【文章编号】 1004-1540(2009)01-0071-05

神经网络对 $C(\overline{\mathbf{R}})$ 中函数的偏差估计

舒 衡,谢庭藩
(中国计量学院 理学院,浙江 杭州 310018)

【摘 要】 利用构造性的方法研究了以 Sigmoidal 函数为激活函数的单隐层前向人工神经网络对 $C(\overline{\mathbf{R}})$ 中函数的偏差估计.

【关键词】 逼近;神经网络;Sigmoidal 函数;连续性模
【中图分类号】 O174.41 【文献标识码】 A

Degree of approximation to a function in $C(\overline{\mathbf{R}})$ by neural network

SHU Heng, XIE Ting-fan
(College of Sciences, China Jiliang University, Hangzhou 310018, China)

Abstract: The degree of approximation to a function in $C(\overline{\mathbf{R}})$ by single layer FANN with sigmoidal function is investigated by the constructive approach.

Key words: approximation; neural network; sigmoidal function; modulus of continuity

在最近这些年里,很多学者研究了用神经网络进行逼近的问题,其中大量的文献都是关于神经网络在连续函数空间中的稠密性的,这些学者包括 Cybenko[1], Funahashi[2], Li[3], Hornik[4], Leshno 等[5], Chen 和 Chen 等[6] 以及 Cui 和 Li[7] 等等.

与稠密性密切相关的问题是神经网络的复杂性问题,也就是刻画逼近偏差与隐层神经元个数以及被逼近函数的连续性或光滑性之间的问题. Chen Debao[8] 给出了以 Sigmoidal 函数为激活函数的单隐层前向人工神经网络对 $C([a,b])$ 中函数逼近的逼近偏差上界估计. Cao 等[9] 描述了以有界的严格单调的奇函数作激活函数的单隐层前向人工神经网络对 $C([a,b])$ 中函数进行逼近的上界估计. 一些学者(Jones[10], Barron[11], Murata[12])研究了以 Sigmoidal 函数作激活函数的神经网络逼近几类充分光滑的函数的问题,得到了以 $O\left(\frac{1}{n}\right)$ 为逼近误差的结果.

1 准备工作

本文的证明中使用如下一些记号. $\mathbf{N}, \mathbf{N}_0, \mathbf{R}$ 分别表示正数集,非负整数集和实数集.

$\Psi_{n,\sigma} := \left\{ a_0 + \sum_{i=1}^{n} a_i \sigma(b_i x + c_i) : a_i, b_i, c_i \in \mathbf{R} \right\}$ 表示以 σ 为激活函数的单隐层前向人工神经网络.

【收稿日期】 2008-11-15
【作者简介】 舒 衡(1982-),男,四川都江堰人,硕士研究生,主要研究方向为人工神经网络.

$\sigma(x)$ 为 Sigmoidal 函数，也即 $\lim\limits_{x\to-\infty}\sigma(x)=0$，$\lim\limits_{x\to+\infty}\sigma(x)=1$. 作为特殊情况若

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

则称 $\sigma(x)$ 为 Heaviside 函数.

$$C(\overline{\mathbf{R}}):=\left\{\begin{array}{l}f:\mathbf{R}\to\mathbf{R}\mid f\in C(\mathbf{R})\text{ 且}\\ \lim\limits_{x\to-\infty}f(x),\lim\limits_{x\to+\infty}f(x)\text{ 存在}\end{array}\right\}$$

表示闭实轴上的连续函数空间.

