中国科学 F辑: 信息科学 2009 年 第 39 卷 第 8 期: 809～814
www.scichina.com info.scichina.com
《中国科学》杂志社
SCIENCE IN CHINA PRESS

神经网络逼近速度下界估计

曹飞龙①*, 张永全②, 徐宗本②

① 中国计量学院信息与数学科学系, 杭州 310018
② 西安交通大学信息与系统科学研究所, 西安 710049
*通信作者 E-mail: flcao@263.net

收稿日期: 2008-05-30; 接受日期: 2009-03-07
国家重点基础研究发展计划 (批准号: 2007CB311002) 和国家自然科学基金 (批准号: 60575045, 60473034) 资助项目

摘要 文中研究一类多元周期 Lebesgue 平方可积函数 SF_d 与神经网络集合 $\Pi_{\varphi,n,d}=\{\sum_{j=1}^{n}b_j\varphi(\omega_j\cdot x+\theta_j):b_j,\theta_j\in R,\omega_j\in\mathbb{R}^d\}$ 之间偏差 $\mathrm{dist}(SF_d,\Pi_{\varphi,n,d})$ 的估计问题. 特别地, 利用 Fourier 变换、逼近论等方法给出 $\mathrm{dist}(SF_d,\Pi_{\varphi,n,d})$ 的下界估计, 即 $\mathrm{dist}(SF_d,\prod_{\varphi,n,d})\geqslant\frac{C}{(n\log_2 n)^{1/2}}$. 所获下界估计仅与神经网络隐层的神经元数目有关, 与目标函数及输入的维数无关. 该估计也进一步揭示了神经网络逼近速度与其隐层拓扑结构之间的关系.

关键词
神经网络
逼近速度
隐层拓扑结构
下界

0 引言

熟知, d 个输入、一个输出的三层前向神经网络数学上可以表示为

$$L_n(X)=\sum_{i=1}^{n}b_i\varphi(w_i\cdot X+\theta_i),\qquad(1)$$

其中输入向量 $X=(x_1,x_2,\ldots,x_d)^{\mathrm{T}}\in\mathbb{R}^d$, 隐层节点个数为 n, $w_i=(\omega_{i1},\omega_{i2},\ldots,\omega_{id})$, 是输入层与隐层之间的连接权, $w_i\cdot X$ 表示 w_i 与 X 的内积, φ 是隐层神经元的激活函数 (传递函数), θ_i 是阈值, 而 b_i 是隐层与输出层之间的连接权. 用 $\Pi_{\varphi,n,d}$ 表示隐层具有 n 个神经元、激活函数为 φ 的形如 (1) 式的三层单隐层前向神经网络的全体, 即

$$\Pi_{\varphi,n,d}=\left\{L_n(X)=\sum_{i=1}^{n}b_i\varphi(w_i\cdot X+\theta_i):\ b_i,\theta_i\in\mathbb{R},w_i\in\mathbb{R}^d\right\}.$$

以 $C(A)$、$L^p(A)(1\leqslant p<\infty)$ 分别表示集 A 上定义的连续函数、p 次 Lebesgue 可积函数全体所构成的 Banach 空间, 并分别赋予通常的上确界范数 $||\cdot||_\infty$ 与 Lebesgue 范数 $||\cdot||_p$. 集合 A 与 B 之间的距离 $\mathrm{dist}(A,B)$ 通常定义为

$$\begin{aligned}\mathrm{dist}(A,B)&=\sup_{f\in A}\mathrm{dist}(f,B)\\&=\sup_{f\in A}\inf_{g\in B}\|f-g\|_p,1\leqslant p<\infty.\end{aligned}$$

一个函数 $\varphi:\mathbb{R}^1\to\mathbb{R}^1$ 被称为 sigmoidal 函数, 如果满足

$$\lim_{x\to-\infty}\varphi(x)=0\ \text{且}\ \lim_{x\to\infty}\varphi(x)=1.$$

目前, 前向神经网络已被广泛应用于各个领域, 其主要原因之一是前向神经网络具有万有逼近性质, 即对于定义在紧集上的任何连续或可积函数, 存在一前向神经网络使它能以任意精度逼近该函数. 前向神经网络逼近的本质是以非线性参数展开来表达与近似一元或多元函数, 与之相关的基本问题有稠密性、复杂性和算法等. 对于稠密性问题, 早在 20 世纪 80 年代末与 90 年代初, 许多学者进行了深入、系统的研究 (典型的结果可参见文献 [1～10]). 1989 年, Cybenko[1] 首次证明了如下的稠密性定理: 如果 φ 是一个连续的 sigmoidal 函数, $I=[0,1]^d$ 是 $\mathbb{R}^d$ 中的 d 维单位立方体, 则 $\Pi_{\varphi,n,d}$ 在 $C(I)$ 中是稠密的. 此

