International Mathematical Forum, 5, 2010, no. 41, 2037 - 2041

A Note on Approximation Problems of Neural Network

Ma Yumei, Liu Lijun

Dalian Nationalities University
Department of Mathematics
Dalian Liaoning, 116600, P. R. China

Nan Dong

Beijing University of Technology
Department of Mathematics
Bejing, 100124, P. R. China

Abstract

In this paper, the proofs of approximation on neural network given by T.Chen and Ch.Jiang was revised.

Mathematics Subject Classification: 92B20, 46T30, 46Fxx

Keywords: Neural Network, Distribution function, Radial Basis Function

1 introduction

In 1994 and 1995, T.Chen[1−3] investigated the approximation property of neural network and proved that an activation function belongs to $L^p_{loc}(R^1)\bigcap S'(R^1)$ can uniformly approximation any integrable function on a compact set if and only if the activate function is not a polynomial. 1998, followed Chen's method, Ch.Jiang[4] get almost similar result in radial basis function (RBF) neural networks.
These results are interesting and important. However, the proofs of theorems in Chen[1]and Jiang[4] is not mathematically accurate. Since these theorems play an important role in neural network approximation theorems, and the idea or way of Chen[1]and Jiang[4] are wonderful, only we try in this note is to revise the proofs of these theorems. Some symbols and notations is as follows: $S(R^n)$ is all infinitely differentiable functions, which are rapidly decreasing at

infinity. $S'(R^n)$ is all linear continuous functionals defined on $S(R^n)$, which is also named tempered distribution.
The following lemmas and definitions are given by Rudin[5] :
Lemma 1.1. A distribution is the Fourier transform of a polynomial if and only if its support is the origin (or empty set).
Lemma 1.2. Suppose $f \in S(R^n), g \in S'(R^n)$. Then $\widehat{g * f} = \hat{g} \cdot \hat{f}$.
Lemma 1.3. If $g, f, h \in D'(R^1)$ and two of them have compact supports, then $f * g = g * f,\ f * (g * h) = (g * f) * h$.
Lemma 1.4. If $g \in S'(R^n)$, then $\hat{\hat{g}}(t) = g(-t)$.
Lemma 1.5. If $f \in L^1(R^n)$, then $\hat{f} \in C_0(R^n)$, and $||\hat{f}||_\infty \le ||f||_1$.
Definition 1.6. If $\phi \in S(R^n)$, then its fourier transform $\hat{\phi}(t) = \int \phi(x) e^{it\cdot x} dx$, $t \in R^n$. If $u \in S'(R^n)$ and $\phi \in S(R^n)$, then $\hat{u}(\phi) = u(\hat{\phi})$, where $\hat{u}$ denotes the Fourier transform of u.

T. Chen:

Theorem 1.7[1]. Suppose $g \in L^p_{Loc}(R^1) \cap S'(R^1)$, then $\{\sum_{i=1}^N c_i g(\lambda_i x + \theta_i)\}$ is dense in $L^p[a, b]$ if and only if g is not a polynomial.
Theorem 1.8[2]. Suppose $g \in C(R^1) \cap S'(R^1)$, then $\{\sum_{i=1}^N c_i g(\lambda_i \|x - \theta_i\|)\}$ is dense in $C(K)$ if and only if g is not an even polynomial (K is a compact set of R^1).
Theorem 1.9[3]. Suppose $g \in C(R^1) \cap S'(R^1)$, then $\{\sum_{i=1}^N c_i g(\lambda_i x + \theta_i)\}$ is dense in $C(K)$ if and only if g is not a polynomial K is a compact set of R^1.

Ch. Jiang:

