Bull. Korean Math. Soc. **38** (2001), No. 2, pp. 293–302

A NOTE ON GREEDY ALGORITHM

NAHMWOO HAHM* AND BUM IL HONG**

ABSTRACT. We improve the greedy algorithm which is one of the general convergence criterion for certain iterative sequence in a given space by building a constructive greedy algorithm on a normed linear space using an arithmetic average of elements. We also show the degree of approximation order is still $\mathcal{O}(1/\sqrt{n})$ by a bounded linear functional defined on a bounded subset of a normed linear space, which offers a good approximation method for neural networks.

1. Introduction

The greedy algorithm is the algorithm which solves a problem by making a sequence of local decisions and makes decisions based solely on local information. This algorithm is one of the important tools in many areas, especially in neural network approximation. In recent years, there has been a great deal of research in the theory of approximation of real valued functions using artificial neural networks with one hidden layer [1, 4, 7, 10, 11]. Mathematically, a neural network can evaluate a special function, depending upon its architecture. Many different types of neural network models are studied, but we describe just one in this paper, called a "feedforward network with one hidden layer". Feedforward neural network with one hidden layer is of the form

$$\sum_{i=1}^{m} \beta_i \sigma(\mathbf{a}_i \mathbf{x} + b_i)$$

Received August 22, 2000.

2000 Mathematics Subject Classification: 41A29, 41A46.

Key words and phrases: approximation with constraints, approximation by arbitrary nonlinear expressions; widths and entropy.

*This research was supported by Kyung Hee University Research Fund, 1999.
**This research was supported by Kyung Hee University Research Fund, 2000.

where σ is a nonlinear activation function, $\mathbf{x} \in \mathbb{R}^n$, the weights $\mathbf{a}_i \in \mathbb{R}^n$, and the thresholds b_i and the coefficients β_i are real numbers for $i = 1, \dots, m$.

There are many different proofs in the area of neural networks related to the proposition that any continuous real functions over arbitrary compact subsets of R^n can be approximated arbitrarily well by neural networks with one hidden layer. This is called the density problem [8, 11]. In the theory of approximation by neural networks, a major problem beyond the universal approximation problem is the complexity. This problem is related to determining the number of neurons necessary to a given order of approximation for all target functions in a certain class. The complexity problem has been studied by many authors [1, 4, 6, 9, 10].

Chakravathy and Ghosh [2], and Ghosh and Tumer [5] indicated that the problem of determining the number of hidden nodes in order to achieve a given degree of approximation bound is very important in practice even though they do not show any theoretical error bounds. Barron [1] showed the dimension-independent bounds for a certain class of functions defined by the Fourier transform properties for certain L_2 approximation. Using radial basis function approximation, Mhaskar and Micchelli [10] obtained dimension-independent bounds similar to those obtained by Barron [1], but with a more general class of activation functions and much easier proofs. Jones [8] showed a method for constructing certain approximations to a general element in the closure of the convex hull of a subset of an inner product space and this is a connection with neural networks.

In this paper, we construct a greedy algorithm using the arithmetic average of elements of a given subset in a normed linear space. We also build upon this work on approximation by functionals defined on a normed linear space to study the complexity problem for neural networks and show the approximation rate of order $\mathcal{O}(1/\sqrt{n})$. This shows our proofs are constructive.

2. Greedy algorithm on an inner product space

DEFINITION 2.1. Let S be a given subset in a normed linear space.

Then the convex hull of S is the set

$$co(S) = \left\{ \sum_{i=1}^{n} \alpha_i s_i \mid \alpha_i \geq 0, \quad \sum_{i=1}^{n} \alpha_i = 1, \quad s_i \in S \right\}$$

and $\bar{co}(S)$ denotes the closure of $co(S)$.

In a subset S of a normed linear space, it is well known that if $f \in \bar{co}(S)$, there exist a sequence $\{r_n\} \in S$ and a sequence $\{s_n\}$ with $s_n \in co\{r_1, r_2, \ldots, r_n\}$ such that $s_n \to f$ where $r_n \in S$ for a positive integer n. However we do not know the exact form of s_n.