引用格式: 曹飞龙, 张永全, 徐宗本. 神经网络逼近速度下界估计. 中国科学 F 辑: 信息科学, 2009, 39(8): 809–814

后, 关于该稠密性定理的一些不同证明方法被设计出来, 同时, 一些学者对为了使稠密性定理成立的激活函数 φ 所需的各种条件进行了一系列研究 (参见文献 [2~10]). 例如, 1991 年, Hornik 在文献 [7] 中证明了当 φ 是一个非常数的有界连续函数时, 稠密性定理也成立; Mhaskar 和 Micchelli[5] 证明只要对连续函数在无穷远处的幅度做适当的限制, 那么任意非多项式函数都可以充当激活函数; 陈天平等人对这一问题做了深入细致的研究 (参见文献 [8~10]), 得到了具有一般意义的结果: 稠密性定理成立的充分必要条件是激活函数 φ 不是多项式函数, 该结果为激活函数在更大范围内的选取提供了理论依据.

神经网络逼近复杂性主要指人工神经网络拓扑结构 (如隐层神经元个数、权值的大小) 与网络逼近能力、逼近速度之间的关系. 从应用角度看, 人们更加关心该问题. 近年来, 该问题引起了国内外学者的广泛关注 (如文献 [11~30]), 人们从不同角度揭示由不同种类激活函数构成的三层前向网络隐层拓扑结构与逼近速度之间的关系, 较多的工作研究了为达到规定的逼近精度所需的隐层神经元最少个数问题. 然而, 我们注意到, 已有的有关神经网络逼近复杂性研究基本上限于对网络逼近速度的上界估计. 很明显, 仅有上界估计对于刻画神经网络的逼近能力与隐层拓扑结构是远远不够的, 因为我们无法判断所给出的上界是否最小. 因此, 为了准确地反映所构造网络的逼近性能, 我们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. 自然地, 这样的下界估计是很困难的, 但却是非常重要而有意义的.

在文献 [25] 中, 我们对一类激活函数, 利用著名的 Bernstein 多项式的波峰展开与张量积方法构造出有明确隐层单元下界的多输入单隐层神经网络, 并以此为工具逼近多元连续或可积函数, 以构造网络逼近一元多项式作为过渡, 用广义连续模为度量尺度估计了逼近速度的上界与下界, 提出了本质的逼近阶的概念, 并刻画了所构造网络的本质逼近阶. 特别地, 证明了如下的逼近速度的下界估计: 假定激活函数 φ 是定义在实数域 $\mathbb{R}$ 上具有 $n+1$ 阶连续有界导数的一元函数, 并且满足: 存在 $\theta \in \mathbb{R}$ 使得 $\varphi^{(k)}(\theta) \neq 0$, $1 \leqslant k \leqslant n+1$, 则对任何 $f \in L^p(K)(K$ 是 $\mathbb{R}^d$ 中的某一紧集), $1 \leqslant p \leqslant \infty$, 可以构造出以 φ 为激活函数, d 个输入, 一个输出, 隐层神经元个数为 $m = N_0 + 2N_1 + \cdots + (n+1)N_n$ 的三层前向神经网络 $L_n(X)$ 使得

$$\omega_\varphi^2\left(f, \frac{1}{\sqrt{n}}\right)_p \leqslant \frac{C}{n}\sum_{i=1}^{n}\|L_i - f\|_p. \tag{2}$$

这里及以下 C 表示与 f、n、X 无关的正常数 (在不同处其值可以不同), 而 $N_i = \binom{i+d-1}{d-1}$ $(0 \leqslant i \leqslant n)$, $\omega_\varphi^2(f,t)_p$ 是函数 f 的二阶加权光滑模 [25,31].