Theorem 1.10[4]. Suppose $g \in L^p_{Loc}(R^n) \cap S'(R^n)$, then $\{\sum_{i=1}^N c_i g(\lambda_i \rho_i x + \theta_i)\}$ is dense in $L^p(K)$ if and only if g is not an even polynomial. (Here, $c_i, \lambda_i \in R^1, \rho_i$ is a rotation, $b_i \in R^n$) K is a compact set of R^n. There are some defects in the proofs of Chen [1−3] and Jiang[4]. Here we only discuss Theorem 1.7 and Theorem 1.10, the other problems are similarity.
Remark 1.11. In Chen[1], the key points to the proof of theorem 1.7 are the statements that "we have $\int_{R^1} g(u) du \int_{R^1} W(u - \lambda x) h(x) dx = 0$ and $< g(t), W(t) * h(\lambda t) >= 0$". However "which is equivalent to "$< \hat{g}(t), \hat{W}(t)\hat{h}(\lambda t) >= 0$" is no reasonable according to the definition of Fourier transform of $S'(R^n)$:if $u \in S'(R^n)$ and $\phi \in S(R^n)$, then $\hat{u}(\phi) = u(\hat{\phi})$. This means we can not do Fourier transform to $g(t)$ and $W(t) * h(\lambda t)$ simultaneous.
Remark 1.12 In Jiang[4], the point to proof of his theorem is the statements that in the eighth line of the proof: for any $p(x) \in D(R^n)$, define $f(x) = p * h(x)$, which follows $\int_{R^n} g^\rho_\lambda(x - t) f(x) dx = 0$. and,

$$g^\rho_\lambda * f = \int_{R^n} g^\rho_\lambda(x - t) f(x) dx = 0.$$

But

$$g^\rho_\lambda * f \neq \int_{R^n} g^\rho_\lambda(x - t) f(x) dx$$

for

$$g_\lambda^\rho * f = \int_{R^n} g_\lambda^\rho(t-x) f(x)dx.$$

2 New Proof of Theorems

2.1 A New Proof of Theorem 1.7. Assuming that $\sum_{i=0}^{N} c_i g(\lambda_i x + \theta_i)$ is not dense in $L^p[a,b]$. By Hahn-Banach Theorem, there exists a function $h \in L^q[a,b](\frac{1}{p}+\frac{1}{q}=1)$, for any $\lambda, \theta \in R^1$, such that

$$\int_a^b g(\lambda x + \theta)h(x) = 0$$

Then for any $w \in S(R^1)$ we have

$$\int_{R^1} w(\theta)d\theta \int_a^b g(\lambda x + \theta)h(x)dx = 0. \tag{1}$$

Let that $h(x) = 0$ if $x \in R^1 \backslash [a,b]$ and that $u = \lambda x + \theta$. The (1) implies that

$$\int_{R^1} g(u)du \int_{R^1} w(u - \lambda x)h(x)dx = 0 \tag{2}$$

Assuming that $w_\lambda(t) = w(\lambda t)$, then (2) is as following

$$< g(u), (w_\lambda * h)(\frac{u}{\lambda}) >= 0 \tag{3}$$

denote that $\lambda = -\xi$, then

$$< g(u), (w_{-\xi} * h)(-\frac{u}{\xi}) >= 0$$

According to Lemma 1.4, i.e. $\hat{\hat{\varphi}}(u) = \varphi(-u)$ for any $\varphi(u) \in S'(R^n)$, we have (3) that

$$< g(u), (w_{-\xi} * h)\hat{}\hat{}(\tfrac{u}{\xi}) >= 0$$

and let $g(u) = g_\xi(\frac{u}{\xi})$, this implies that

$$< g_\xi(\frac{u}{\xi}), (w_{-\xi} * h)\hat{}(\frac{u}{\xi}\hat{}>) = 0. \tag{4}$$

By the definition of Fourier, (4) is that

$$< \hat{g}_\xi(t), (w_{-\xi} * h)\hat{}(t) >= 0 \tag{5}$$

Thus by Lemma 1.2, (5) is really to be

$$< \hat{g}_\xi(t), \hat{w}_{-\xi}(t) \cdot \hat{h}(t) >= 0 \tag{6}$$

Since $w(t)$ and λ are both arbitrary, $w_{-\xi}(t)$ is also arbitrary. (6) hence that $supp\{\hat{g}\} \subset \{0\}$. Then g is a polynomial by Lemma 1.1.

2.2 A New Proof of Theorem 1.10.

Assume that $\{\sum_{i=1}^{N} c_i g(\lambda_i \rho_i x + b_i)\}$ is not dense in $L^p(K)$. According to H-B Theorem, there exists a nonzero $h(x) \in L^q(K)(\frac{1}{p} + \frac{1}{q} = 1)$, for any $\lambda \in R^1$, any $t \in R^n$ and any rotation ρ of R^n, such that

$$\int_K g(\lambda \rho x - t) h(x) dx = 0$$

Suppose that $\lambda > 0$ and that $g_\lambda^\rho(x) = g(\lambda \rho x)$. For $h(x) \neq 0$, $supp\{h\} \subset K$ and $h(x) \in L^q(K)$, then

