数 学 年 刊
2009,30A(6):741–750

神经网络的加权本质逼近阶 ***

王 建 军*　徐 宗 本**

提要 证明了具有单一隐层的神经网络在 L^q_ω 的逼近，获得了网络逼近的上界估计和下界估计．这一结果揭示了神经网络在加权逼近的意义下，网络的收敛阶与隐层单元个数之间的关系，为神经网络的应用提供了重要的理论基础．

关键词 逼近估计，神经网络， Jacobi 权
MR (2000) 主题分类 41A36, 41A25
中图法分类 O175.29
文献标志码 A
文章编号 1000-8314(2009)06-0741-10

1 引　言

近年来，有许多学者对神经网络逼近问题进行了研究，取得了一系列成果．神经网络已经在工程、计算机、物理、生物等科学中有了广泛的应用．它之所以得到广泛应用，其主要原因之一是它具有一定意义上的万有逼近性 (见 [1–12])．所有这些研究的一个典型推论是：定义在 R^d 中紧集上的任何一个连续函数可以通过一个单一隐层的三层前向神经网络任意逼近．一个具有单一隐层，含 d 个输入， 1 个输出的三层前向神经网络数学上可表示为

$$\mathcal{N}(x)=\sum_{i=1}^{m}c_i\sigma\Big(\sum_{j=1}^{d}w_{ij}x_j+\theta_i\Big),\quad x\in\mathcal{R}^d,\ d\geqslant 1,\tag{1.1}$$

其中 $1\leqslant i\leqslant m$, $\theta_i\in\mathcal{R}$ 是阈值， $w_i=(w_{i1},w_{i2},\cdots,w_{id})^{\mathrm{T}}\in\mathcal{R}^d$ 是输入层与隐层第 i 个神经元的连接权值， c_i 是隐层与输出层之间的连接权值， σ 是隐层节点的激活函数 (传递函数). 通常情况下，网络激活函数 σ 取为 sigmoid 型函数，即满足 $\sigma(t)\to 1$ $(t\to\infty)$, $\sigma(t)\to 0$ $(t\to-\infty)$ 的函数.

许多应用问题都被转化为利用神经网络逼近多元函数的问题．关于神经网络逼近多元函数方面的工作在过去的一些年得到了广泛的研究，得到了一些重要的关于逼近稠密性的结果．例如，文 [1] 中，在 sigmoid 型激活函数等弱型条件下，对于定义在 R^d 中紧子集 K 上的任何连续多元函数，存在一个形式如 (1.1) 所示的三层前向神经网络，使它以任何精度逼近该函数．此后，许多学者建立了各种不同形式的逼近多元函数的稠密性和复杂性结果 (见 [2, 5, 10–13])．这些研究基本上都是关于神经网络逼近能力的定性研究，有关神经网络逼近的定量研究，特别是反映网络的逼近速度与隐层拓扑之间关系的研究，最近开始引起人们的强烈关注 (见 [14–22])．已有的有关神经网络逼近的定量研究基本上

本文 2009 年 4 月 27 日收到.
*西南大学数学与统计学院，重庆 400715; 西安交通大学理学院信息与系统科学研究所，西安 710049.
E-mail: wangjianjun@mail.xjtu.edu.cn
**西安交通大学理学院信息与系统科学研究所，西安 710049. E-mail: zbxu@mail.xjtu.edu.cn
***国家 973 计划 (No. 2007CB311000), 国家自然科学基金 (No. 10726040, No. 10701062, No. 10826081), 教育部科学技术重点项目 (No. 108176), 中国博士后基金 (No. 20080431237) 和重庆市科委自然科学基金 (No. CSTC2009BB2306) 资助的项目.

限于对网络逼近速度的上界估计 (见 [14, 17, 18, 21–24]). 很明显，仅有上界估计对于刻画一个神经网络的逼近能力是远远不够的，它并不能准确反映网络的本质逼近阶 (换言之，一个网络能够且仅能够达到的逼近精度). 我们应该不仅要对网络逼近速度的上界进行估计，而且更希望对其下界作出估计. 特别地，我们希望获得神经网络逼近速度的上下界具有相同阶的估计，从而能精确确定网络的本质逼近阶 (见 [25, 26]), 以从根本上澄清网络的逼近阶与网络隐层单元数的估计之间的关系. 文 [25, 26] 对满足一定条件的激活函数，基于多元函数逼近理论，给出该类神经网络逼近的上、下界估计和本质逼近阶估计，澄清了网络的逼近能力与网络拓扑之间的关系. 本文基于函数逼近论的知识，结合多元逼近的思想，得到了神经网络加权逼近的上下界估计和加权本质逼近界估计，同时给出了网络达到一定逼近速度所需的隐层单元数；推广了没有加权情况下的结果.

