Applied Mathematics and Computation 224 (2013) 29–35

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Approximation by neural networks with scattered data ☆

Shaobo Lin [a], Xiaofei Guo [b], Feilong Cao [c], Zongben Xu [a,*]

[a] Institute for Information and System Sciences, School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, Shanxi Province, PR China
[b] Department of Applied Mathematics, Northwestern Polytechnical University, Xian 710129, Shanxi Province, PR China
[c] Institute of Metrology and Computational Science, China Jiliang University, Hangzhou 310018, Zhejiang Province, PR China

ARTICLE INFO

Keywords:
Neural networks
Scattered data
Reproducing kernel
Trigonometric polynomials

ABSTRACT

In this paper, the approximation capability of feed-forward neural networks (FNNs) formed as $\sum_{k=1}^{n} c_k\phi(x - x_k)$ is considered, where ϕ is an even periodic continuous function and x_k's are scattered data. The best approximation error of trigonometric polynomials and the continuous modulus of the activation function ϕ are introduced to describe the approximation capability of FNNs.

1. Introduction

In the past two decades, we have witnessed the enormous emergence concerning the approximation capabilities of feed-forward neural networks (FNNs). FNNs can be formally described as devices producing input–output functions depending on flexible parameters. Generally speaking, the input–output functions possess the form of linear combination 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.

Theoretically, any continuous functions can be approximated by FNNs to arbitrary desired degree provided the number of hidden neurons is large enough. This result is usually called the density problem of FNN approximation. For example, by using standard functional analysis approach, Cybenko [9] established the density theorem of FNNs with the well known sigmoidal activation function. Similar results can also be found in Funahashi [12], Hornik et al. [13], Li [15], Leshno et al. [14], Chen and Chen [6], Chui and Li [7]. and references therein.

Compared with the density problem, a related and more important problem is that of complexity: to determine how many neurons is necessary to yield a prescribed degree of FNN approximation, i.e., one tries to describe the relationship between the rate of approximation and the number of neurons in the hidden layer. Up till now, many authors have published results on the complexity issue of FNNs with various activation functions. For example, by taking the bounded sigmoidal function as the activation function, Chen [5] established a Jackson-type error estimate for FNN approximation on the interval $[0, 1]$. Recently, Cao and Zhang [4] extended it to L^p space, i.e., they give an L^p error estimate for such an approximation. By taking monotone odd function as the activation function, Cao et al. [3] constructed an FNN to approximate the continuous function and established a Jackson-type inequality on the unit interval. For more details about the complexity issue, we refer

☆ The research was supported by the National 973 Programming (2013CB329404), the Key Program of National Natural Science Foundation of China (Grant No. 11131006), and the National Natural Science Foundations of China (Grants No. 61075054).

* Corresponding author.
E-mail address: zbxu@mail.xjtu.edu.cn (Z. Xu).

0096-3003/$ - see front matter
http://dx.doi.org/10.1016/j.amc.2013.08.014

the readers to Barron [2], Mhaskar and Micchelli [19], Maiorov and Meir [17], Makovoz [18], Ferrari and Stengel [11], Xu and Cao [23], etc.

On the other hand, the problem of effectively representing an underlying function based on its values sampled at finitely many distinct scattered sites $X = \{x_i\}_{i=1}^n$ is important and arises in many applications such as machine learning [8], scattered data fitting [22] and differential or integral equations [21]. For example, due to the so-called representation theorem of learning theory [20], the solutions to both the regularized least square algorithm and support vector machine algorithm take the form as

$$\sum_{k=1}^{n} \phi(x - x_k), \tag{1}$$

where x_k's are the sites at which the data are collected. Therefore, it seems natural to consider networks that evaluate a function of the form as (1). In the present paper, by taking the even and periodic function as the activation function, we can construct an FNN formed as (1) to approximate any continuous functions. To this end, we first give an integral representation for any trigonometric polynomials. Then, we construct an FNN to approximate the integral representation. At last, the approximation bound for trigonometric polynomials implies a Jackson-type error estimate for the constructed neural network. The modulus of smoothness of the activation function and the best trigonometric polynomial approximation error are introduced to describe the approximation capability of FNNs.