对于 $f\in C(\overline{\mathbf{R}})$，定义函数 $\Phi(A)$ 如下：

$$\Phi(A):=\max\left\{\begin{array}{l}\sup\limits_{x_1,x_2\geqslant A}|f(x_1)-f(x_2)|,\\ \sup\limits_{x_1,x_2\leqslant -A}|f(x_1)-f(x_2)|\end{array}\right\}$$

用于刻画函数在无穷远处的连续程度，易知 $\lim\limits_{A\to+\infty}\Phi(A)=0$. 如果 $f\in C([a,b])$，其连续性模定义如下：

$$\omega(f,\delta)_{[a,b]}:=\sup_{x,x+h\in[a,b],|h|\leqslant\delta}|f(x+h)-f(x)|.$$

如果存在常数 $M>0$ 使得 $\omega(f,\delta)\leqslant M\delta^{\alpha}$，则称 f 是 Lipschitz α 连续的 $(0<\alpha\leqslant 1)$，记作 $f\in LiP_M\alpha$. 我们也规定，若 $M=1$，则可省去不写. 我们用 $f(x)=O(g(x))$ 表示存在一个常数 $M>0$ 使得 $|f(x)|\leqslant Mg(x)$.

B. I. Hong 和 N. Hahm[13] 研究了用 Sigmoidal 函数作为激活函数的神经网络对 $C(\overline{\mathbf{R}})$ 中函数的逼近度，该文公布的定理如下：

定理 H－H 如果 σ 是 $\mathbf{R}$ 上有界的 Sigmoidal 函数，$f\in C(\overline{\mathbf{R}})$ 即 $f\in C(\mathbf{R})$ 且 $\lim\limits_{x\to+\infty}f(x)=a$，$\lim\limits_{x\to-\infty}f(x)=b$，$a,b\in\mathbf{R}$，那么对于每个正整数 n，都有

$$E^{\Psi}_{n,\sigma}(f)\leqslant\|\sigma\|\,\omega\left(f,\frac{2A_1}{n+1}\right).$$

此处 $\|\sigma\|=\sup\limits_{x\in\mathbf{R}}|\sigma(x)|$，$\omega(f,\cdot)$ 是 f 在 $\mathbf{R}$ 上的连续性模，常数 A_1 依赖于 f 和 σ，其中 $E^{\Psi}_{n,\sigma,\mathbf{R}}(f):=E^{\Psi}_{n,\sigma}(f)=\inf\limits_{g\in\Psi_{n,\sigma}}\sup\limits_{x\in\mathbf{R}}|f(x)-g(x)|$. 该文的证明中是这样取常数 A_1，存在一个常数 $A_1\geqslant 1$，使得对于任意 $x\geqslant A_1$，有 $|f(x)-a|<\frac{\varepsilon}{2}$，$|\sigma(x)-1|<\gamma$. 可见常数 A_1 是依赖于 ε 的，因此不能根据 ε 的任意性，由

$$E^{\Psi}_{n,\sigma}(f)\leqslant\sup_{x\in\mathbf{R}}|f(x)-g(x)|\leqslant$$
$$\|\sigma\|\,\omega\left(f,\frac{2A_1}{n+1}\right)+2\varepsilon$$

推出

$$E^{\Psi}_{n,\sigma}(f)\leqslant\|\sigma\|\,\omega\left(f,\frac{2A_1}{n+1}\right).$$

实际上定理 H－H 是不成立的. 反例如下：今取

$$f(x)=\begin{cases}\dfrac{1}{1+x} & x>0,\\ 1 & x\leqslant 0.\end{cases}$$

显然 $f(x)\in C(\overline{\mathbf{R}})$ 且 $a=0$. 为使 $|f(x)-a|<\varepsilon$，须在 $x>A_1$ 时有 $1+x>\frac{1}{\varepsilon}$ 或者说 $x>\frac{1}{\varepsilon}-1$. 于是必须取 $A_1\geqslant\frac{1}{\varepsilon}-1$，从而有

$$E^{\Psi}_{n,\sigma}(f)\leqslant\|\sigma\|\,\omega\left(f,\frac{2(1-\varepsilon)}{(n+1)\varepsilon}\right)+2\varepsilon.$$

当 $\varepsilon\to 0$ 时，对于给定的 n，$\frac{2(1-\varepsilon)}{(n+1)\varepsilon}\to+\infty$，$\omega\left(f,\frac{2(1-\varepsilon)}{(n+1)\varepsilon}\right)$ 不收敛于 0. 因此定理 H－H 失去意义.