文献 [30] 将文献 [25] 中的结果推广到激活函数为指数型神经网络 R_n^σ(见文献 [20]), 对于 $f \in C(K)$, 给出了逼近速度的一类加权平均逼近下界的估计 (见文献 [30] 之定理 1):

$$\omega_2\left(f, \frac{1}{n+2}\right)_\infty \leqslant \frac{C}{n^2}\left(\sum_{k=1}^{n} k||f - R_k^\sigma(d)||_\infty + ||f||_\infty\right), \tag{3}$$

这里指数型神经网络 R_n^σ 的隐层神经元个数不少于 $\min_{B_d(f,n)<\varepsilon}(n+1)^d$, 其中

$$B_d(f,n) = \frac{1}{2}\left(1 + \frac{\pi^2}{2}\sqrt{d}\right)^2 \omega_2\left(f, \frac{1}{n+2}\right)_\infty,$$

$\omega_2(f,t)_\infty$ 为函数 f 的二阶连续模 [31], ε 为给定的逼近精度.

在文献 [26] 中, 作者采用不同于文献 [25] 的方法, 讨论单隐层神经网络逼近速度与最佳多项式逼近之间的关系, 以函数最佳多项式逼近作为度量尺度, 估计单隐层神经网络逼近连续函数的速度. 在文献 [29] 中, 作者还考虑了同时联立逼近的复杂性问题, 利用光滑模估计网络联立逼近的速度.

上述提及一系列网络逼近复杂性的研究 (特别是网络逼近速度的下界估计是一项关键而颇具难度的工作) 对刻画网络的本质逼近阶与隐层拓扑结构的研究起到了十分重要的作用. 然而, 我们发现已有工作存在如下不足: (1) 正如文献 [25] 中的评注所指出: 所构造网络的逼近能力大多依赖于目标函数的光滑性, 并局限于所采用构造工具逼近能力的饱和性; (2) 所构造的网络的隐层神经元数目往往偏多, 从而导致与逼近速度不相协调或神经元数目难以计算等问题; (3) 刻画逼近误差的尺度大多是函数的连续模 (光滑模) 或最佳多项式逼近, 尽管这些尺度是逼近论中常

用的刻画函数光滑性与逼近误差的传统工具, 但它们不具有直接刻画 "阶" 的优点以及存在不易计算的缺点; (4) 较多研究均是针对某一特殊激活函数类进行或对激活函数的条件有较多的限制; (5) 已有下界估计均是针对逼近速度的某种平均而言, 并且逼近速度往往与维数有关. 因此, 针对上述的诸多不足, 一个重要而且迫需解决的问题是: 对广泛的激活函数类以及给定的某一目标函数类 F, 如何估计 $\Pi_{\varphi,n,d}$ 与 F 之间的偏差, 即如何利用隐层神经元数目 n 直接估计 $(F,\Pi_{\varphi,n,d})$ 的上界与下界. 本文致力于该问题的研究. 特别地, 我们毋须对激活函数太多的要求 (如 sigmodial 型、可导、指数型等), 利用 Fourier 变换、逼近论等某些方法, 给出 $\mathrm{dist}(F,\Pi_{\varphi,n,d})$ 的一个简洁下界估计.

1 主要结果

令 $d\geqslant 1$ 为一个固定的整数, 记 $Q^d=[-\pi,\pi]^d$, 用 $L^2(Q^d)$ 表示由 $\mathbb{R}^d\to\mathbb{R}$ 之 Lebesgue 平方可积、每个变量均以 2π 为周期的函数 f 所组成的 Banach 空间, 其范数 $||f||_2$ 满足

$$||f||_2=\left\{\left(\frac{1}{2\pi}\right)^d\int_{Q^d}|f(t)|^2\mathrm{d}t\right\}^{\frac{1}{2}}<+\infty.$$

对于 $f\in L^2(Q^d)$, 定义 f 的 Fourier 变换

$$\hat{f}(k)=\left(\frac{1}{2\pi}\right)^d\int_{Q^d}f(t)\mathrm{e}^{-\mathrm{i}k\cdot t}\mathrm{d}t,\quad k\in\mathbb{Z}^d.$$