2. Approximation trigonometric polynomials by FNNs

In this section, we construct an FNN formed as (1) to approximate trigonometric polynomials. To this end, some definitions concerning reproducing kernel Hilbert space (RKHS) are required.

Let $\phi(x, y) : [-\pi, \pi] \times [-\pi, \pi] \to \mathbb{R}$ be continuous, symmetric and positive definite, i.e., for any finite set of distinct points $X = \{x_i\}_{i=1}^l \subset [-\pi, \pi]$, the matrix $\{\phi(x_i, y_j)\}_{1 \leqslant i,j \leqslant l}$ is positive definite. Then the RKHS, $\mathcal{H}_\phi$, associated with the kernel ϕ is defined (see [1]) to be the closure of the linear span of the set of functions $\{\phi_x := \phi(x, \cdot) : x \in X\}$ with the inner product $\langle \cdot, \cdot \rangle_\phi$ and norm $\| \cdot \|_\phi$ satisfying $\langle \phi_x, \phi_y \rangle_\phi = \phi(x, y)$,

$$\langle \phi_x, g \rangle_\phi = g(x), \quad \forall x \in X, \; g \in \mathcal{H}_\phi,$$

and $\|g\|_\phi := \langle g, g \rangle_\phi$, respectively. According to the definition of $\|g\|_\phi$, we can obtain

$$\|g\| \leqslant \kappa \|g\|_\phi, \quad \forall g \in \mathcal{H}_\phi, \tag{2}$$

where $\| \cdot \|$ denotes the uniform norm and $\kappa = \sup_{x \in [-\pi,\pi]} \sqrt{\phi(x, x)}$.

The next thing we do is to give an integral representation for arbitrary trigonometric polynomials. Let $\mathcal{T}_n$ be the collection of trigonometric polynomial defined on $[-\pi, \pi]$ with degree not larger than n. The kernel

$$K_n(x, y) := \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{n} \cos k(x - y), \quad x, y \in [-\pi, \pi],$$

will play an important role in our proof. The following Lemma 1 which can be found in [10] shows the reproducing property of $K_n(\cdot, \cdot)$.

Lemma 1. *The kernel $K_n(\cdot, \cdot)$ is continuous, symmetric and positive definite. Furthermore, it is the reproducing kernel of $\mathcal{T}_n$, i.e.,*

(i) $K_n(x, \cdot) \in \mathcal{T}_n$ with fixed x.
(ii) $\int_{-\pi}^{\pi} T_n(y) K_n(x - y) dy = T_n(x)$.

For further use, we also introduce the following Lemma 2.

Lemma 2. *If ϕ is an even and periodic function with 2π being its period, then for arbitrary $a, b \in \mathbb{R}$, there holds*

$$\int_{-\pi}^{\pi} \phi(x - y)(a \cos ky + b \sin ky) dy = \hat{\phi}_k (a \cos kx + b \sin kx), \quad k = 0, 1, \ldots,$$

where

$$\hat{\phi}_k := \int_{-\pi}^{\pi} \phi(t) \cos kt dt. \tag{3}$$

Proof. Since ϕ is even and 2π-periodical function, we have

$$\begin{aligned}
&\int_{-\pi}^{\pi}\phi(x-y)(a\cos ky+b\sin ky)dy \\
&= a\int_{-\pi}^{\pi}\phi(y-x)\cos kydy + b\int_{-\pi}^{\pi}\phi(y-x)\sin kydy \\
&= a\int_{-\pi}^{\pi}\phi(y)\cos k(y+x)dy + b\int_{-\pi}^{\pi}\phi(y)\sin k(y+x)dy \\
&= a\int_{-\pi}^{\pi}\phi(y)\cos kydy\cos kx - a\int_{-\pi}^{\pi}\phi(y)\sin kydy\sin kx \\
&\quad + b\int_{-\pi}^{\pi}\phi(y)\cos kydy\sin kx + b\int_{-\pi}^{\pi}\phi(y)\sin kydy\cos kx \\
&= \hat{\phi}_k(a\cos kx + b\sin kx).
\end{aligned}$$

This completes the proof of Lemma 2. □

Based on the above two lemmas, we give the integral representation for any $T_n \in \mathcal{T}_n$.