In the following lemma, Barron [1] indicated that s_n in $co(S)$ could be represented as

$$\alpha s_{n-1} + (1-\alpha) r_n$$

for some $\alpha \in [0,1]$ and $r_n \in S$ when the given space is an inner product space. Note that in an inner product space, we define $\|f\| = \sqrt{< f, f >}$.

LEMMA 2.2. *If S is a bounded subset in an inner product space and $b = \sup_{s \in S} \|s\|$, then for $f \in \bar{co}(S)$ and $h \in co(S)$,*

$$\inf_{s \in S} \|f - \alpha h - (1-\alpha)s\|^2 \leq \alpha^2 \|f - h\|^2 + (1-\alpha)^2 (b^2 - \|f\|^2)$$

where $0 \leq \alpha \leq 1$.

Proof. See Barron [1]. □

We define the Chebyshev radius and the Chebyshev center of a bounded subset S of an inner product space.

DEFINITION 2.3. Let S be a subset of an inner product space X.Then

$$r_* = \inf_{x \in X} \sup_{s \in S} \|s - x\|$$

is called the Chebyshev radius of S and an element $x_* \in X$ satisfying

$$r_* = \sup_{s \in S} \|s - x_*\|$$

is called the Chebyshev center of S.

REMARK. If a given space in Lemma 2.2 is a Hilbert space, we obtain the following result which is more generalized. The following lemma is just mentioned in [3] and we prove it in detail.

LEMMA 2.4. *Let S be a bounded set in a Hilbert space X. If $f \in \bar{co}(S)$ and $h \in co(S)$, then*

$$\inf_{s \in S} \|f - \alpha h - (1-\alpha)s\|^2 \leq \alpha^2 \|f - h\|^2 + (1-\alpha)^2 \gamma$$

where $\gamma = r_*^2 - \|f - x_*\|^2$ such that r_* and x_* are the Chebyshev radius and the center of S, respectively.

Proof. If we replace f, h and S in Lemma 2.2 with $f - x, h - x$ and $S - x$, respectively where x is an element in a Hilbert space, then

$$\begin{aligned} &\inf_{s \in S} \|f - \alpha h - (1-\alpha)s\|^2 \\ &\leq \alpha^2 \|f - h\|^2 + (1-\alpha)^2 \sup_{s \in S} (\|s - x\|^2 - \|f - x\|^2). \end{aligned}$$

Therefore we have, by taking the infimum over a Hilbert space,

$$\begin{aligned} &\inf_{s \in S} \|f - \alpha h - (1-\alpha)s\|^2 \\ &= \inf_{x \in X} \inf_{s \in S} \|f - \alpha h - (1-\alpha)s\|^2 \\ &\leq \inf_{x \in X} [\alpha^2 \|f - h\|^2 + (1-\alpha)^2 \sup_{s \in S} (\|s - x\|^2 - \|f - x\|^2))] \\ &= \alpha^2 \|f - h\|^2 + (1-\alpha)^2 \inf_{x \in X} \sup_{s \in S} (\|s - x\|^2 - \|f - x\|^2) \\ &= \alpha^2 \|f - h\|^2 + (1-\alpha)^2 (r_*^2 - \|f - x_*\|^2). \end{aligned}$$

The last equation is true since every bounded set in a Hilbert space has a unique Chebyshev center. Thus we complete the proof. □

We have the greedy algorithm on an inner product space suggested by Barron [1] as follows:

Let S be a bounded subset in an inner product space X and $f \in \bar{co}(S)$. Let $\gamma = \inf_{x \in X} \sup_{s \in S} \{\|s - x\|^2 - \|f - x\|^2\}$ and let C be a positive constant with $C > \gamma$ and $\epsilon_n = C(C - \gamma)/n(C - \gamma + n\gamma)$. We choose $s_1 \in S$ such that

$$\|f - s_1\|^2 \leq \inf_{s \in S} \|f - s\|^2 + \epsilon_1.$$

Inductively, we choose $s_n = \beta s_{n-1} + (1-\beta)s^*$ for some suitable $\beta \in [0,1]$ and $s^* \in S$ such that

$$\|f - s_n\|^2 \le \inf_{0\le\alpha\le 1} \inf_{s\in S} \|f - \alpha s_{n-1} - (1-\alpha)s\|^2 + \epsilon_n.$$

In this algorithm, we do not know the exact form of suitable β and so it is hard to apply this algorithm in practice. However, using this algorithm and Lemma 2.2, Barron [1] showed the following result by mathematical induction.