2 记号和主要结果

本文将采用如下记号： $\mathbf{N},\mathbf{R},\mathbf{R}+$ 分别表示非负整数，实数和非负实数. R^d 表示 d 维实 Euclid 空间；假设 S 是 d 维欧氏空间 R^d 中的单纯形

$$S=S_d=\Big\{\mathbf{x}=(x_1,x_2\cdots,x_d):x_i\geqslant 0,\quad |\mathbf{x}|=\sum_{i=1}^d x_i\leqslant 1\Big\}.$$

令 V_S 是 S 的边界方向集, 它的每个向量具有单位长度, 其方向平行于 S 的边界 (若两个方向完全相反向量，视作同一个向量), 即 $V_S=\Big\{\mathbf{e}_i,\frac{\mathbf{e}_i-\mathbf{e}_j}{\sqrt{2}},\mathbf{e}_i=(0,0,\cdots,0,\overset{\text{第 } i \text{ 个}}{1},0,\cdots,0)\Big\}$ $(i=1,\cdots,d)$. 对任意 $\mathbf{x}\in S$, $\mathbf{e}\in V_S$, 权函数 $\varphi_{\mathbf{e}}(\mathbf{x})$ 定义为

$$\varphi_{\mathbf{e}}(\mathbf{x})=\begin{cases}\sqrt{x_i(1-|\mathbf{x}|)}, & \mathbf{e}=\mathbf{e}_i, \quad 1\leqslant i\leqslant d,\\ \sqrt{2x_ix_j}, & \mathbf{e}=\frac{\mathbf{e}_i-\mathbf{e}_j}{\sqrt{2}}, \quad 1\leqslant i<j\leqslant d.\end{cases}$$

我们用 $L^q_\omega(S),1\leqslant q<\infty$ 表示 S 上带 Jacobi 权 $\omega(\mathbf{x})$ $(\omega(\mathbf{x})=\mathbf{x}^\alpha(1-|\mathbf{x}|)^\beta)$ 意义下的 Lebesgue 可积函数空间， $L^\infty_\omega(S)=C(S)$ 表示 S 上带 Jacobi 权意义下的连续函数空间, 分别赋予如下范数：

$$\|f\|_{q,\omega}=\begin{cases}\Big(\displaystyle\int_S|\omega(\mathbf{t})f(\mathbf{t})|^q\mathrm{d}\mathbf{t}\Big)^{1/q}, & 1\leqslant q<\infty,\\ \displaystyle\max_{\mathbf{t}\in S}|\omega(\mathbf{t})f(\mathbf{t})|, & q=\infty,\end{cases}$$

其中 $\mathbf{x}=(x_1,x_2,\cdots,x_d)$, $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_d),0<\alpha_i,\beta<1,1\leqslant i\leqslant d$, $\mathbf{x}^\alpha=x_1^{\alpha_1}x_2^{\alpha_2}\cdots x_d^{\alpha_d}$, $|\mathbf{x}|=x_1+x_2+\cdots+x_d$. 任意函数 $f\in L^q_\omega(S),1\leqslant q\leqslant\infty$, f 的沿方向 $\mathbf{e}$, 步长为 h 的 r-阶对称差分为

$$\Delta^r_{h,\mathbf{e}}f(\mathbf{x})=\begin{cases}\displaystyle\sum_{i=0}^r(-1)^{r-i}\binom{r}{i}f\Big(\mathbf{x}+\Big(\frac{r}{2}-i\Big)h\mathbf{e}\Big), & \mathbf{x},\mathbf{x}\pm\dfrac{\mathbf{rh}}{\mathbf{2}}\in S,\\ 0, & \text{否则}.\end{cases}$$

于是 f 的 r-阶加权连续模定义如下：

$$\Omega^r_\varphi(f,t)_{q,\omega}=\sup_{0<h\leqslant t}\sum_{\mathbf{e}\in S}\|\Delta^r_{\varphi_{\mathbf{e}}(\mathbf{x})h,\mathbf{e}}f(\mathbf{x})\|_{q,\omega}.$$

函数的连续模和高阶连续模是分别刻画函数连续性程度和光滑性程度的工具，它在逼近论中起非常重要的作用. 本文中将利用其刻画神经网络的加权逼近阶. 我们将考虑一类激活函数 Υ_M 的神经网络，其中 Υ_M 中的元素 ϕ 满足： (1) $\phi:R\to[0,1]$, 并且 ϕ 具

有 $M+1$ 阶连续导函数 $\phi^k, k=1,2,\cdots,M+1$; (2) ϕ^k 一致有界，并且至少存在一个点 $\theta_0 \in R$, 使得 $\phi^k(\theta_0) \neq 0$, $1 \leqslant k \leqslant M+1$. 满足这样条件的激活函数很多，例如，标准的 sigmoidal 函数 $\phi(x) = (1+\mathrm{e}^{-\alpha x})^{-1}$ $(\alpha > 0)$. 基于上述记号，本文所获结果如下：

定理 2.1　对任何 $f \in L_w^q(K), 1 \leqslant q \leqslant \infty, \phi \in \Upsilon_M$, 存在一个以 ϕ 为激活函数，d 个输入，一个输出，隐层神经元个数为 $m = N_0 + 2N_1 + \cdots + (n+1)N_n$ (或大于这个下界) 的形式如 (1.1) 的三层前向神经网络 $\mathcal{N}_n(\mathbf{x})$, 使得

$$\|\mathcal{N}_n - f\|_{q,w} \leqslant C\Big(\Omega_\varphi^2\Big(f, \frac{1}{\sqrt{n}}\Big)_{q,w} + \frac{1}{n}\|f\|_{q,w}\Big), \tag{2.1}$$