Lemma 3. *If ϕ is continuous, even, 2π-periodic, and satisfies $\hat{\phi}_k \neq 0, k = 0, 1, \ldots$, then for arbitrary $T_n \in \mathcal{T}_n$, we have*

$$T_n(x) = \frac{1}{\pi}\sum_{k=1}^{n}\frac{1}{\hat{\phi}_k}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}T_n(y)\cos k(y-z)\phi(x-z)dydz + \frac{1}{2\pi\hat{\phi}_0}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}T_n(y)\phi(x-z)dydz.$$

Proof. From Lemma 1, we know that for arbitrary $T_n \in \mathcal{T}_n$, there holds that

$$T_n(x) = \int_{-\pi}^{\pi}T_n(y)K_n(x-y)dy = \frac{1}{\pi}\sum_{k=1}^{n}\int_{-\pi}^{\pi}T_n(y)\cos k(x-y)dy + \frac{1}{2\pi}\int_{-\pi}^{\pi}T_n(y)dy.$$

It follows from Lemma 2 that for arbitrary even function $\phi \in C([-\pi,\pi])$, there holds

$$\begin{aligned}
\int_{-\pi}^{\pi}\phi(x-z)\cos k(y-z)dz &= \left(\cos ky\int_{-\pi}^{\pi}\phi(x-z)\cos kzdz + \sin ky\int_{-\pi}^{\pi}\phi(x-z)\sin kzdz\right) \\
&= \hat{\phi}_k(\cos ky\cos kx + \sin ky\sin kx) = \hat{\phi}_k\cos k(x-y).
\end{aligned}$$

Since ϕ satisfying $\hat{\phi}_k \neq 0$, then for any $k = 0, 1, \ldots$, there holds

$$\cos k(x-y) = \frac{1}{\hat{\phi}_k}\int_{-\pi}^{\pi}\phi(x-z)\cos k(y-z)dz.$$

Thus, we have

$$T_n(x) = \frac{1}{\pi}\sum_{k=1}^{n}\frac{1}{\hat{\phi}_k}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}T_n(y)\cos k(y-z)\phi(x-z)dydz + \frac{1}{2\pi\hat{\phi}_0}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}T_n(y)\phi(x-z)dydz.$$

This finishes the proof of Lemma 3. □

Let $\sigma(\cdot)$ be the Heaviside function, i.e.,

$$\sigma(t) := \begin{cases} 1, & t \geqslant 0, \\ 0, & t < 0. \end{cases}$$

Denote

$$c_1(x) := 1 - \sigma(x - x_1),$$

$$c_i(x) := \sigma(x - x_{i-1}) - \sigma(x - x_i), \quad 2 \leqslant i \leqslant n-1,$$

$$c_n(x) := \sigma(x - x_{n-1}).$$

Now we define the neural network as

$$N_\phi^n(x) := \sum_{i=1}^{n}\left(\frac{1}{\pi}\sum_{k=1}^{n}\frac{1}{\hat{\phi}_k}\int_{-\pi}^{\pi}c_i(z)\int_{-\pi}^{\pi}T_n(y)\cos k(y-z)dydz + \frac{1}{2\pi\hat{\phi}_0}\int_{-\pi}^{\pi}c_i(z)\int_{-\pi}^{\pi}T_n(y)dydz\right)\phi(x-x_i).$$