本文将对该逼近度另行刻画，我们的证明是构造性的.

2 主要结论及其证明

定理 2.1 如果 σ 是 $\mathbf{R}$ 上有界的 Sigmoidal 函数，$f\in C(\overline{\mathbf{R}})$，$n\in\mathbf{N}$，那么对于任意的 $\varepsilon>0$，$A>0$，存在单隐层前向人工神经网络

$$\Psi_{n,\sigma}:=\left\{a_0+\sum_{i=1}^{n}a_i\sigma(b_ix+c_i):a_i,b_i,c_i\in\mathbf{R}\right\},$$

使得

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x)-f(x)|\leqslant$$
$$\|\sigma\|\,\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]}+\Phi(A)+\varepsilon.$$

此处 $\|\sigma\|=\sup\limits_{x\in(-\infty,+\infty)}|\sigma(x)|$.

证明 对于任意的 $A>0$，将 $(-\infty,+\infty)$ 分为三部分 $(-\infty,-A)$，$[-A,A]$，$(A,+\infty)$. 固定 n，取 $x_k=-A+\frac{2A}{n+1}k$，$k=0,1,2,\cdots,n+1$. 对于任意的 $\varepsilon>0$，取

$$C=\frac{\varepsilon}{n\left(\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]}+1\right)}.$$

由 Sigmoidal 函数的定义，存在 $B>0$ 使得

$$x\geqslant B\text{ 时，}|\sigma(x)-1|<C,$$

$x \leqslant -B$ 时，$|\sigma(x)| < C$.

取 $\delta \in \left(0, \frac{A}{2(n+1)}\right)$，构造单隐层前向人工神经网络

$$\Psi_{n,\sigma}(x) = f(x_1) + \sum_{i=1}^{n}(f(x_{i+1}) - f(x_i))\sigma(\delta^{-1}B(x - x_i)).$$

如果 $x \in [x_k - \delta, x_k + \delta]$，此处 $k = 1,2,\cdots,n+1$（对于 $k = n+1$，$x \in [x_{n+1} - \delta, x_{n+1}]$），那么

$i < k$ 时，$x - x_i \geqslant \delta$，

$i > k$ 时，$x - x_i \leqslant -\delta$.

于是

$$\Psi_{n,\sigma}(x) - f(x) = f(x_k) - f(x) + \sum_{i=1}^{k-1}(f(x_{i+1}) - f(x_i))(\sigma(\delta^{-1}B(x - x_i)) - 1) + (f(x_{k+1}) - f(x_k))\sigma(x - x_k) + \sum_{i=k+1}^{n}(f(x_{i+1}) - f(x_i))\sigma(\delta^{-1}B(x - x_i)).$$

因此

$$|\Psi_{n,\sigma}(x) - f(x)| \leqslant \omega(f,\delta)_{[-A,A]} + (k-1)\omega\left(f, \frac{2A}{n+1}\right)_{[-A,A]}C + \|\sigma\|\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]} + (n-k)\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]}C \leqslant \omega(f,\delta)_{[-A,A]} + \|\sigma\|\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]} + \varepsilon.$$

如果 $x \in [x_k + \delta, x_{k+1} - \delta]$，此处 $k = 0,1,\cdots,n$（对于 $k = 0$，$x \in [x_0, x_1 - \delta]$），那么

$i \leqslant k$ 时，$x - x_i \geqslant \delta$，

$i > k$ 时，$x - x_i \leqslant -\delta$.