以及存在 $0 < \delta < 1$, 使得

$$\Omega_\varphi^2\Big(f, \frac{1}{\sqrt{n}}\Big)_{q,w} \leqslant \frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^{\delta}\|\mathcal{N}_i - f\|_{q,w}. \tag{2.2}$$

特别地，当且仅当 $\omega_\varphi^2(f,t)_{p,w} = O(t^\alpha)$, $0 < \alpha \leqslant 2$ 时，有

$$\|\mathcal{N}_n - f\|_{p,w} = O(n^{-\alpha/2}), \tag{2.3}$$

这里及以下 C 表示与 $f, n, \mathbf{x}$ 无关的正常数 (在不同处其值可以不同), 而 $N_i = \binom{i+d-1}{d-1}$, $0 \leqslant i \leqslant n$.

上述定理中的 (2.1) 是对网络逼近速度的上界估计，反映了网络加权逼近速度上界与隐层神经元个数之间的关系．它说明网络加权逼近速度的上界受控于被逼近函数的二阶加权连续模及隐层神经元个数．特别地，它显示了前向神经网络加权逼近速度达到预定程度时，网络所需的隐层单元个数．显然当 $n \to \infty$ 时，$\|\mathcal{N}_n - f\|_{q,w} \to 0$, 因此，它也说明对任意连续或可积函数 f, 存在三层前向神经网络对其任意加权逼近．定理中的 (2.2) 是神经网络加权逼近速度的下界估计，它表明网络加权逼近函数的速度的平均 $\frac{1}{n}\sum\limits_{i=1}^{n}(\frac{n}{i})^\delta\|\mathcal{N}_i - f\|_{q,w}$ 不会低于所被逼近函数的二阶加权连续模. 定理中的 (2.3) 刻画了神经网络的加权本质逼近阶，它表明当被逼近函数满足 $\omega_\varphi^2(f,t)_{p,w} = O(t^\alpha)$, $0 < \alpha \leqslant 2$ 时，神经网络的加权本质逼近阶是 $O(n^{-\alpha/2})$. 同时，它也表明被逼近函数的光滑性与网络加权逼近速度的关系，即被逼近函数的光滑性越高，网络加权逼近速度越快；反之亦然.

3　神经网络模型

本节将在 $L_\omega^q(S), 1 \leqslant q \leqslant \infty$ 空间构造实现函数加权逼近的三层前向神经网络. 我们用 Z_+^d 表示 R^d 中所有非负多重整数集. 对任意 $\mathbf{x} = (x_1, x_2, \cdots, x_d), \mathbf{y} = (y_1, y_2, \cdots, y_d) \in R^d$, 若 $x_i \leqslant y_i$ $(1 \leqslant i \leqslant d)$, 则 $\mathbf{x} \leqslant \mathbf{y}$. 记 $\mathbf{k} = (k_1, k_2, \cdots, k_d) \in Z_+^d$, $\mathbf{x}^{\mathbf{k}} = x_1^{k_1}x_2^{k_2}\cdots x_d^{k_d}$, $\mathbf{k}! = k_1!k_2!\cdots k_d!$. 对正整数 $n \geqslant 1$, Bernstein 算子 $\mathcal{B}_n$ 定义为

$$\mathcal{B}_n(f, \mathbf{x}) = \sum_{|\mathbf{k}| \leqslant n} P_{n,\mathbf{k}}(\mathbf{x}) f\Big(\frac{\mathbf{k}}{n}\Big), \tag{3.1}$$

其中 $\mathbf{x} \in S$, $f \in \mathbf{C}(S)$,

$$P_{n,\mathbf{k}}(\mathbf{x}) = \frac{n!}{\mathbf{k}!(n-|\mathbf{k}|)!}\mathbf{x}^{\mathbf{k}}(1-|\mathbf{x}|)^{n-|\mathbf{k}|}. \tag{3.2}$$

对于函数 $f\in L(S)$, Bernstein-Durrmeyer 算子定义为

$$\mathcal{D}_n(f,\mathbf{x})=\mathcal{D}_{n,d}(f,\mathbf{x})=\sum_{|\mathbf{k}|\leqslant n}\Phi_{n,\mathbf{k}}(f)P_{n,\mathbf{k}}(\mathbf{x}),\quad f\in L(S),$$

其中

$$\Phi_{n,\mathbf{k}}(f)=\frac{(n+d)!}{n!}\int_S P_{n,\mathbf{k}}(\mathbf{t})f(\mathbf{t})\mathrm{d}\mathbf{t}.$$

在文 [27] 中，作者获得了关于 Bernstein 算子加 Jacobi 权逼近的结果．对于 Bernstein-Durrmeyer 算子，我们已在文 [28] 中建立了其加权逼近的结果.