To characterize the approximation degree of the constructed FNN, we need introduce the modulus of smoothness of a continuous function f as

$$\omega(f,t) = \sup_{0\leqslant h\leqslant t} \max_{x,x+h\in[-\pi,\pi]} |f(x+h)-f(x)|.$$

The smoothness modulus is often used as a tool for measuring approximation error. It is also used to measure the smoothness of a function and its accuracy between approximation theory and Fourier analysis (see [16]). The function f is called (M,α)-Lipschitz continuous $(0<\alpha\leq 1)$, which can be written as $f\in \mathrm{Lip}(M,\alpha)$, if and only if there exists a constant $M>0$ such that

$$\omega(f,\delta)\leqslant M\delta^{\alpha}.$$

Now we are in a position to give the main result of this section.

Theorem 1. *Let ϕ be continuous, even, 2π-periodic and satisfy $\hat{\phi}_k>0,\ k=0,1,\ldots$, then for arbitrary $T_n\in\mathcal{T}_n$ there holds*

$$\|T_n - N_\phi^n\|_\phi^2 \leqslant 2\frac{\|T_n\|_1^2}{\hat{\phi}_0^2}\omega(\phi,\Delta),$$

where $\|g\|_1$ is the usual L^1 norm of the integrable function g, $\hat{\phi}_0 := \int_{-\pi}^{\pi}\phi(x)dx$ and $\Delta := \max_{1\leqslant i\leqslant n}|x_{i+1}-x_i|$ denotes the maximum separation distance of X.

Proof. Since $\hat{\phi}_k>0$ for all $k\geqslant 0$, it follows from [8, Proposition 2.12] that the kernel defined by $\phi(\cdot,\cdot):=\phi(\cdot-\cdot)$ is positive definite. Denote

$$h(z) := \frac{1}{\pi}\sum_{k=1}^{n}\frac{1}{\hat{\phi}_k}\int_{-\pi}^{\pi}T_n(y)\cos k(y-z)dy + \frac{1}{2\pi\hat{\phi}_0}\int_{-\pi}^{\pi}T_n(y)dy.$$

Since

$$\langle T_n, T_n\rangle_\phi = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}h(z)\phi(z-y)h(y)dzdy,$$

and

$$\langle T_n, \phi(\cdot - x_i)\rangle_\phi = T_n(x_i),$$

it follows that

$$\begin{aligned}
\|T_n - N_\phi^n\|_\phi^2 &= \langle T_n - N_\phi^n, T_n - N_\phi^n\rangle_\phi = \left\langle T_n - \sum_{i=1}^{n}\phi(\cdot - x_i)\int_{-\pi}^{\pi}c_i(z)h(z)dz, T_n - \sum_{j=1}^{n}\phi(\cdot - x_j)\int_{-\pi}^{\pi}c_j(y)h(y)dy\right\rangle_\phi \\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}h(z)h(y)\left\langle \phi(\cdot - z) - \sum_{i=1}^{n}c_i(z)\phi(\cdot - x_i), \phi(\cdot - y) - \sum_{j=1}^{n}c_j(y)\phi(\cdot - x_j)\right\rangle_\phi dzdy \\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Bigg(\langle\phi(z-\cdot),\phi(\cdot - y)\rangle_\phi - \sum_{j=1}^{n}c_j(y)\langle\phi(\cdot - z),\phi(x_j-\cdot)\rangle_\phi \\
&\quad -\sum_{i=1}^{n}c_i(z)\left(\langle\phi(\cdot - x_i),\phi(\cdot - y)\rangle_\phi - \sum_{j=1}^{n}c_j(y)\langle\phi(\cdot - x_i),\phi(\cdot - x_j)\rangle_\phi\right)\Bigg)h(z)h(y)dzdy \\
&= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\phi(y-z) - \sum_{j=1}^{n}c_j(y)\phi(z-x_j) - \sum_{i=1}^{n}c_i(z)\left(\phi(y-x_i) - \sum_{j=1}^{n}c_j(y)\phi(x_i-x_j)\right)\right)h(z)h(y)dzdy.
\end{aligned}$$

