$. This implies that

$$E_n(f) \leqslant C\omega\left(f, n^{-\frac{1}{d}}\right).$$

4. Numerical results

In this section, we present two numerical experiments to demonstrate the validity of the obtained results.
For the first one, we select the target function as $f = x_{(1)}\exp\left(-x_{(1)}^2 - x_{(2)}^2\right) + \sin\left(x_{(1)} \cdot x_{(2)}\right)$. We use the neural network

$$N_n^\sigma(\mathbf{x}) = f(\mathbf{x}_1) + \sum_{k=1}^{n^2-1} (f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k))\sigma\left(An(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k))\right),$$

to approximate f. Set $n = 50$, $A = 2$, and σ be the well known Heaviside function, i.e., $\sigma(t) = 1$ for $t \geqslant 0$ and $\sigma(t) = -1$ for $t < 0$. The following Fig.4 shows that the constructed neural network can approximate the target function very well.

For the second one, we select the target function as $f = x_{(1)}\log\left(10\left(x_{(1)}^2 + x_{(2)}^2\right)\right) + \cos(x_{(1)} + x_{(2)}) - x_{(2)}$, and the activation function of the neural network as the well known logistic squasher $\sigma(t) = \frac{1}{1+e^{-t}}$. Similarly, we can construct a neural network

Please cite this article in press as: S. Lin et al., Multivariate Jackson-type inequality for a new type neural network approximation, Appl. Math. Modell. (2014), http://dx.doi.org/10.1016/j.apm.2014.05.018

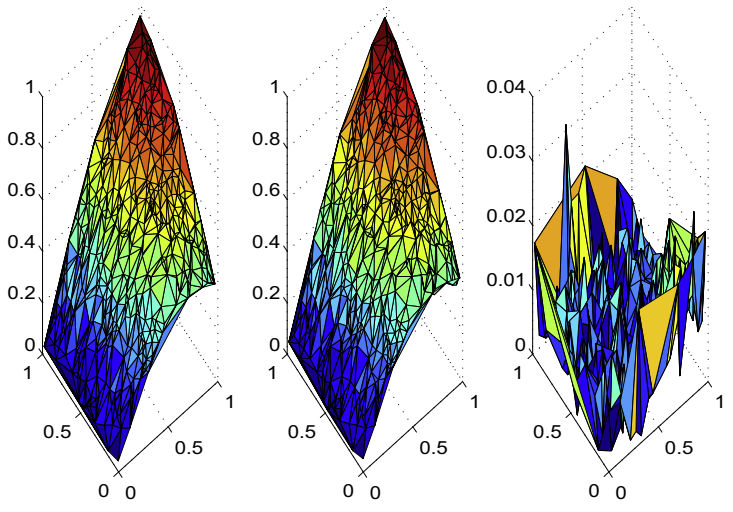

Fig. 4. The left figure is the target function, the middle one is the constructed neural network, and the right one is the error function, i.e., $f - N_n^\sigma$.

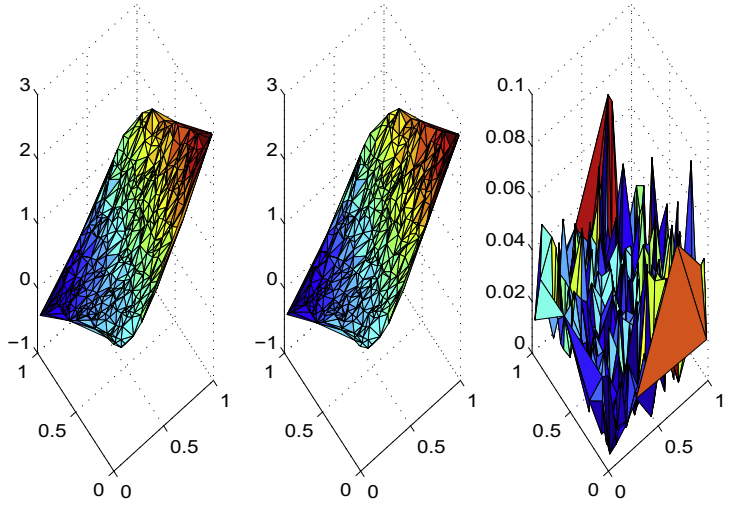

Fig. 5. The left figure is the target function, the middle one is the constructed neural network, and the right one is the error function, i.e., $f - N_n^\sigma$.