命题 3.1 (见 [28]) 对 $f\in L^q(S), 1\leqslant q\leqslant\infty$, 有

$$\|\mathcal{D}_nf-f\|_{q,w}\leqslant C\Big(\Omega_\varphi^2\Big(f,\frac{1}{\sqrt{n}}\Big)_{q,w}+\frac{1}{n}\|f\|_{q,w}\Big),$$

$$\Omega_\varphi^2\Big(f,\frac{1}{\sqrt{n}}\Big)_{q,w}\leqslant\frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^{\delta}\|\mathcal{D}_if-f\|_{q,w}.$$

令

$$N_p=\begin{pmatrix}p+d-1\\ d-1\end{pmatrix},\quad I_p=\begin{pmatrix}p+d-2\\ d-2\end{pmatrix}.$$

注意到，存在 Z_+^d 中的多重整数 $\mathbf{i}$, 使 $i_1+i_2+\cdots+i_d=p$, 以及存在 Z_+^{d-1} 中的多重整数 $\mathbf{j}'$, 使 $j_1'+j_2'+\cdots+j_{d-1}'=p$. 显然 $N_p=N_{p-1}+I_p$. 用 $\mathbf{j}'_{N_{p-1}+l}, 1\leqslant l\leqslant I_p$, 表示 Z_+^{d-1} 中的多重整数 $\mathbf{j}'$, 其满足 $j_1'+j_2'+\cdots+j_{d-1}'=p$, 则 $\{\mathbf{j}_l':1\leqslant l\leqslant N_p\}$ 是 Z_+^{d-1} 中多重整数 $\mathbf{j}'$ 的一个集合，其满足 $j_1'+j_2'+\cdots+j_{d-1}'\leqslant p$. 对于每一个 p, 令

$$\mathbf{i}_l^{(p)}=(p-|\mathbf{j}_l'|,\mathbf{j}_l'),\quad 1\leqslant l\leqslant N_p.$$

每一 $\mathbf{i}_l^{(p)}$ 是 Z_+^d 中的一个多重整数，并且 $|\mathbf{i}_l^{(p)}|=p, 1\leqslant l\leqslant N_p$. 定义 $\mathbf{p}_l=(1,\mathbf{j}_l'), 1\leqslant l$. 则对于每一固定的整数 $p\geqslant 0$, 令

$$\mathbf{p}_l^{(p)}=\frac{1}{2(1+p)}\mathbf{p}_l,\quad 1\leqslant l\leqslant N_p,$$

则有 $|\mathbf{p}_l^{(p)}|\leqslant\frac{1}{2}$, $1\leqslant l\leqslant N_p$.

由文 [8, 9], Bernstein-Durrmeyer 算子 $\mathcal{D}_nf$ 可以表示成下面命题中的形式.

命题 3.2 对 $f\in L(S)$, 有

$$\mathcal{D}_nf(\mathbf{x})=\sum_{p=0}^{n}\sum_{l=1}^{N_p}d_l^{(p)}\langle\mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle^p,\tag{3.3}$$

其中 $(d_1^{(p)},d_2^{(p)},\cdots,d_{N_p}^{(p)})^{\mathrm{T}}$ 满足如下方程

$$\begin{pmatrix}(\mathbf{j}_1')^{\mathbf{j}_1'} & (\mathbf{j}_2')^{\mathbf{j}_1'} & \cdots & (\mathbf{j}_{N_p}')^{\mathbf{j}_1'}\\ \vdots & \vdots & & \vdots\\ (\mathbf{j}_1')^{\mathbf{j}_{N_p}'} & (\mathbf{j}_2')^{\mathbf{j}_{N_p}'} & \cdots & (\mathbf{j}_{N_p}')^{\mathbf{j}_{N_p}'}\end{pmatrix}\begin{pmatrix}d_1^{(p)}\\ \vdots\\ d_{N_p}^{(p)}\end{pmatrix}=\frac{(2(1+n))^p}{p!}\begin{pmatrix}\mathbf{i}_1^{(p)}!c_1^{(p)}(f)\\ \vdots\\ \mathbf{i}_{N_p}^{(p)}!c_{N_p}^{(p)}(f)\end{pmatrix},$$

并且

$$c_l^{(p)}(f)=\frac{n!}{(n-p)!}\sum_{\mathbf{q}\leqslant\mathbf{i}_l^{(p)}}\Phi_{n,\mathbf{q}}(f)\frac{(-1)^{|\mathbf{i}_l^{(p)}-\mathbf{q}|}}{\mathbf{q}!(\mathbf{i}_l^{(p)}-\mathbf{q})!},\quad 1\leqslant l\leqslant N_p.$$

通过以上讨论，我们确定逼近任何连续或可积函数的三层前向神经网络

$$\mathcal{N}_n(\mathbf{x}) = \sum_{p=0}^{n} \sum_{l=1}^{N_p} d_l^{(p)} \sum_{i=1}^{p+1} c_{i,p} \varphi(w_{i,p} \langle \mathbf{x} \cdot \mathbf{p}_l^{(n)} \rangle + \theta).$$