Set

$$g(z) := \phi(y-z) - \sum_{j=1}^{n}c_j(y)\phi(z-x_j).$$

Since for arbitrary $x\in[-\pi,\pi]$, there exists a $j_0\in\mathbb{N}$ such that $x\in[x_{j_0},x_{j_0+1}]$ (without loss of generality, we assume $2\leqslant j_0\leqslant n-1$), it follows from the definition of σ that

$$
\begin{aligned}
g(x) &- \sum_{i=1}^{n} g(x_i)c_i(x) \\
&= g(x) - \left(g(x_1) + \sum_{i=1}^{n-1} (g(x_{i+1}) - g(x_i))\sigma(x - x_i) \right) \\
&= g(x) - g(x_1) - \sum_{i=1}^{n-1} (g(x_{i+1}) - g(x_i))\sigma(x - x_i) \\
&= g(x) - g_(x_1) - \sum_{i=1}^{j_0-1} (g(x_{i+1}) - g(x_i))(\sigma(x - x_i) - 1) \\
&\quad + g(x_1) - g(x_{j_0}) - (g(x_{j_0+1}) - g(x_{j_0}))\sigma(x - x_{j_0}) \\
&\quad - \sum_{i=j_0+1}^{n-1} (g(x_{i+1}) - g(x_i))\sigma(x - x_i)) \\
&= g(x) - g(x_{j_0}) - (g(x_{j_0+1}) - g(x_{j_0}))\sigma(x - x_{j_0})
\end{aligned}
$$

Therefore

$$
\left| g(x) - \sum_{i=1}^{n} g(x_i)c_i(x) \right| \leqslant \omega(g, \Delta).
$$

From the definition of $\omega(g, t)$, we obtain

$$
\begin{aligned}
\omega(g, \Delta) &\leqslant \sup_{-\pi \leqslant z, z+\Delta \leqslant \pi} |g(z + \Delta) - g(z)| = \sup_{-\pi \leqslant t, t+\Delta \leqslant \pi} \left| \phi(z + \Delta - y) - \sum_{i=1}^{n} c_i(y)\phi(z + \Delta - x_i) - \phi(z - y) + \sum_{i=1}^{n} c_i(y)\phi(z - x_i) \right| \\
&\leqslant \sup_{-\pi \leqslant z, z+\Delta \leqslant \pi} |\phi(z - y) - \phi(z + \Delta - y)| + \sup_{-\pi \leqslant z, z+\Delta \leqslant \pi} \left| \sum_{i=1}^{n} c_i(y)(\phi(z + \Delta - x_i) - \phi(z - x_i) \right| \\
&\leqslant \omega(\phi, \Delta) + \omega(\phi, \Delta) \sum_{i=1}^{n} c_i(y),
\end{aligned}
$$

where in the last inequality we use the fact that $c_i(y) \geqslant 0$ for arbitrary $i = 1, \ldots, n$. Furthermore, from the definition of $c_i(\cdot)$, there holds that

$$
\sum_{i=1}^{n} c_i(y) = 1.
$$

Thus we obtain

$$
\left| g(x) - \sum_{i=1}^{n} g(x_i)c_i(x) \right| \leqslant 2\omega(\phi, \Delta).
$$

Therefore, we have

$$
\|T_n - N_\phi^n\|_\phi^2 \leq 2 \left(\int_{-\pi}^{\pi} h(z)dz \right)^2 \omega(\phi, \Delta).
$$

