J. Appl. Math. & Informatics Vol. 27(2009), No. 1 - 2, pp. 139 - 147
Website: http://www.kcam.biz

APPROXIMATION ORDER TO A FUNCTION IN $C^1[0,1]$ AND ITS DERIVATIVE BY A FEEDFOWARD NEURAL NETWORK

NAHMWOO HAHM† AND BUM IL HONG*

ABSTRACT. We study the neural network approximation to a function in $C^1[0,1]$ and its derivative. In [3], we used even trigonometric polynomials in order to get an approximation order to a function in L_p space. In this paper, we show the simultaneous approximation order to a function in $C^1[0,1]$ using a Bernstein polynomial and a feedforward neural network. Our proofs are constructive.

AMS Mathematics Subject Classification :41A10, 41A24, 41A29
Key words and phrases : Approximation order, Bernstein polynomial, feedforward neural network

1. Introduction

Approximation by a feedforward neural network has been studied by many researchers since it has many applications in engineerings and computer science. In neural network theory, basically we have two problems. The first problem is the density problem which is related to the question of representing a target function arbitrarily closed by a neural network. The second problem is the complexity problem that is related to the degree of approximation. If we have an approximation order to a complexity problem, then a density result by neural network is trivial.

Hong and Hahm [2] showed complexity results to a continuous function by a neural network with sigmoidal activation function. In [7], Mhaskar and Hahm

Received April 24, 2008. Accepted September 7, 2008. *Corresponding author.
†This research was supported by the University of Incheon Research Fund, 2007.

introduced the generalized translation networks. A feedforward neural network with n neurons is of the form

$$\sum_{i=1}^{n} c_i \sigma(a_i \cdot x + b_i) \tag{1.1}$$

where a_i, b_i and c_i are real numbers for $1 \leq i \leq n$ and σ is a real-valued function defined on $\mathbb{R}$. Using the generalized translation network, Hahm and Hong [3] obtained the complexity result to a function in L_p space.

Recently, the simultaneous approximation of a function and its derivatives has been studied by some researchers. In [1], Gallent and White introduced the problem of the simultaneous approximation to the neural network and they proved the density result using Maclaurin theorem. Li [5] also showed the density result for a multivariate function and its derivative. In [1, 5], they only showed the density results related to simultaneous approximation.

In this paper, we show the simultaneous approximation order to a $C^1[0,1]$ function and its derivative by a feedforward neural network. First of all, we approximate a function and its derivative by a Bernstein polynomial and then approximate that Bernstein polynomial by a feedforward neural network. Our proofs are constructive and we examine our results using numerical examples.

2. Preliminaries

If a function $f(x)$ is defined in $[0,1]$, then the Bernstein polynomial of degree n is given by

$$B_n(f,x) := \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k}. \tag{2.1}$$

From the binomial theorem [8], we can easily derive the followings.

Lemma 2.1. *For $x \in [0,1]$, we have*

(1) $\displaystyle\sum_{k=0}^{n-1} \binom{n-1}{k} x^k (1-x)^{n-k-1} = 1.$

(2) $\displaystyle\sum_{k=0}^{n-1} k \binom{n-1}{k} x^k (1-x)^{n-k-1} = (n-1)x.$

(3) $\displaystyle\sum_{k=0}^{n-1} k^2 \binom{n-1}{k} x^k (1-x)^{n-k-1} = (n-1)x((n-1)x + 1 - x).$

From Lemma 2.1, we can easily obtain the following.

Lemma 2.2. *If* $T(x) = \sum_{k=0}^{n-1}(k-(n-1)x)^2 \binom{n-1}{k} x^k(1-x)^{n-k-1}$ *for* $x \in [0,1]$, *then* $||T||_{\infty,[0,1]} \leq \frac{n-1}{4}$.

Proof. By Lemma 2.1, we have

$$\begin{aligned} T(x) &= \sum_{k=0}^{n-1}(k-(n-1)x)^2 \binom{n-1}{k} x^k(1-x)^{n-k-1} \\ &= \sum_{k=0}^{n-1}\left\{k^2 - 2(n-1)x + (n-1)^2x^2\right\} \binom{n-1}{k} x^k(1-x)^{n-k-1} \\ &= (n-1)x\left\{(n-2)x + 1 - (n-1)x\right\} \\ &= (n-1)(x-x^2). \end{aligned}$$

Thus $|T(x)| \leq \frac{n-1}{4}$ for any $x \in [0,1]$. □

In order to represent the simultaneous approximation order, we use the following definition.