容易看出，上述神经网络 $\mathcal{N}_n(\mathbf{x})$ 的隐层单元数目为

$$m = N_0 + 2N_1 + \cdots + (n+1)N_n, \quad \text{其中 } N_i = \binom{i+d-1}{d-1},\ 0 \leqslant i \leqslant n.$$

4 一元多项式的神经网络加权逼近

对任意的 $r \in Z$, 令

$$\mathcal{P}_r(x) = a_0 + a_1 x + \cdots + a_r x^r, \quad x \in \mathcal{R} \tag{4.1}$$

为 r-阶一元多项式. 我们将研究区间 $[0,1]$ 上的多项式 $\mathcal{P}_r(x)$ 的神经网络加 Jacobi 权逼近. 关于神经网络逼近多项式也有很多研究，如文 [3, 14] 等. 我们将采用不同于上述文献的方法，给出神经网络带权逼近一元多项式的结果，即

命题 4.1 设 $\phi \in \Upsilon_M$ 是网络的激活函数，$\mathcal{P}_r(x)$ 为 (4.1) 中的一元多项式，$x \in [0,1]$. 对任意的 $1 \leqslant q \leqslant \infty$, $\varepsilon > 0$, 存在一个形如 (1.1) 的隐层单元数不少于 $\sum\limits_{0 \leqslant j \leqslant r} (j+1) = \frac{(r+1)(r+2)}{2}$ (独立于 ε), 使得

$$\|\mathcal{N}_n - \mathcal{P}_r\|_{q,\omega} < \varepsilon, \tag{4.2}$$

其中

$$N_n(x) = \sum_{0 \leqslant j \leqslant r} \sum_{0 \leqslant i \leqslant j} c_{i,j} \phi(wx + \theta), \quad n \geqslant \frac{(r+1)(r+2)}{2}, \tag{4.3}$$

$$w = h(2i - j), \tag{4.4}$$

$$c_{i,j} = a_j \frac{1}{\phi^{(|j|)}(\theta)} (2h)^{-j} (-1)^i \binom{j}{i}. \tag{4.5}$$

证 首先，对 $q = \infty$, $x \in [0,1]$, 我们证明 (4.2). 令 $M_1 = \max |x|$, 因为

$$\phi^{(j)}(wx + \theta) = \frac{\partial^j}{\partial \omega^j} [\phi(wx + \theta)] = x^j \phi^{(|j|)}(wx + \theta), \tag{4.6}$$

于是

$$\phi_{j,x}(\theta) \doteq \phi^{(j)}(\omega x + \theta)|_{\omega = 0} = x^j \phi^{(|j|)}(\theta), \tag{4.7}$$

因而

$$x^j = \frac{\phi_{j,x}(\theta)}{\phi^{(|j|)}(\theta)}. \tag{4.8}$$

由 (4.1), (4.6),

$$p_r(x) = \sum_{0 \leqslant j \leqslant r} a_j \frac{\phi_{j,x}(\theta)}{\phi^{(|j|)}(\theta)}, \tag{4.9}$$

对任意固定的 $b\in R$, 我们考虑差分

$$\triangle_{h,x}^{j}\phi(\theta)=\sum_{0\leqslant i\leqslant j}(-1)^{i}\binom{j}{i}\phi(h(2i-j)x+\theta). \tag{4.10}$$

由三角不等式和差分的表示，得到

$$\begin{aligned}
&\Big|p_r(x)-\sum_{0\leqslant j\leqslant r}a_j\frac{1}{\phi^{(|j|)}(\theta)}(2h)^{-j}\triangle_{h,x}^{j}\phi(\theta)\Big|\\
&=\Big|\sum_{0\leqslant j\leqslant r}a_j\frac{\phi_{j,x}(\theta)}{\phi^{(|j|)}(\theta)}-\sum_{0\leqslant j\leqslant r}a_j\frac{1}{\phi^{(|j|)}(\theta)}(2h)^{-j}\triangle_{h,x^j}\phi(\theta)\Big|\\
&\leqslant\sum_{0\leqslant j\leqslant r}|a_j|\frac{1}{|\phi^{(|j|)}(\theta)|}|\phi_{j,x}(\theta)-(2h)^{-|j|}\triangle_{h,x}^{j}\phi(\theta)|\\
&=\sum_{0\leqslant j\leqslant r}|a_j|\frac{1}{|\phi^{(|j|)}(\theta)|}\\
&\quad\cdot\Big|x^j\phi^{(|j|)}(\theta)-x^j(2h)^{-j}\int_{-h}^{h}\int_{-h}^{h}\cdots\int_{-h}^{h}\phi^{(|j|)}(\theta+(\tau_1+\tau_2+\cdots+\tau_j)x)\mathrm{d}\tau_1\mathrm{d}\tau_2\cdots\mathrm{d}\tau_j\Big|\\
&=\sum_{0\leqslant j\leqslant r}|a_j|\frac{1}{|\phi^{(|j|)}(\theta)|}x^j\Big|(2h)^{-j}\int_{-h}^{h}\int_{-h}^{h}\cdots\int_{-h}^{h}\phi^{(|j|)}(\theta)\mathrm{d}\tau_1\mathrm{d}\tau_2\cdots\mathrm{d}\tau_j\\
&\quad-(2h)^{-j}\int_{-h}^{h}\int_{-h}^{h}\cdots\int_{-h}^{h}\phi^{(|j|)}(\theta+(\tau_1+\tau_2+\cdots+\tau_j)x)\mathrm{d}\tau_1\mathrm{d}\tau_2\cdots\mathrm{d}\tau_j\Big|\\
&\leqslant\sum_{0\leqslant j\leqslant r}|a_j|\frac{1}{|\phi^{(|j|)}(\theta)|}x^jL(\phi^{(|j|},2M_1jh)\\
&\leqslant 2M_1h\sum_{0\leqslant j\leqslant r}|a_j|jM_1^j\frac{\|\phi^{(|j|+1)}\|_\infty}{|\phi^{(|j|)}(\theta)|},
\end{aligned}\tag{4.11}$$