Finally we only need to give an upper bound estimate for $\left| \int_{-\pi}^{\pi} h(z)dz \right|$. From the definition of $h(z)$, we deduce

$$
\begin{aligned}
\left| \int_{-\pi}^{\pi} h(z)dz \right| &= \left| \int_{-\pi}^{\pi} \left(\frac{1}{\pi} \sum_{k=1}^{n} \frac{1}{\hat{\phi}_k} \int_{-\pi}^{\pi} T_n(y) \cos k(y - z)dy + \frac{1}{2\pi\hat{\phi}_0} \int_{-\pi}^{\pi} T_n(y)dy \right) dz \right| \\
&= \left| \frac{1}{\pi} \sum_{k=1}^{n} \frac{1}{\hat{\phi}_k} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \cos k(z - y)dzT_n(y)dy + \frac{1}{\hat{\phi}_0} \int_{-\pi}^{\pi} T_n(y)dy \right| = \left| \frac{1}{\hat{\phi}_0} \int_{-\pi}^{\pi} T_n(y)dy \right|
\end{aligned}
$$

Thus,

$$
\left\| T_n - N_\phi^n \right\|_\phi^2 \leq 2 \frac{\|T_n\|_1^2}{\hat{\phi}_0^2} \omega(\phi, \Delta).
$$

This completes the proof of Theorem 1. $\square$

3. Approximation and interpolation by FNNs

In this section, we deduce an upper bound of approximation by FNNs formed as (1) with the help of Theorem 1.

Theorem 2. *Under the assumptions of Theorem 1, for $f \in \mathcal{H}_\phi$, there exists a constant C independent of n such that*

$$\|f - N_\phi^n\|_\phi \leqslant E_n(f)_\phi + C(\omega(\phi,\Delta))^{1/2},$$

where $E_n(f)_\phi := \inf_{T_n \in \mathcal{T}_n} \|f - T_n\|_\phi$ is the best approximation error of the trigonometric polynomials under the RKHS norm.

Proof. For arbitrary $\varepsilon > 0$, it is obvious that there exists a $T_n \in \mathcal{T}_n$ such that

$$\|f - T_n\|_\phi < E_n(f)_\phi + \varepsilon.$$

Therefore, it follows from Theorem 1 that

$$\|f - N_\phi^n\|_\phi \leqslant \|f - T_n\|_\phi + \|T_n - N_\phi^n\| \leqslant E_n(f)_\phi + \varepsilon + C(\omega(\phi,\Delta))^{1/2}.$$

Thus the arbitrariness of ε finishes the proof of Theorem 2. □

Remark 1. Since $\phi(\cdot,\cdot)$ is the reproducing kernel of the RKHS, $\mathcal{H}_\phi$, the best approximation of the FNN formed as (1) in the RKHS norm is also the exact interpolation FNN. Thus Theorem 2 also gives an RKHS norm error estimate of the FNN interpolant.

It is obvious that the smoother ϕ is, the smaller $\mathcal{H}_\phi$ is. It is natural to arise the question: can we get a similar error estimate for $f \in C([-\pi,\pi])$? The following Theorem 3 focuses on giving an affirmative answer to this question.

Theorem 3. *If the assumptions of Theorem 1 holds, then for arbitrary $f \in C([-\pi,\pi])$, there exists a constant C independent of n such that*

$$\left\|f - N_\phi^n\right\| \leqslant C\left(E_{n/2}(f) + \frac{\|f\|_1}{\hat{\phi}_0}(\omega(\phi,\Delta))^{1/2}\right). \tag{4}$$

Proof. From (2) and Theorem 1 it follows that

$$\left\|T_n - N_\phi^n\right\| \leqslant \phi(0)\left\|T_n - N_\phi^n\right\|_\phi \leqslant C\frac{\|T_n\|_1}{\hat{\phi}_0}(\omega(\phi,\Delta))^{1/2},$$

where $E_n(f) := \inf_{T_n \in \mathcal{T}_n}\|f - T_n\|$ is the best approximation error of trigonometric polynomials under the uniform norm. Let $V_{[n/2]}f$ be the well known de la Vallée Poussin operator, i.e.

$$V_{[n/2]}(f)(x) := \frac{1}{[n/2]}\sum_{k=[n/2]}^{2[n/2]-1} S_k(f)(x),$$

where

$$S_k(f)(x) := \frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)\left(\frac{1}{2} + \sum_{j=1}^{k}\cos jt\right)dt.$$