并令

$$||f||_{SF,d}=\sum_{k\in\mathbb{Z}^d}|\hat{f}(k)|,$$

定义函数类

$$SF_d=\{f\in L^2(Q^d):||f||_{SF,d}\leqslant 1\}.$$

Mhaskar 在文献 [12] 中研究了神经网络逼近 SF_d 中函数速度的上界, 获得如下结果: 假定 $\varphi\in L^2(Q^1)$, 且 $\hat{\varphi}(1)\neq 0$, 则对任意 $f\in SF_d$, 成立

$$\inf_{g\in\Pi_{\varphi,n,d}}||f-g||\leqslant\frac{\delta_n}{\sqrt{n+1}}+2\frac{E_n(\varphi)}{|\hat{\varphi}(1)|}.$$

其中数列 $\{\delta_n\}$ 满足 $\lim_{n\to\infty}\delta_n=0, 0<\delta_n\leqslant\sqrt{2}$, $E_n(\varphi)=\inf_{p\in\pi_n}||\varphi-p||_2$ 是函数 φ 的最佳多项式逼近, π_n 是阶数不超过 n 的三角多项式集合. 进一步, 文献 [12] 获得:

$$\mathrm{dist}(SF_d,\Pi_{\varphi,n,d})\leqslant\frac{1}{\sqrt{n}}+\frac{E_n(\varphi)}{|\hat{\varphi}(1)|}.\tag{4}$$

显然, 上述的上界估计与自变量的维数、目标函数无关, 仅与隐层神经元数 n 及激活函数 φ 有关, 网络的激活函数 φ 也不必为 sigmodial 函数, 该上界估计在一定程度上反映了逼近速度与隐层拓扑结构之间的关系. 我们也注意到 Kåurková在文献 [16] 中, 利用泛函分析等方法也得到与文 [12] 相类似的逼近的结果. 自然要问: $\mathrm{dist}(F,\Pi_{\varphi,n,d})$ 下界如何估计? 我们至今尚未见到相关研究. 本文得到如下结果:

定理 1 假设 $\varphi\in L^2(Q^1)$, n 为任一自然数, 则

$$\mathrm{dist}\left(SF_d,\prod_{\varphi,n,d}\right)\geqslant\frac{C}{(n\log_2 n)^{1/2}},$$

其中 C 为某一正常数.

可以看出: (1) 定理 1 对激活函数 ϕ 的限制是非常宽松的, 仅要求周期平方可积即可, 去掉了上界估计中 $\hat{\varphi}\neq 0$ 的条件, 更不需通常的 sigmodial 与光滑性要求. 因此, 定理 1 具有比较广泛的适用性. (2) 定理 1 所给出的下界与被逼近的目标函数 f 及输入的维数 d 无关, 仅与隐层神经元数 n 有关, 毋须借助函数的连续模 (光滑模)、最佳逼近多项式等度量尺度, 其逼近速度的 "阶" 是显而易见的.

2 定理 1 证明

在证明定理 1 之前, 先介绍一些概念与记号. 令 m 为一自然数, 考虑由向量 $t=(t_1,t_2,\ldots,t_m)\in\mathbb{R}^m$ 且以范数 $||t||_{l_2^m}=(\sum_{j=1}^m t_j^2)^{\frac{1}{2}}$ 构成的 Hilbert 空间 l_2^m 设 $H\subset l_2^m$, 定义 t 到集合 H 的距离为

$$\mathrm{dist}(t,H)=\inf_{h\in H}||t-h||_{l_2^m}.$$

记集合 m

$$E^m=\{\varepsilon=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_m):\varepsilon_i=\pm 1,i=1,2,\ldots,m\}.$$

对于任一向量 $t=(t_1,t_2,\ldots,t_m)\in\mathbb{R}^m$, 符号向量 $\mathrm{sgn}\ t=(\mathrm{sgn}\ t_1,\mathrm{sgn}\ t_2,\ldots,\mathrm{sgn}\ t_m)$, 其中当 $\tau\geqslant 0$ 时, $\mathrm{sgn}\ \tau=1$; 当 $\tau<0$ 时, $\mathrm{sgn}\ \tau=0$. 对于 $T\subset\mathbb{R}^m$, 我们用 $\mathrm{sgn}T$ 表示符号向量集 $\{\mathrm{sgn}\ t:t\in T\}$.