其中 $L(\phi,\delta)=\sup\limits_{|t-x|<\delta}|\phi(x)-\phi(t)|$ 是函数 ϕ 的连续模. 当 ϕ 有连续导数时，有 $L(\phi,\delta)\leqslant\delta|\phi'|_\infty$. 令

$$M_0=\max\{|\phi^{(i)}(\theta)|,i=0,1,\cdots,r+1\}.$$

由于 $|\omega(x)|\leqslant 1$ $(x\in[0,1])$, 故

$$\begin{aligned}
\Big|\omega(x)(p_r(x)-\sum_{0\leqslant j\leqslant r}a_j\frac{1}{\phi^{(|j|)}(\theta)}(2h)^{-j}\triangle_{h,x}^{j}\phi(\theta))\Big|&\leqslant 2M_1M_0h\sum_{0\leqslant j\leqslant r}\frac{|a_j|j}{|\phi^{(|j|)}(\theta)|}M_1^j\\
&\leqslant Mh,
\end{aligned}\tag{4.12}$$

其中 $M=2M_1M_0\sum\limits_{0\leqslant j\leqslant r}\frac{|a_j|j}{|\phi^{(|j|)}(\theta)|}M_1^j$. 由 (4.11),(4.12), 我们构造网络如下：

$$N_n(x)=\sum_{0\leqslant j\leqslant r}\sum_{0\leqslant i\leqslant j}c_{i,j}\phi(wx+\theta), \tag{4.13}$$

其中

$$w=h(2i-j), \tag{4.14}$$

$$c_{i,j}=a_j\frac{1}{\phi^{(|j|)}(\theta)}(2h)^{-j}(-1)^i\binom{j}{i}. \tag{4.15}$$

从 (4.12), 得到

$$|\omega(x)(p_{\mathbf{r}}(\mathbf{x})-N_n(\mathbf{x}))|\leqslant Mh.$$

令 $h<\frac{\varepsilon}{M}$, 则

$$|\omega(x)(p_{\mathbf{r}}(\mathbf{x})-N_n(\mathbf{x}))|<\epsilon.$$

即 (4.2) 对 $q=\infty$ 是成立的.

对 $1\leqslant q<\infty$, 从 $q=\infty$ 的结果，有

$$\begin{aligned}\|\mathcal{N}_n-\mathcal{P}_r\|_{q,\omega}&=\Big\{\int_0^1|\omega(x)(\mathcal{N}_n(x)-\mathcal{P}_r(x))|^q\Big\}^{1/q}\mathrm{d}x\\&\leqslant\Big\{\max_{x\in[0,1]}|\omega(x)(\mathcal{N}_n(x)-\mathcal{P}_r(x))|^q\Big\}^{1/q}\\&\leqslant\max_{x\in[a,b]}|\omega(x)(\mathcal{N}_n(x)-\mathcal{P}_r(x))|\\&\leqslant\varepsilon.\end{aligned}$$

由 ε 的任意性，即可得到 (4.2). 至于所构造网络的隐层单元数目，由 (4.13) 知，隐层单元数目不小于 $\sum\limits_{0\leqslant j\leqslant r}(j+1)=\frac{(r+1)(r+2)}{2}$.

从命题 4.1, 我们很容易得到如下推论.

推论 4.1 对固定的整数 r, $P_r(x)=a_rx^r$, $x\in[0,1]$ 是 r 次齐次多项式. 对任意的 $\varepsilon>0$, 存在一个形如 (1.1), 隐层单元数目不少于 $r+1$ 的神经网络，使得

$$\|\mathcal{N}_n-\mathcal{P}_r\|_{q,\omega}<\varepsilon,\tag{4.16}$$

其中

$$N_n(x)=\sum_{i=0}^{r}c_{i,r}\phi(wx+\theta),\quad n\geqslant r+1,\tag{4.17}$$

$$w=h(2i-r),\tag{4.18}$$

$$c_{i,r}=a_r\frac{1}{\phi^{(r)}(\theta)}(2h)^{-r}(-1)^i\binom{r}{i}.\tag{4.19}$$