Definition 2.3. *For* $f \in C^1[0,1]$, *we define*

$$\omega_0(f,\delta) := \sup\left\{|f(x) - f(y)| : |x-y| < \delta\right\} \tag{2.3}$$

and

$$\omega_1(f,\delta) := \sup\left\{|f'(x) - f'(y)| : |x-y| < \delta\right\} \tag{2.4}$$

for $\delta > 0$ *and* $x, y \in [0,1]$.

Note that ω_0 and ω_1 satisfy the followings.
(1) ω_0 and ω_1 are non-decreasing.
(2) $\omega_i(f,\alpha\delta) \leq (\alpha+1)\omega_i(f,\delta)$ for a positive real number α and $i = 0,1$.
(3) $\omega_i(f,\alpha+\beta) \leq \omega_i(f,\alpha)+\omega_i(f,\beta)$ for positive real numbers α, β and $i = 0,1$.
Throughout the paper, the letters $d, c, c_1, c_2, c_3, \cdots$ will denote positive constants and their values may be different at different occurrences. In addition, n denotes a natural number that is greater than 1.

3. Main results

In this section, we show the simultaneous approximation order to a function in $C^1[0,1]$ by a Bernstein polynomial.

Theorem 3.1. *For $f \in C^1[0,1]$, we have*

$$||f - B_n(f)||_{\infty,[0,1]} \leq c_1\omega_0\left(f, \frac{1}{\sqrt{n}}\right) \tag{3.1}$$

and

$$||f' - B_n'(f)||_{\infty,[0,1]} \leq c_2\omega_1\left(f, \frac{1}{\sqrt{n-1}}\right) \tag{3.2}$$

where c_1 and c_2 are positive constants which are independent of n.

Proof. The proof of $||f - B_n(f)||_\infty \leq c_1\omega_0\left(f, \frac{1}{\sqrt{n}}\right)$ is given in [6].

By mean value theorem, we have

$$\begin{aligned}
&B_n'(f,x)\\
=&\sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k}\left[kx^{k-1}(1-x)^{n-k} + (n-k)x^k(1-x)^{n-1-k}\right]\\
=&n\sum_{k=0}^{n-1}\left[f(\frac{k+1}{n}) - f(\frac{k}{n})\right]\binom{n-1}{k}x^k(1-x)^{n-1-k}\\
=&\sum_{k=0}^{n-1} f'(\theta_k)\binom{n-1}{k}x^k(1-x)^{n-1-k}
\end{aligned} \tag{3.3}$$

where $\frac{k}{n} < \theta_k < \frac{k+1}{n}$. For any $x \in [0,1]$,

$$\begin{aligned}
&|f'(x) - B_n'(f,x)|\\
\leq&\sum_{k=0}^{n-1}|f'(x) - f'(\theta_k)|\binom{n-1}{k}x^k(1-x)^{n-1-k}\\
\leq&\sum_{k=0}^{n-1}\left|f'(x) - f'(\frac{k}{n-1})\right|\binom{n-1}{k}x^k(1-x)^{n-1-k}\\
+&\sum_{k=0}^{n-1}\left|f'(\frac{k}{n-1}) - f'(\theta_k)\right|\binom{n-1}{k}x^k(1-x)^{n-1-k}.
\end{aligned} \tag{3.4}$$

For a given $\delta > 0$ and $x_0, y_0 \in [0,1]$ with $x_0 < y_0$, we set that $\alpha := \alpha(x_0, y_0, \delta)$ is an integer $\left[\frac{y_0 - x_0}{\delta}\right]$ where $[\cdot]$ is the Gauss function. For $i = 1, 2, \cdots, \alpha + 1$, we set $\beta_i = x_0 + \left(\frac{y_0 - x_0}{\alpha + 1}\right)i$. Then

$$|f'(x_0) - f'(y_0)| \leq \sum_{i=0}^{\alpha}|f'(\beta_{i+1}) - f'(\beta_i)| \leq (\alpha+1)\omega_1(f,\delta). \tag{3.5}$$