现令 $u=(u_1,u_2,\ldots,u_n)\in\mathbb{R}^n$, 并令 $p_1(u)$, $p_2(u),\ldots,p_m(u)$ 为总次数不超过 n 的代数多项式,

定义如下的 m 维集合:

$$\Phi_{m,n,s}=\{(p_1(u),p_2(u),\ldots,p_m(u)):u\in\mathbb{R}^n\}.$$

我们还需如下引理:

引理 1[21] 设 m,n,s 为自然数, 且满足 $m\geqslant 2n$ 及 $n\log\frac{4ems}{n}\leqslant\frac{m}{4}$, 则存在一向量 $\varepsilon\in E^m$ 与某一绝对常数 $C>0$, 使得 $\mathrm{dist}(\varepsilon,\Phi_{m,n,s})\geqslant C\sqrt{m}$.

现在证明定理 1. 设 n,s 为自然数, 以 $m=|\{k\in\mathbb{Z}^d:|k|\leqslant s\}|$ 表示集合 $\{k\in\mathbb{Z}^d:|k|\leqslant s\}$ 的元素个数, 则存在常数 C',C'', 使得 $C's^d\leqslant m\leqslant C''s^d$. 考虑如下由三角多项式组成的集合:

$$F_{n,d}=\left\{\frac{1}{m}\sum_{|k|\leqslant s}\varepsilon_k\mathrm{e}^{\mathrm{i}k\cdot X}:\varepsilon_k=\pm1\right\},$$

则对任何三角多项式 $f\in F_{n,d}$, 有 $||f||_{SF_d}=\frac{1}{m}\sum_{|k|\leqslant s}1=1$. 因此 $f\in SF_d$, 并由 f 的任意性知 $F_{n,d}\subset SF_d$.

下面考虑集合 $F_{n,d}$ 与 $\Pi_{\varphi,n,d}$ 之间偏差 $\mathrm{dist}(F_{n,d},\Pi_{\varphi,n,d})$ 的下界. 设 $f(X)$ 为集合 $F_{n,d}$ 中的任意一个多项式函数, $h(X)=\sum_{j=1}^n a_j\varphi(w_j\cdot X+\theta_j)$ 为集合 $\Pi_{\varphi,n,d}$ 中任一元素, 即隐层具有 n 个神经元, 激活函数为 φ 的三隐层前向神经网络. 由于 $\varphi\in L^2_{[-\pi,\pi]}$, 我们可以计算 $h(X)$ 的 Fourier 变换

$$\hat{h}(k)=\left(\frac{1}{2\pi}\right)^d\sum_{j=1}^n a_j\int_{[-\pi,\pi]^d}\varphi(w_j\cdot X+\theta_j)\mathrm{e}^{-\mathrm{i}k\cdot X}\mathrm{d}X,$$

其中 $w_j=(w_{j_1},w_{j_2},\ldots,w_{j_d})\in\mathbb{R}^d$.

当 $w_j\neq0(1\leqslant j\leqslant n)$ 时, 则存在 $t_j=(t_{j_1},t_{j_2},\ldots,t_{j_d})\in\mathbb{R}^d$, 使得 $w_j\cdot t_j=-\theta_j$. 从而

$$\begin{aligned}\hat{h}(k)&=\left(\frac{1}{2\pi}\right)^d\sum_{j=1}^n a_j\int_{[-\pi,\pi]^d}\varphi(w_j(X-t_j))\cdot\mathrm{e}^{-\mathrm{i}k\cdot X}\mathrm{d}X\\&=\left(\frac{1}{2\pi}\right)^d\sum_{j=1}^n a_j\int_{[-\pi,\pi]^d}\varphi(w_jX)\cdot\mathrm{e}^{-\mathrm{i}k\cdot(X-t_j)}\mathrm{d}X\\&=\left(\frac{1}{2\pi}\right)^d\sum_{j=1}^n a_j\mathrm{e}^{\mathrm{i}k\cdot t_j}\int_{[-\pi,\pi]^d}\varphi(w_jX)\cdot\mathrm{e}^{-\mathrm{i}k\cdot X}\mathrm{d}X.\end{aligned}$$

令 $d_{k,j}=(\frac{1}{2\pi})^d\int_{[-\pi,\pi]^d}\phi(w_j\cdot X)\mathrm{e}^{-\mathrm{i}k\cdot X}\mathrm{d}X$, 得到

$$\hat{h}(k)=\sum_{j=1}^n d_{k,j}a_j\mathrm{e}^{\mathrm{i}k\cdot t_j}=\sum_{j=1}^n b_{k,j}\mathrm{e}^{\mathrm{i}k\cdot t_j},$$

其中 $b_{k,j}=a_jd_{k,j}$.