Then we have

$$\|f - V_{[n/2]}f\| \leqslant CE_{[n/2]}(f), \tag{5}$$

and

$$\|V_{[n/2]}f\| \leqslant C\|f\|. \tag{6}$$

Therefore, by setting $T_n := V_{[n/2]}f$ in Theorem 1, we get

$$\left\|f - N_\phi^n\right\| \leqslant \|f - V_{[n/2]}f\| + \|V_{[n/2]}f - N_\phi^n\| \leqslant CE_{[n/2]}(f) + \frac{C\|V_{[n/2]}f\|_1}{\hat{\phi}_0}(\omega(\phi,\Delta))^{1/2}.$$

The above inequality together with (6) yields (4) immediately. □

Acknowledgements

Two anonymous referees have carefully read the paper and have provided to us numerous constructive suggestions. As a result, the overall quality of the paper has been noticeably enhanced, to which we feel much indebted and are grateful.

References

[1] N. Arouszajiu, Theory of reproducing kernels, Trans. Math. Soc. 68 (1950) 337–404.
[2] A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, IEEE Trans. Inf. Theor. 39 (1993) 930–945.
[3] F.L. Cao, T.F. Xie, Z.B. Xu, The estimate for approximation error of neural networks: a constructive approach, Neurocomputing 71 (2008) 626–630.
[4] F.L. Cao, R. Zhang, The errors of approximation for feedforward neural networks in L^p metric, Math. Comput. Modell. 49 (2009) 1563–1572.
[5] D.B. Chen, Degree of approximation by superpositions of a sigmoidal function, Approx. Theor. Appl. 9 (1993) 17–28.
[6] 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.
[7] C.K. Chui, X. Li, Approximation by ridge functions and neural networks with one hidden layer, J. Approx. Theor. 70 (1992) 131–141.
[8] F. Cucker, D.X. Zhou, Learning Theory: an Approximation Theory Viewpoint, Cambridge University Press, Cambridge, 2007.
[9] G. Cybenko, Approximation by superpositions of sigmoidal function, Math. Control Signals Syst. 2 (1989) 303–314.
[10] R. DeVore, G. Lorentz, Constructive Approximation, Springer-Verlag, Berlin, 1993.
[11] S. Ferrari, R.F. Stengel, Smooth function approximation using neural networks, IEEE Trans. Neural Networks 16 (2005) 24–38.
[12] K.I. Funahashi, On the approximate realization of continuous mapping by neural networks, Neural Networks 2 (1989) 183–192.
[13] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximation, Neural Networks 2 (1989) 359–366.
[14] 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.
[15] X. Li, On simultaneous approximation of by radial basis function neural networks, Appl. Math. Comput. 95 (1998) 75–89.
[16] G.G. Lorentz, Approximation of Functions, Rinehart and Winston, New York, 1966.
[17] V. Maiorov, R.S. Meir, Approximation bounds for smooth function in $C(\mathbb{R}^d)$ by neural and mixture networks, IEEE Trans. Neural Networks 9 (1998) 969–978.
[18] Y. Makovoz, Uniform approximation by nerual networks, J. Approximate Theor. 95 (1998) 215–228.
[19] H.N. Mhaskar, C.A. Micchelli, Degree of approximation by neural networks with a single hidden layer, Adv. Appl. Math. 16 (1995) 151–183.
[20] I. Steinwart, A. Christmann, Support Vector Machines, Springer, New York, 2008.
[21] A.N. Tikhonov, V.Y. Arsenin, Solutions of III-Posed Problems, Wiley, New york, 1977 (translated form the Russian).
[22] H. Wendland, Scattered Data Approximation, Cambridge University Press, Cambridge, 2005.
[23] Z.B. Xu, F.L. Cao, The essential order of approximation for neural networks, Sci. China Ser. F 47 (2004) 97–112.