于是

$$\Psi_{n,\sigma}(x) - f(x) = f(x_{k+1}) - f(x) + \sum_{i=1}^{k}(f(x_{i+1}) - f(x_i))(\sigma(\delta^{-1}B(x - x_i)) - 1) + \sum_{i=k+1}^{n}(f(x_{i+1}) - f(x_i))\sigma(\delta^{-1}B(x - x_i)).$$

因此

$$|\Psi_{n,\sigma}(x) - f(x)| \leqslant \omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]} + n\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]}C \leqslant \omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]} + \varepsilon.$$

注意到 $\|\sigma\| \geqslant 1$，f 的连续性含有 $\delta \to 0$ 时 $\omega(f,\delta)_{[-A,A]} \to 0$，所以对于 $x \in [-A,A]$ 有

$$|\Psi_{n,\sigma}(x) - f(x)| \leqslant \|\sigma\|\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]} + \varepsilon.$$

当 $x \in (-\infty, -A)$ 时

$$\Psi_{n,\sigma}(x) - f(x) = f(x_1) - f(x) + \sum_{i=1}^{n}(f(x_{i+1}) - f(x_i))\sigma(\delta^{-1}B(x - x_i)).$$

即有

$$|\Psi_{n,\sigma}(x) - f(x)| \leqslant |f(x) - f(x_0)| + |f(x_0) - f(x_1)| + n\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]}C \leqslant \Phi(A) + \omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]} + \varepsilon.$$

当 $x \in (A, +\infty)$ 时

$$\Psi_{n,\sigma}(x) - f(x) = f(x_{n+1}) - f(x) + \sum_{i=1}^{n}(f(x_{i+1}) - f(x_i))(\sigma(\delta^{-1}B(x - x_i)) - 1).$$

会有

$$|\Psi_{n,\sigma}(x) - f(x)| \leqslant |f(x) - f(x_{n+1})| + n\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]}C \leqslant \Phi(A) + \varepsilon.$$

综上所述，对于 $x \in (-\infty, +\infty)$，

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x) - f(x)| \leqslant \|\sigma\|\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]} + \Phi(A) + \varepsilon.$$

定理 2.2　如果 H 是 Heaviside 函数，$f \in C(\overline{\mathbf{R}})$，$n \in \mathbf{N}$，那么对于任意的 $A > 0$，存在单隐层前向人工神经网络

$$\Psi_{n,\sigma}(x) = a_0 + \sum_{i=1}^{n} a_i H(b_i x + c_i)$$

使得

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x) - f(x)| \leqslant \omega\left(f,\frac{A}{n+1}\right)_{[-A,A]} + \Phi(A).$$

证明　对于任意的 $A > 0$，将 $(-\infty, +\infty)$ 分为三部分 $(-\infty, -A)$，$[-A, A]$，$(A, +\infty)$. 固定 n，取

$$x_i = -A + \frac{2A}{n+1}i, i = 0,1,2,\cdots,n+1.$$

$$t_i = \frac{1}{2}(x_{i-1} + x_i) = -A + \frac{2i-1}{2(n+1)}A,$$

$i = 1,\cdots,n+1$.

构造单隐层前向人工神经网络

$$\Psi_{n,\sigma}(x)=f(t_1)+\sum_{i=1}^{n}(f(t_{i+1})-f(t_i))H(x-x_i).$$

如果 $x\in[x_k,t_{k+1}]$, $k=0,1,2,\cdots,n+1$(对于 $k=n+1$, $x=x_{n+1}=A$),那么

$$\Psi_{n,\sigma}(x)-f(x)=f(t_1)+\sum_{i=1}^{k}(f(t_{i+1})-f(t_i))H(x-x_i)-f(x)=f(t_{k+1})-f(x).$$

于是

$$|\Psi_{n,\sigma}(x)-f(x)|=|f(t_{k+1})-f(x)|\leqslant\omega\left(f,\frac{A}{n+1}\right)_{[-A,A]}.$$