$$N_n^\sigma(\mathbf{x}) = f(\mathbf{x}_1) + \sum_{k=1}^{2500-1} (f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k))\sigma\Big(100(\overline{d}(\mathbf{x}_1, \mathbf{x}) - \overline{d}(\mathbf{x}_1, \mathbf{x}_k))\Big),$$

such that $\|f - N_n^\sigma\|$ is very small (see Fig. 5).

5. Conclusion

In this paper, we have proposed a new type of FNN and studied its approximation properties. In general, for the usual FNN formed as (2.2) with sigmoidal activation function, it is difficult to guarantee that Jackson-type inequality holds for FNN approximation when $d \geqslant 2$. To establish such an inequality, we introduce an FNN with new structure in the hidden layer, which maintains that there is an order between different points in $\mathbf{I}^d$. Based on this, we have constructed a feasible FNN and established a Jackson-type error estimate for approximating continuous functions by the constructed FNN. Our methods of proof are constructive. The numerical experiments have verified our theoretical results.

References

[1] T.P. Chen, H. Chen, Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical system, IEEE Trans. Neural Networks 6 (1995) 911–917.

Please cite this article in press as: S. Lin et al., Multivariate Jackson-type inequality for a new type neural network approximation, Appl. Math. Modell. (2014), http://dx.doi.org/10.1016/j.apm.2014.05.018

[2] C.K. Chui, X. Li, Approximation by ridge functions and neural networks with one hidden layer, J. Approx. Theory 70 (1992) 131–141.
[3] G. Cybenko, Approximation by superpositions of sigmoidal function, Math. Control Signals Syst. 2 (1989) 303–314.
[4] K.I. Funahashi, On the approximate realization of continuous mapping by neural networks, Neural Networks 2 (1989) 183–192.
[5] K. Hornik, M. Stinchombe, H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Neural Networks 3 (1990) 551–560.
[6] M. Leshno, V.Y. Lin, A. Pinks, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, Neural Networks 6 (1993) 861–867.
[7] X. Li, On simultaneous approximation of by radial basis function neural networks, Appl. Math. Comput. 95 (1998) 75–89.
[8] G.A. Anastassiou, Multivariate sigmoidal neural network approximation, Neural Networks 24 (2011) 378–386.
[9] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inf. Theory 39 (1993) 930–945.
[10] S. Ferrari, R.F. Stengel, Smooth function approximation using neural networks, IEEE Trans. Neural Networks 16 (2005) 24–38.
[11] V. Maiorov, R.S. Meir, Approximation bounds for smooth function in $C(\mathbf{R}^d)$ by neural and mixture networks, IEEE Trans. Neural Networks 9 (1998) 969–978.
[12] Y. Makovoz, Uniform approximation by neural networks, J. Approx. Theory 95 (1998) 215–228.
[13] H.N. Mhaskar, C.A. Micchelli, Degree of approximation by neural networks with a single hidden layer, Adv. Appl. Math. 16 (1995) 151–183.
[14] H.N. Mhaskar, Neural networks for optimal approximation of smooth and analytic functions, Neural Comput. 8 (1996) 164–177.
[15] D.B. Chen, Degree of approximation by superpositions of a sigmoidal function, Approx. Theory Appl. 9 (1993) 17–28.
[16] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numer. 8 (1999) 143–195.
[17] R.A. DeVore, G.G. Lorentz, Constructive Approximation, Springer-Verlag, Berlin, 1993.

Please cite this article in press as: S. Lin et al., Multivariate Jackson-type inequality for a new type neural network approximation, Appl. Math. Modell. (2014), http://dx.doi.org/10.1016/j.apm.2014.05.018