若存在 $j_0\in N,1\leqslant j_0\leqslant n$, 使得 $w_{j_0}=0$, 则 $\hat{h}(k)$ 中的第 j_0 加项为

$$\begin{aligned}&\left(\frac{1}{2\pi}\right)^d a_{j_0}\int_{[-\pi,\pi]^d}\varphi(\theta_{j_0})\mathrm{e}^{-\mathrm{i}k\cdot X}\mathrm{d}X\\&=\left(\frac{1}{2\pi}\right)^d a_{j_0}\varphi(\theta_{j_0})\int_{[-\pi,\pi]^d}\mathrm{e}^{-\mathrm{i}k\cdot X}\mathrm{d}X\\&=a_{j_0}\varphi(\theta_{j_0})\mathrm{e}^{\mathrm{i}k\cdot0}=b_{k,j_0}\mathrm{e}^{\mathrm{i}k\cdot t_{j_0}},\end{aligned}$$

其中 $b_{k,j_0}=a_{j_0}\phi(\theta_{j_0}),t_{j_0}=0$. 由此, 利用 Fourier 展开及正交性质, 我们得到

$$\begin{aligned}||f-h||_2^2&=\int_{[-\pi,\pi]^d}\left|\sum_{k\in\mathbb{Z}^d}(\hat{f}(k)-\hat{h}(k))\mathrm{e}^{\mathrm{i}k\cdot X}\right|^2\mathrm{d}X\\&=\sum_{k\in\mathbb{Z}^d}\left|\hat{f}(k)-\hat{h}(k)\right|^2\\&\geqslant\sum_{|k|\leqslant s}\left|\hat{f}(k)-\hat{h}(k)\right|^2\\&=\sum_{|k|\leqslant s}\left|\hat{f}(k)-\sum_{j=1}^n b_{k,j}\mathrm{e}^{\mathrm{i}k\cdot t_j}\right|^2.\end{aligned}$$

由于 $\hat{f}(k)=\frac{\varepsilon_k}{m}$, 则有

$$\begin{aligned}||f-h||_2^2&\geqslant\sum_{|k|\leqslant s}\left|\frac{\varepsilon_k}{m}-\sum_{j=1}^n b_{k,j}\mathrm{e}^{\mathrm{i}k\cdot t_j}\right|^2\\&=\frac{1}{m^2}\sum_{|k|\leqslant s}\left|\varepsilon_k-m\sum_{j=1}^n b_{k,j}\mathrm{e}^{-\mathrm{i}k\cdot t_j}\right|^2.\end{aligned}$$

因为 $b_{k,j}$ 是复数, 故令 $b_{k,j}=b'_j+ib''_j$, 其中 b'_j,b''_j 为实数, 所以

$$\begin{aligned}&m\sum_{j=1}^n b_{k,j}\mathrm{e}^{\mathrm{i}k\cdot t_j}\\&=m\sum_{j=1}^n(b'_j+\mathrm{i}b''_j)(\cos k\cdot t_j+\mathrm{i}\sin k\cdot t_j)\\&=m\sum_{j=1}^m(b'_j\cos k\cdot t_j-b''_j\sin k\cdot t_j)\\&\quad+m\mathrm{i}\sum_{j=1}^m(b''_j\cos k\cdot t_j+b'_j\sin k\cdot t_j).\end{aligned}$$

又因为 $\varepsilon_k=\pm1$, 所以

$$\begin{aligned}&||f-h||_2^2\\&\geqslant\frac{1}{m^2}\sum_{|k|\leqslant s}\left|\varepsilon_k-m\sum_{j=1}^m(b'_j\cos k\cdot t_j-b''_j\sin k\cdot t_j)\right|^2\end{aligned}$$

$$+\frac{1}{m^2}\sum_{|k|\leqslant s} m^2\left|\sum_{j=1}^{m} b''_j\cos k\cdot t_j + b'_j\sin k\cdot t_j\right|^2$$

$$\geqslant \frac{1}{m^2}\sum_{|k|\leqslant s}\left|\varepsilon_k - m\sum_{j=1}^{m}(b'_j\cos k\cdot t_j - b''_j\sin k\cdot t_j)\right|^2.$$