5 定理的证明

下面我们将证明定理 2.1. 令 $\mathbf{x}\cdot\mathbf{p}_l^{(n)}=u$, 将 $\langle\mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle^p$ 看作定义在 $[0,1]$ 上的关于 u 的阶数为 p 的齐次多项式. 由推论 4.1, $\langle\mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle^p$ 可以被如下神经网络

$$\mathcal{N}_{M_p}=\sum_{i=1}^{M_p}c_{i,p}\phi(w_{i,p}\langle\mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle+\theta),\quad c_{i,p},w_{i,p}\in R,\ M_p\geqslant p+1\tag{5.1}$$

任意加权逼近，即

$$\|\mathcal{N}_{M_p}-\langle\mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle^p\|_{q,\omega}<\varepsilon,\tag{5.2}$$

因而构造神经网络

$$\mathcal{N}_n(\mathbf{x})=\sum_{p=0}^{n}\sum_{l=1}^{N_p}d_l^{(p)}\sum_{i=1}^{M_p}c_{i,p}\varphi(w_{i,p}\langle\mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle+\theta),\quad c_{i,p},w_{i,p}\in R,\ M_p\geqslant p+1.\tag{5.3}$$

利用命题 3.1, 有

$$\begin{aligned}\|\mathcal{N}_n - f\|_{q,\omega} &\leqslant \|\mathcal{N}_n - \mathcal{D}_n f\|_{q,\omega} + \|\mathcal{D}_n f - f\|_{q,\omega} \\ &\leqslant C\Big(\Omega_\varphi^2\Big(f, \frac{1}{\sqrt{n}}\Big)_{q,w} + \frac{1}{n}\|f\|_{q,\omega}\Big) + \|\mathcal{N}_n - \mathcal{D}_n f\|_{q,\omega}. \end{aligned} \tag{5.4}$$

由 (5.2), (5.3),

$$\begin{aligned}\|\mathcal{N}_n - \mathcal{D}_n f\|_{q,\omega} &= \Big\| \sum_{p=0}^{n}\sum_{l=1}^{N_p} d_l^{(p)}\Big(\langle \mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle^p - \sum_{i=1}^{M_p} c_{i,p}\phi(w_{i,p}\langle \mathbf{x}\cdot\mathbf{p}_l^{(n)}\rangle + \theta)\Big)\Big\|_{q,\omega} \\ &\leqslant \varepsilon \sum_{p=0}^{n}\sum_{l=1}^{N_p} |d_l^{(p)}|. \end{aligned} \tag{5.5}$$

于是， (5.4) 中的第 3 项可以任意小．因此， (2.1) 成立.

下面我们证明 (2.2). 由命题 3.1, 有

$$\begin{aligned}\Omega_\varphi^2\Big(f, \frac{1}{\sqrt{n}}\Big)_{q,\omega} &\leqslant \frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^\delta \|\mathcal{D}_i f - f\|_{q,\omega} \\ &\leqslant \frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^\delta \|\mathcal{D}_i f - \mathcal{N}_i\|_{q,\omega} + \frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^\delta \|\mathcal{N}_i - f\|_{q,\omega} \\ &\leqslant C\Big(\varepsilon\frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^\delta \sum_{p=0}^{i}\sum_{l=1}^{N_p}|d_l^{(p)}| + \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^\delta \|\mathcal{N}_i - f\|_{q,\omega}\Big). \end{aligned}$$

令 $\epsilon \to 0$, 得到

$$\Omega_\varphi^2\Big(f, \frac{1}{\sqrt{n}}\Big)_{q,\omega} \leqslant \frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^\delta \|\mathcal{N}_i - f\|_{q,\omega}. \tag{5.6}$$

最后，如果 $\Omega_\varphi^2(f,t)_{q,w} = O(t^\alpha)$, $0 < \alpha \leqslant 2$, 由 (5.4), (5.5),

$$\|\mathcal{N}_n - f\|_{q,w} \leqslant \Big(n^{-\alpha/2} + \frac{1}{n}\|f\|_{q,w}\Big) = O(n^{-\alpha/2}). \tag{5.7}$$

相反地，如果 (5.7) 成立，由 (5.6) 获得

$$\Omega_\varphi^2\Big(f, \frac{1}{\sqrt{n}}\Big)_{q,\omega} \leqslant \frac{C}{n}\sum_{i=1}^{n}\Big(\frac{n}{i}\Big)^\delta i^{-\alpha/2} = O\Big(\frac{1}{n^{1-\delta}}\sum_{i=1}^{n} i^{-1-\alpha/2}\Big) = O(n^{-\alpha/2}). \tag{5.8}$$

上述 (5.8) 中的最后一个不等式通过数学归纳法很容易证明.

参 考 文 献

[1] Cybenko G., Approximation by superpositions of sigmoidal function [J], *Math. of Control Signals and Sys.*, 1989, 2:303–314.

[2] Funahashi K. I., On the approximate realization of continuous mappings by neural networks [J], *Neural Networks*, 1989, 2:183–192.

[3] Hornik K., Stinchombe M. and White H., Multilayer feedforward networks are universal approximation [J], *Neural Networks*, 1989, 2:359–366.