By Lemma 2.2 and Lemma 3.1, we have

$$\begin{aligned}
&\sum_{k=0}^{n-1} |f'(x) - f'(\frac{k}{n-1})| \binom{n-1}{k} x^k (1-x)^{n-1-k} \\
\leq& \omega_1(f,\delta) \sum_{k=0}^{n-1} \left[\alpha(x, \frac{k}{n-1}, \delta) + 1 \right] \binom{n-1}{k} x^k (1-x)^{n-1-k} \\
\leq& \omega_1(f,\delta) \left[\sum_{k=0}^{n-1} \alpha(x, \frac{k}{n-1}, \delta) \binom{n-1}{k} x^k (1-x)^{n-1-k} + 1 \right] \qquad (3.6) \\
\leq& \omega_1(f,\delta) \left[\frac{1}{\delta^2} \sum_{k=0}^{n-1} (\frac{k}{n-1} - x)^2 \binom{n-1}{k} x^k (1-x)^{n-1-k} + 1 \right] \\
\leq& \omega_1(f,\delta) \Big(\frac{1}{4(n-1)\delta^2} + 1 \Big).
\end{aligned}$$

If we choose $\delta = \dfrac{1}{\sqrt{n-1}}$, then

$$\sum_{k=0}^{n-1} |f'(x) - f'(\frac{k}{n-1})| \binom{n-1}{k} x^k (1-x)^{n-1-k} \leq c\omega_1\Big(f, \frac{1}{\sqrt{n-1}}\Big). \qquad (3.7)$$

for some positive constant c which is indepedent of n. Now we compute the last part of (3.4).

$$\begin{aligned}
&\sum_{k=0}^{n-1} |f'(\frac{k}{n-1}) - f'(\theta_k)| \binom{n-1}{k} x^k (1-x)^{n-1-k} \\
\leq& \sum_{k=0}^{n-1} \omega_1(f, \frac{1}{n-1}) \binom{n-1}{k} x^k (1-x)^{n-1-k} \qquad (3.8) \\
=& \omega_1\Big(f, \frac{1}{n-1}\Big).
\end{aligned}$$

By (3.7) and (3.8), we have

$$||f' - B'_n(f)||_{\infty,[0,1]} \leq c\omega_1\Big(f, \frac{1}{\sqrt{n-1}}\Big) + \omega_1\Big(f, \frac{1}{n-1}\Big) := c_2\omega_1\Big(f, \frac{1}{\sqrt{n-1}}\Big)$$

where c_2 is a positive constant which is independent of n. □

Now we show the density result for a target function and its derivative using a feedforward neural network.

Theorem 3.2. *Let σ be a C^∞ function. Suppose that there exists $a \in \mathbb{R}$ such that $\sigma^{(m)}(a) \neq 0$ for any nonnegative intger m. Then for a given $\epsilon > 0$ and a polynomial $P_n(x) = \sum_{i=0}^{n} \alpha_i x^i$, there exists a feedforward neural network*

$$N_{n,h}(\sigma, x) := \sum_{i=0}^{n} \alpha_i \frac{1}{h^i \sigma^{(i)}(a)} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \sigma(jhx + a)$$

such that

$$||P_n - N_{n,h}(\sigma)||_{\infty,[0,1]} < \epsilon \quad \text{and} \quad ||P_n' - N_{n,h}'(\sigma)||_{\infty,[0,1]} < \epsilon \tag{3.9}$$

for sufficiently small $h > 0$.

Proof. Let $\epsilon > 0$ be given. Note that $\frac{1}{h^i \sigma^{(i)}(a)} \sum_{j=0}^{i} (-1)^{i-j} \sigma(jhx + a) \to x^i$ as $h \to 0$. Therefore we can easily see that $||P_n - N_{n,h}(\sigma)||_{\infty,[0,1]} < \epsilon$ for sufficiently small $h > 0$. Note that

$$\begin{aligned} &N_{n,h}'(\sigma, x) \\ &= \sum_{i=0}^{n} \alpha_i \frac{1}{h^i \sigma^{(i)}(a)} \sum_{j=0}^{i} (-1)^{i-j} \binom{i}{j} \sigma'(jhx + a) jh \\ &= \sum_{i=0}^{n} \alpha_i \frac{i}{h^{i-1} \sigma^{(i)}(a)} \sum_{j=0}^{i-1} (-1)^{i-j-1} \binom{i-1}{j} \sigma'((j+1)hx + a). \end{aligned} \tag{3.10}$$