令实变量 $b=(b_1,b_2,\ldots,b_{2n})\in\mathbb{R}^{2n}, X=(x_{j_1},\ldots,x_{j_q}), j=1,2,\ldots,2n$, 得出

$$P_k(u)=P_k(b,X)=m\sum_{j=1}^{2n} b_j x_{j_1}^{k_1}x_{j_2}^{k_2}\cdots x_{j_d}^{k_d},\quad |k|\leqslant s,$$

其中当 $1\leqslant j\leqslant n$ 时, $b_j=b'_j$, 当 $n+1\leqslant j\leqslant 2n$ 时, $b_j=b''_j$, $u=(b,X)\in\mathbb{R}^{2n+2dn}$, 则每个多项式 $P_k(u)$ 的总次数不超过 s. 现令 $m=C_2 n\log_2 n$, 其中 $C_2=(4eC')^{-1/2}$ 并且 $n\geqslant 10$, 那么 $n\log_2\frac{4ems}{n}\leqslant\frac{m}{4}$. 故由引理 1 知: 存在 $\varepsilon\in E^m$, 使得

$$\inf_{b_{k,j}\in\mathbb{R},t_j\in\mathbb{R}^d}\sum_{|k|\leqslant s}\left|\varepsilon_k - m\sum_{j=1}^{n} b_{k,j}\mathrm{e}^{\mathrm{i}k\cdot t_j}\right|^2$$

$$\geqslant \inf_{b'_j,b''_j\in\mathbb{R};t_j\in\mathbb{R}^d}\sum_{|k|\leqslant s}\left|\varepsilon_k - m\sum_{j=1}^{n}(b'_j\cos k\cdot t_j - b''_j\sin k\cdot t_j)\right|^2$$

$$\geqslant \inf_{u\in\mathbb{R}^{2n+2dn}}\sum_{|k|\leqslant s}|\varepsilon_k - p_k(u)|^2\geqslant C'm.$$

综合上述不等式, 得到 $||f-h||_2^2\geqslant\frac{1}{m^2}C'm=\frac{C'}{m}$, 所以, $||f-h||_2\geqslant\frac{C}{\sqrt{m}}=\frac{C}{\sqrt{n\log_2 n}}$, 即

$$\mathrm{dist}(SF_d,\Pi_{\varphi,n,d})\geqslant\frac{C}{(n\log_2 n)^{1/2}}.$$

定理 1 证毕.

3 评注

本文给出了三层前向神经网络集合与目标函数集之间距离的下界估计. 其主要意义如下: (1) 所获得估计与激活函数及维数无关, 该下界估计可由神经元数 n 直接刻画, 并且该估计简洁、明了, 易于计算, 反映了速度下界与隐层神经元之间的关系; (2) 目标函数集、神经网络集均比较广泛. 构成神经网络的激活函数仅要求 Lebesgue 平方可积, 而目标函数仅是对其 Fourier 变换作一简单要求 (即 Fourier 变换之和有界). 因此, 本文结果具有广泛性.

此外, 我们由上界估计 (4) 知道: 当激活函数满足一定光滑性条件时, 由逼近论的 Jackson 定理 (见文献 [32]) 知道: $E_n(\varphi)\leqslant\frac{C}{\sqrt{n}}$, 由此得到

$$\frac{C_1}{(n\log_2 n)^{1/2}}\leqslant\mathrm{dist}(SF_d,\Pi_{\varphi,n,d})\leqslant\frac{C_2}{n^{1/2}}.\quad(5)$$

自然地, 我们希望上下界的 "阶" 是一致的, 即能够准确刻画 $\mathrm{dist}(SF_d,\Pi_{\varphi,n,d})$ 的本质逼近阶, 这是一个富有挑战、值得进一步研究的问题. 我们猜测: 在一定条件下 (5) 式中的上界估计可改进为 $\frac{C}{(n\log_2 n)^{1/2}}$.