如果 $x\in[t_k,x_k)$, $k=1,2,\cdots,n+1$,那么

$$\Psi_{n,\sigma}(x)-f(x)=f(t_1)+\sum_{i=1}^{k-1}(f(t_{i+1})-f(t_i))H(x-x_i)-f(x)=f(t_k)-f(x).$$

于是

$$|\Psi_{n,\sigma}(x)-f(x)|=|f(t_k)-f(x)|\leqslant\omega\left(f,\frac{A}{n+1}\right)_{[-A,A]}.$$

因此对于 $x\in[-A,A]$,有

$$|\Psi_{n,\sigma}(x)-f(x)|\leqslant\omega\left(f,\frac{A}{n+1}\right)_{[-A,A]}.$$

当 $x\in(-\infty,-A)$ 时

$$\Psi_{n,\sigma}(x)-f(x)=f(t_1)-f(x).$$

于是

$$|\Psi_{n,\sigma}(x)-f(x)|\leqslant|f(t_1)-f(x_0)|+|f(x_0)-f(x)|\leqslant\omega\left(f,\frac{A}{n+1}\right)_{[-A,A]}+\Phi(A).$$

当 $x\in(A,+\infty)$ 时

$$\Psi_{n,\sigma}(x)-f(x)=f(t_{n+1})-f(x).$$

于是

$$|\Psi_{n,\sigma}(x)-f(x)|\leqslant|f(t_{n+1})-f(x_{n+1})|+|f(x_{n+1})-f(x)|\leqslant\omega\left(f,\frac{A}{n+1}\right)_{[-A,A]}+\Phi(A).$$

综上所述,对于 $x\in(-\infty,+\infty)$,

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x)-f(x)|\leqslant\omega\left(f,\frac{A}{n+1}\right)_{[-A,A]}+\Phi(A).$$

下面的定理说明上述逼近偏差是最优的.

定理 2.3　对于每个 $n\geqslant1$ 及 $A>0$,存在 $f_n\in C(\overline{\mathbf{R}})$ 使得

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x)-f_n(x)|\geqslant\omega\left(f_n,\frac{A}{n+1}\right)_{[-A,A]}+\Phi(A).$$

证明　对于任意的 $A>0$,固定 n,取

$$x_i=-A+\frac{2A}{n+1}i,\ i=0,1,2,\cdots,n+1.$$

定义如下分段函数 f_n

$$f_n(x):=\begin{cases}(-1)^i & x=x_i,\\ \frac{n+1}{2A}[(-1)^{i+1}-(-1)^i]\left[x-\left(-A+\frac{2A}{n+1}i\right)\right]+(-1)^i & x\in(x_i,x_{i+1}),\\ 1 & x<x_0,\\ (-1)^{n+1} & x>x_{n+1}.\end{cases}$$

显然 $f_n\in C(\overline{\mathbf{R}})$. 根据 f_n 的定义有如下事实,

$$\omega\left(f_n,\frac{A}{n+1}\right)_{[-A,A]}\leqslant1,\ \Phi(A)=0.$$

由于

$$\Psi_{n,\sigma}(x)=a_0+\sum_{i=1}^{n}a_iH(b_ix+c_i)=a_0+\sum_{i=1}^{n}a_iH\left(b_i\left(x+\frac{c_i}{b_i}\right)\right).$$

是至多有 $n+1$ 个不同值的步函数,且 $(-\infty,+\infty)$ 被分为 $(-\infty,-A)$, $\bigcup_{i=0}^{n}[x_i,x_{i+1}]$, $(A,+\infty)$ 这样 $n+3$ 个子区间,所以至少存在某一个子区间 $I=[x_i,x_{i+1}]$ $(i=0,1,\cdots,n)$(或者 $(-\infty,-A)$, $(A,+\infty)$)有

$$\sup_{x\in I}|\Psi_{n,\sigma}(x)-f_n(x)|\geqslant1.$$

因此

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x)-f_n(x)|\geqslant\sup_{x\in I}|\Psi_{n,\sigma}(x)-f_n(x)|\geqslant1\geqslant\omega\left(f_n,\frac{A}{n+1}\right)_{[-A,A]}+\Phi(A).$$