Now we use the Taylor formula. For any natural number $l > n-1$, the equation (3.10) can be rewritten as

$$\begin{aligned} &N_{n,h}'(\sigma, x) \\ &= \sum_{i=0}^{n} \alpha_i \frac{i}{h^{i-1} \sigma^{(i)}(a)} \sum_{j=0}^{i-1} (-1)^{i-j-1} \binom{i-1}{j} \sigma'(jhx + a) \qquad (3.11) \\ &+ \sum_{i=0}^{n} \alpha_i \frac{i}{h^{i-1} \sigma^{(i)}(a)} \sum_{j=0}^{i-1} (-1)^{i-j-1} \binom{i-1}{j} \left(\sum_{r=1}^{l-1} \frac{\sigma^{(r+1)}(jhx + a)}{r!} (jx)^r \right) \\ &+ \sum_{i=0}^{n} \alpha_i \frac{i}{h^{i-1} \sigma^{(i)}(a)} \sum_{j=0}^{i-1} (-1)^{i-j-1} \binom{i-1}{j} \left(\frac{\sigma^{(l+1)}(jhx + a + \theta)}{l!} (hx)^l \right) \end{aligned}$$

where θ is a point between $jhx + a$ and $(j+1)hx + a$. Since

$$\sum_{i=0}^{n} \alpha_i \frac{i}{h^{i-1}\sigma^{(i)}(a)} \sum_{j=0}^{i-1} (-1)^{i-j-1} \binom{i-1}{j} \sigma'(jhx+a) = \sum_{i=0}^{n} \alpha_i i N_{m-1,h}(\sigma', x)$$

and the last two terms of (3.11) have the accuracy $\mathcal{O}(h)$ respectively, we have

$$||P_n' - N_{n,h}'(\sigma)||_{\infty,[0,1]} < \epsilon$$

for sufficiently small $h > 0$. □

Theorem 3.3. *Let $f \in C^1[0,1]$ and σ be a C^∞ function. Suppose that there exists $a \in \mathbb{R}$ such that $\sigma^{(m)}(a) \neq 0$ for any nonnegative integer m. Then there exists a feedforward neural network $N_{n,h}(\sigma, x)$ such that*

$$||f - N_{n,h}(\sigma)||_{\infty,[0,1]} \leq c_1 \omega_0\Big(f, \frac{1}{\sqrt{n}}\Big) \tag{3.12}$$

and

$$||f' - N_{n,h}'(\sigma)||_{\infty,[0,1]} \leq c_2 \omega_1\Big(f, \frac{1}{\sqrt{n-1}}\Big) \tag{3.13}$$

for sufficiently small $h > 0$ and some positive constants c_1 and c_2 which are independent of n.

Proof. Let $\epsilon > 0$ be given. By Theorem 3.1, there exists a Bernstein polynomial $B_n(f, x)$ such that

$$||f - B_n(f)||_{\infty,[0,1]} \leq c_1 \omega_0\Big(f, \frac{1}{\sqrt{n}}\Big) \tag{3.14}$$

and

$$||f' - B_n'(f)||_{\infty,[0,1]} \leq c_2 \omega_1\Big(f, \frac{1}{\sqrt{n-1}}\Big). \tag{3.15}$$

Moreover, by (3.9), there exists a feedforward neural network $N_{n,h}(\sigma, x)$ such that

$$||B_n - N_{n,h}(\sigma)||_{\infty,[0,1]} < \epsilon \quad \text{and} \quad ||B_n' - N_{n,h}'(\sigma)||_{\infty,[0,1]} < \epsilon. \tag{3.16}$$

for sufficiently small $h > 0$. Therefore, from (3.14), (3.15) and (3.16), we have

$$||f - N_{n,h}(\sigma)||_{\infty,[0,1]} < c_1 \omega_0\Big(f, \frac{1}{\sqrt{n}}\Big) + \epsilon$$

and

$$||f' - N'_{n,h}(\sigma)||_{\infty,[0,1]} < c_2\omega_1\Big(f, \frac{1}{\sqrt{n-1}}\Big) + \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we get (3.12) and (3.13). □

4. Numerical results

We demonstrate numerical results to support our theoretical results. We proved in Theorem 3.3 that the differentiable target function over $[0, 1]$ can be simultaneously approximated by a neural network. To show this, we select e^{-x^2} as a target function and $cosx$ as an activation function of the neural network. Note that $cos^{(n)}(\pi/4) \neq 0$ for any nonnegative integer n.