参考文献

1 Cybenko G. Approximation by superpositions of a single function. Math Contr Signal, 1989, 2: 303–314

2 Funahashi K. On the approximate realization of continuous mappings by neural networks. Neur Netw, 1989, 2: 183–192

3 Hornik K, Stinchombe M, White H. Multilayer feedforward networks are universal approximators. Neur Netw, 1989, 2: 359–366

4 Leshno M, Lin V Y, Pinkus A, et al. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neur Netw, 1993, 6: 861–867

5 Mhaskar H N, Micchelli C A. Approximation by superposition of sigmoidal and radial basis functions. Adv Appl Math, 1992, 13: 350–373

6 Yoshifusa I. Approximation of functions on a compact set by finite sums of sigmoid function without scaling. Neur Netw, 1991, 4: 817–826

7 Hornik H. Approximation capabilities of multilayer feedforward networks. Neur Netw, 1991, 4: 251–257

8 Chen T P, Chen H. Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. IEEE Trans Neur Netw, 1995, 6: 904–910

9 Chen T P, Chen H. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. IEEE Trans Neur Netw, 1995, 6: 911–917

10 陈天平. 神经网络及其在系统识别应用中的逼近问题. 中国科学 A 辑, 1994, 37(1): 1–7

11 Mhaskar H N, Micchelli C A. Degree of approximation by neural and translation networks with a single hidden layer. Adv Appl Math, 1995, 16: 151–183
12 Mhaskar H N, Micchelli C A. Dimension independent bounds on the degree of approximation by neural networks. IBM J Res Dev, 1994, 38(3): 277–284
13 Chui C K, Li X. Approximation by ridge functions and neural networks with one hidden layer. J Approx Theory, 1992, 70: 131–141
14 Maiorov V, Meir R S. Approximation bounds for smooth functions in $C(R^d)$ by neural and mixture networks. IEEE Trans Neur Netw, 1998, 9: 969–978
15 Kåurkova V, Kainen P C, Kreinovich V. Estimates of the number of hidden units and variation with respect to half-space. Neur Netw, 1997, 10: 1068–1078
16 Kåurková V, Sanguineti M. Bounds on rates of variable-basis and neural network approximation. IEEE Trans Inform Theory, 2001, 47(6): 2659–2665
17 Li X. Simultaneous approximation of multivariate function and their derivatives by neural networks with one hidden layer. Neurocomputing, 1996, 12: 327–343
18 Chen X H, White H. Improve rates and asymptotic normality for nonparametric neural network estimators. IEEE Trans Inform Theory, 1999, 45: 682–691
19 Suzuki S. Constructive function approximation by three-layer artificial neural networks. Neur Netw, 1998, 11: 1049–1058
20 Ritter G. Efficient estimation of neural weights by polynomial approximation. IEEE Trans Inform Theory, 1999, 45: 1541–1550
21 Maiorov V E. On best approximation by ridge function. J Approx Theory, 1999, 99: 68–94
22 Kåurkova V, Sanguineti M. Comparison of worst case errors in linear and neural networks approximation. IEEE Trans Inform Theory, 2002, 48: 264–275
23 Lavretsky E. On the geometric convergence of neural approximations. IEEE Trans Neur Netw, 2002, 13: 274–282
24 Barron A. Universal approximation bounds for superposition of a sigmoidal function. IEEE Trans Inform Theory, 1993, 39: 930–945
25 Xu Z B, Cao F L. The essential order of approximation for neural networks. Sci China Ser F-Inf Sci, 2004, 47(1): 97–112
26 曹飞龙, 张永全, 张卫国. 单隐层神经网络与最佳多项式逼近. 数学学报, 2007, 50(2): 385–392
27 曹飞龙, 张永全. 距离空间中的神经网络插值与逼近. 数学学报, 2008, 51(1): 91–98
28 Cao F L, Xie T F, Xu Z B. The estimate of approximation error for neural networks: a constructive approach. Neurocomputing, 2008, 71: 626–630
29 Xu Z B, Cao F L. Simultaneous L^p-approximation order for neural networks. Neur Netw, 2005, 18: 914–923
30 Xu Z B, Wang J J. The essential order of approximation for nearly exponential type neural networks. Sci China Ser F-Inf Sci, 2006, 49(4): 446–460
31 Ditzian Z, Totik V. Moduli of smoothness. Berlin/New York: Springer-Verlag, 1987. 1–35
32 谢庭藩, 周颂平. 实函数逼近论. 杭州: 杭州大学出版社, 1998. 63–71