$$\int_K g(\lambda \rho x - t) h(x) = 0$$

Let $h(x) = 0$ if $x \in R^n \backslash K$. For any $w \in S(R^n)$, we have that

$$\int_{R^n} w(t) dt \int_{R^n} g(\lambda \rho x - t) h(x) dx = 0 \tag{7}$$

Assuming that $u = t - \lambda \rho x$, (7) implies that

$$\int_{R^n} \int_{R^n} g(-u) w(u + \lambda \rho x) h(x) dx du = 0 \tag{8}$$

Denote that $\tilde{w}(\cdot) = w(-\cdot)$, then $w(u + \lambda \rho x) = \tilde{w}(-u - \lambda \rho x)$. Thus

$$\int_{R^n} \int_{R^n} g(-u) \tilde{w}(-u - \lambda \rho x) h(x) dx du = 0$$

and

$$\int_{R^n} g(-u)(\tilde{w} * h) \left(-\rho^{-1} \frac{u}{\lambda}\right) du = 0$$

Replacing u with $-v$ and λ with $-\xi$, we have that

$$< g(v), (\tilde{w} * h) \left(-\rho^{-1} \frac{v}{\xi}\right) >= 0. \tag{9}$$

Let $g_\xi^\rho(\cdot) = g(\xi \rho \cdot)$, (9) is as fowling form:

$$< g_\xi^\rho \left(\rho^{-1} \frac{v}{\xi}\right), (\tilde{w} * h) \left(-\rho^{-1} \frac{v}{\xi}\right) >= 0$$

i.e.

$$< g_\xi^\rho \left(\rho^{-1} \frac{v}{\xi}\right), (\tilde{w} \hat{\hat{*}} h) \left(-\rho^{-1} \frac{v}{\xi}\right) >= 0$$

Thus by Lemma 1.6,

$$< \hat{g}_\xi^\rho, (\tilde{w} \hat{*} h) >= 0$$

This also is that

$$< \hat{g}^{\rho}_{\xi}, \hat{\tilde{w}} \cdot \hat{h} >= 0$$

Next, we use Chen's [1–3] way to show that $g^{\rho}_{\xi}(u)$ is a polynomial. In fact, since $\hat{h}(t) \in C_0(R^n)$, there exists $t_0 \in R^n \backslash \{0\}$ and $O(t_0, \delta) = \{t : |t - t_0| < \delta\}$ such that $\forall t \in O(t_0, \delta)$ and $|\hat{h}(t)| > c > 0|$

For $t_1 \in R^n \backslash \{0\}$, let $t_0 = \lambda\rho(t_1)$ i.e. $x \cdot \rho^{-1}t = t \cdot \rho x$, thus $|\hat{h}(\lambda\rho^{-1}t))| > c$ for any $t \in O(t_1, \frac{\delta}{\lambda})$, which implies that

$$< \hat{g}^{\rho}_{\xi}(u), \hat{w}(u) >=< \hat{g}^{\rho}_{\xi}(u), \frac{\hat{w} \cdot \hat{h}(\lambda\rho^{-1}u)}{\hat{h}(\lambda\rho^{-1}u)} >= 0$$

Since that $w(t) \in D(R^n)$ is arbitrary and that $t \neq 0$, $\tilde{w}$ is arbitrary function in $S(R^n)$. Then $supp\{\hat{g}^{\rho}_{\xi}(u)\} \subset \{0\}$ or $\emptyset$. By Rudin's theorem, we have that $g^{\rho}_{\xi}(u)$ is a polynomial. Further more $g(u)$ is a polynomial, which completes the proof.

References

[1] Chen Tian-Ping, *Approximate Problem on Neural Networks and its Application in System Recognize*[J]. Science in China (Series A), 1994, **24**(1): 1–7. (in Chinese)

[2] Chen Tian-Ping, Chen Hong, *Approximation capability to functions of several variables nonlinear functionals, and operators by radial basis function neural networks*[J]. IEEE Trans. Neural Networks, 1995, **6**(4): 904–910

[3] Chen Tian-Ping, Chen Hong, *Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems*[J]. IEEE Trans. Neural Networks, 1995, **6**(4): 911–917

[4] Jiang Chuan-Hai, *The approximate problem on neural network*[J]. Annal of Math, 1998, **19A**(3): 295–300 (in Chinese)

[5] W. Rudin, Functional analysis[M] Springer-verlag, New York. 1973.

Received: January, 2010