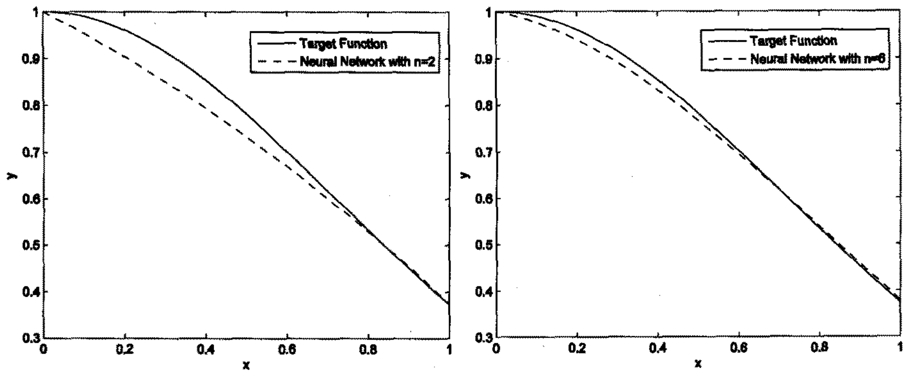

Figure 1. The target function and a neural network with $n = 2$ and $n = 6$.

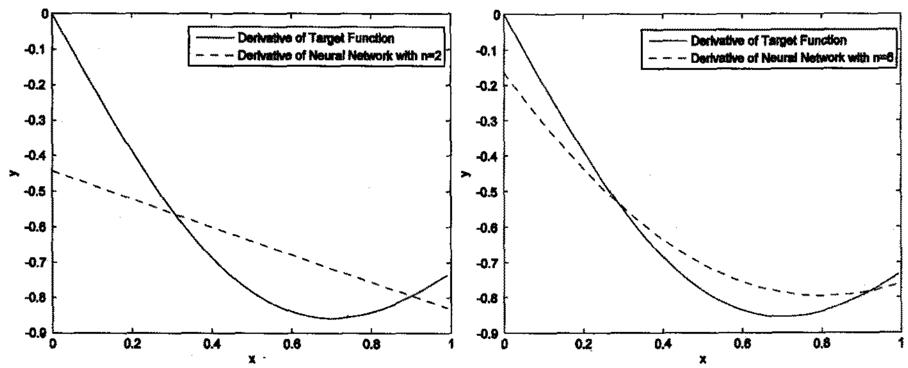

Figure 2. The derivative of the target function and the derivative of a neural network with $n = 2$ and $n = 6$.

Even though Figure 1 and Figure 2 show that the approximation order to the derivative of a target function by the derivative of a neural network is lower than that to a target function by a neural network, Figure 2 clearly shows that the derivative of a target function can be approximated well by the derivative of a neural network as the number of hidden units increase. We will investigate the approximation order to the nth derivative of a target function by the nth derivative of a neural network in the future.

REFERENCES

1. Gallant and H. White, *On Learning the Derivatives of an Unknown Mapping with Multilayer feedforward Networks,* Neural Networks 5(1992), 129-138.
2. B. I. Hong and N. Hahm, *Approximation Order to a Function in $\bar{C}(\mathbb{R})$ by Superposition of a Sigmoidal Function*, Applied Mathematics Letters 15(2002), 591 - 597.
3. N. Hahm and B. I. Hong, *Approximation Order to a Function in L_p Space by Generalized Translation Networks*, Honam Mathematical Journal 28(2006), 125-133.
4. Fengjun Li and Zongben Xu, *The Estimation of Simultaneous Approximation Order for Neural Networks*, Chaos Solutions and Fractals 36(2008), 572-580.
5. X. Li, *Simultaneous Approximation of a Multivariate Functions and Their Derivatives by Neural Networks with One Hidden Layer*, Neurocomputing 12(1996), 327-343.
6. G. G. Lorentz, *Bernstein Polynomials*, Chelsea, (1986).
7. H. N. Mhaskar and N. Hahm, *Neural Networks for Functional Approximation and System Identification*, Neural Computation 9(1997), 143-159.
8. I. P. Natanson, *Constructive Function Theory*, Frederick Ungar Publ., (1964).

Nahmwoo Hahm received his Ph.D. at the University of Texas at Austin in 2000. Since 2000, he has been a professor of Univeristy of Incheon. His research interests: numerical analysis and neural networks.

Department of Mathematics, University of Incheon, Incheon 402-749, Korea
e-mail: nhahm@incheon.ac.kr

Bum Il Hong received his Ph.D. at Purdue University in 1990. Since 1996, he has been in the department of mathematics at Kyung Hee Univeristy as a professor. His research interests are numerical analysis, numerical p.d.e. and neural networks.

Department of Mathematics, Kyung Hee University, Yongin, Kyunggi 446-701, Korea
e-mail: bihong@khu.ac.kr