3　定理 2.1 的应用

若 $f\in C(\overline{\mathbf{R}})$,对于任意的 $\varepsilon>0$,取 A 使得 $\Phi(A)<\varepsilon$,然后取 n 使得 $\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]}<\varepsilon$,即有 $|\Psi_{n,\sigma}(x)-f(x)|<3\varepsilon$. 因此由 ε 的任意性可知

$$\inf_{n}\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x)-f(x)|=0.$$

若 $f\in C(\overline{\mathbf{R}})$ 且在 $\mathbf{R}$ 上 $f\in Lip\alpha$, $0<\alpha<1$

取 A 使得

$$\Phi(A)=\omega\left(f,\frac{2A}{n+1}\right)_{[-A,A]},$$

则有常数 $C>0$ 使得

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x)-f(x)|\leqslant C\left(\frac{1}{\sqrt{n+1}}\right)^{\alpha}.$$

例如对于函数

$$f(x)=\begin{cases}\dfrac{1}{1+x} & x>0,\\ 1 & x\leqslant 0.\end{cases}$$

在任何区间$[-A,A]$上 $f\in Lip1$,而 $\Phi(A)=\dfrac{1}{1+A}$,取 A 使得$\dfrac{2A}{n+1}=\dfrac{1}{1+A}$,则有常数 $C>0$ 使得

$$\sup_{x\in(-\infty,+\infty)}|\Psi_{n,\sigma}(x)-f(x)|\leqslant C\left(\frac{1}{\sqrt{n+1}}\right).$$

【参　考　文　献】

[1] CYBENKO G. Approximation by superpositions of sigmoidal function[J]. Math Control Signals Syst, 1989(2):303-314.

[2] FUNAHASHI K I. On the approximate realization ofcontinuous mappings by neural networks[J]. Neural Networks, 1989(2):183-192.

[3] LI X. On simultaneous approximations by radial basis function neural networks[J]. Appl Math Comput, 1998(95):75-89.

[4] HORNIK K, STINCHOMBE M, WHITE H. Universal approximation of an unknown mapping and its derivertives using multilayer feedforward networks[J]. Neural Networks, 1990(3):551-560.

[5] LESHNO M, LIN V Y, PINKUS A, et al, Multilayer-feedforward networks with a nonpolynominal activiation function can approximate any function[J]. Neural Networks, 1993(6):861-867.

[6] CHEN T P, CHEN H. Universal approximation to nonlinear operators by neural networks with arbitrary activition functions and its application to dynamical system[J]. IEEE Trans Neural Networks, 1995(6):911-917.

[7] CUI C K, LI X. Approximation by ridge function and neural networks with one hidden layer[J]. J Approx Theory, 1992(70):131-141.

[8] CHEN D B. Degree of approximation by superpositons of a sigmoidal function, Approx[J]. Theory and its Appl, 1993,3(9):17-28.

[9] CAO F L, XIE T F, XU Z B. The estimate for approximation error of neural networks: A constructive approach[J]. Neurocomputing, 2008(71):626-630.

[10] JONES L K. A simple lemma on greedy approximation in Hilber space and convergence rates for projection pursuit regression and neural network training[J]. Ann Stat, 1992(20):765-772.

[11] BARRON A R. Universal approximation bounds for superpositions of a sigmoidal function[J]. IEEE Trans Inf Theory, 1993(39):930-945.

[12] MURATA N. An integral representation of functions using three layered networks and their approximation-bounds[J]. Neural Networks, 1996(9):947-956.

[13] HONG B I, HAHM N. Approximation order to a function in $C(\overline{\mathbf{R}})$ by superposition of a sigmoidal function[J]. Applied Mathematics Letters, 2002(15):591-597.