[4] Leshno M., Lin V. Y., Pinks A. and Schocken S., Multilayer feedforward networks with a nonpolynomial activation function can approximate any function [J], *Neural Networks*, 1993, 6:861–867.

[5] Yoshifusa I., Approximation of functions on a compact set by finite sums of sigmoid function without scaling [J], *Neural Networks*, 1991, 4:817–826.

[6] Attali J. G. and Pages G., Approximation of functions by a multilayer perceptron: a new approach [J], *Neural Networks*, 1997, 10:1069–1081.

[7] Cardaliaguet P. and Euvrard G., Approximation of a function and its derivatives with a neural networks [J], *Neural Networks*, 1992, 5:207–220.

[8] Chui C. K. and Li X., Approximation by ridge functions and neural networks with one hidden layer [J], *J. Approx. Theory*, 1992, 70:131–141.

[9] Chui C. K. and Li X., Neural networks with one hidden layer [M]//Jetter K. and Utreras F. I., eds., Multivariate Approximation: From CAGD to Wavelets, Singapore: World Scientific Press, 1993:77–89.

[10] Chen T. P. and Chen H., Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to a dynamic system [J], *IEEE Transactions on Neural Networks*, 1995, 6:911–917.

[11] 陈天平，神经网络及其在系统识别应用中的逼近问题 [J], 中国科学 (A 辑), 1994, 24(1):1–7.

[12] Chen T. P. and Chen H., Approximation capability to functions of several variables, nonlinear functions, and operators by radial function neural networks [J], *IEEE Trans. Neural Networks*, 1995, 6:904–910.

[13] Chen T. P., A unified approach for neural networks-like approximation of nonlinear functions [J], *Neural Networks*, 1998, 11:981–983.

[14] 曹飞龙, 徐宗本, 多变元周期函数的神经网络逼近：逼近阶估计 [J], 计算机学报, 2001, 24(9):903–908.

[15] Kůrková V., Kainen P. C. and Kreinovich V., Estimates for the number of hidden units and variation with respect to half-space [J], *Neural Networks*, 1997, 10:1068–1078.

[16] Mhaskar H. N. and Micchelli C. A., Approximation by superposition of a sigmoidal function [J], *Advances in Applied Mathematics*, 1992, 13:350–373.

[17] Mhaskar H. N., Neural networks for optimal approximation for smooth and analytic functions [J], *Neural Computing*, 1996, 8:164–177.

[18] Mhaskar H. N. and Micchelli C. A., Degree of approximation by neural networks with a single hidden layer [J], *Advances in Applied Mathematics*, 1995, 16:151–183.

[19] Mhaskar H. N. and Khachikyan L., Neural networks for functions approximation [M]//Makhoul J., Manolakos E. and Wilson E., eds., Neural Networks for Signal Processing, New York: IEEE Press, 1995:21–29.

[20] Mhaskar H. N. and Micchelli C. A., Dimension-independent bounds on the degree of approximation by neural networks [J], *IBM J. Res. Develop.*, 1994, 38:277–284.

[21] Maiorov V. and Meir R. S., Approximation bounds for smooth functions in $C(\mathcal{R}^d)$ by neural and mixture networks [J], *IEEE Transactions on Neural Networks*, 1998, 9:969–978.

[22] Ritter G., Efficient estimation of neural weights by polynomial approximation [J], *IEEE Transactions on Information Theory*, 1999, 45:1541–1550.

[23] Suzuki S., Constructive function approximation by three-layer artificial neural networks [J], *Neural Networks*, 1998, 11:1049–1058.

[24] 谢庭藩，曹飞龙，关于插值神经网络的构造性 [J], 自然科学进展, 2008, 18(3):334–340.

[25] Xu Z. B. and Cao F. L., The essential order of approximation for neural networks [J], *Science in China, Series F*, 2004, 47(2):97–112.

[26] Xu Z. B. and Wang J. J., The essential order of approximation for nearly exponential type neural networks [J], *Science in China, Series F*, 2006, 49(4):446–460.

[27] 曹飞龙，张学东， d 维 Bernstein 算子加 Jacobi 权的收敛阶 [J], 计算数学, 2001, 23(4):407–416.

[28] 王建军，若干线性算子的加权逼近 [D], 银川：宁夏大学， 2003.

The Essential Order of Approximation with Weights of Neural Networks

WANG Jianjun* XU Zongben**

*School of Mathematics and Statistics, Southwest University, Chongqing 400715, China; Institute for Information and System Sciences, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: wangjianjun@mail.xjtu.edu.cn

**Institute for Information and System Sciences, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: zbxu@mail.xjtu.edu.cn

Abstract This paper presents the approximation ability of a feedforward neural network with a single hidden layer in L^q_w, including the estimation of its approximation upper and lower bounds. Under the principle of the weighted approximation, the work shows the relationship between the approximation precision of an underlying feedforward neural network and the number of hidden nodes. The crucial point provides a theoretical foundation for the applications of feedforward neural networks.

Keywords Approximation estimation, Neural networks, Jacobi weights

2000 MR Subject Classification 41A36, 41A25