THEOREM 2.5. *Let S be a bounded set in an inner product space. If $f \in \bar{co}(S)$, then there exists a sequence $\{s_n\}$ in $co(S)$ such that*

$$\|f - s_n\| < \frac{C}{\sqrt{n}}$$

where C is a constant such that $C > \inf_{x\in X} \sup_{s\in S}\{\|s-x\|^2 - \|f-x\|^2\}$.

As we mentioned before, the constant C in Theorem 2.5 is complicated and we do not know the exact form of $s_n \in co(S)$ and so it is hard to use it in practice.

COROLLARY 2.6. *If S be a bounded set in an inner product space and $f \in \bar{co}(S)$, then*

$$\inf \|f - \sum_{i=1}^{n} \alpha_i s_i\| \le \frac{C}{\sqrt{n}}$$

where the infimum is taken over all $s_1, \dots, s_n \in S$ and all nonnegative $\alpha_1, \dots, \alpha_n$ whose sum is 1, and C is a constant such that $C > \inf_{x\in X}$ $\sup_{s\in S}\{\|s-x\|^2 - \|f-x\|^2\}$.

These results can be improved when the given space is finite dimensional. The proof can be easily obtained by the famous Caratheodory Theorem.

THEOREM 2.7. *If S is a subset of an n-dimensional space and $f \in \bar{co}(S)$, then*

$$\inf \|f - \sum_{i=0}^{n} \alpha_i s_i\| = 0,$$

where the infimum is taken over all $s_i \in S$ and $\alpha_i \ge 0$ for $0 \le i \le n$ with $\sum_{i=0}^{n} \alpha_i = 1$.

3. Greedy algorithm on a normed linear space

As we pointed out before, the algorithm suggested by Barron [1] proved the error in approximating an element of $\bar{C}o(S)$ with a convex combination of n elements in S. But, the proof in Barron [1] is non-constructive and gives no information concerning the implementation or needed complexity of the structure in practice. We make an observation concerning the proof of Theorem 2.5 and we obtain a corresponding simple algorithm for achieving the same bounds. We show that each s_n in Theorem 2.5 can be taken to be an arithmetic average of n points in S. This result gives some applications to neural networks. A neural network with a single hidden layer with n neurons proves approximation to a function of the form

$$\sum_{i=1}^{m} \beta_i \sigma(\mathbf{a}_i \mathbf{x} + b_i)$$

as output when x is the network input. The activation function σ is fixed and the network is trained by selecting the parameter $\mathbf{a}_i, b_i, \beta_i, m$. We train with the relaxed algorithm by setting $P_n = \{\beta\sigma(\mathbf{a}\mathbf{x} + b) : \mathbf{a} \in \mathbb{R}^n \text{ and } b, \beta \in \mathbb{R}\}$. when we consider the linear functional on a normed linear.

THEOREM 3.1. *Let S be a bounded subset of a real normed linear space and let Φ be a continuous linear functional of norm 1. If $f \in \bar{co}(S)$, then*

$$\inf_{s \in S} |\Phi(\alpha h + (1-\alpha)s - f)|^2 \le \alpha^2 \|h - f\|^2$$

for $h \in co(S)$ and $0 \le \alpha \le 1$.

Proof. It is enough to show that

$$\inf_{s \in S} |\Phi(\alpha h + (1-\alpha)s - f)| \le \alpha \|h - f\|.$$

Let $\epsilon > 0$ be given. Since $f \in \bar{co}(S)$, there exists $p \in co(S)$ such that $\|f - p\| \le \epsilon$. Thus $|\Phi(f - p)| \le \|f - p\| \le \epsilon$ since Φ is a bounded linear functional of norm 1. If we write $p = \sum_{i=1}^{n} \beta_i s_i$ where $\beta_i \ge 0, s_i \in S$ for

$i = 1, 2, \ldots, n$ and $\sum_{i=1}^{n} \beta_i = 1$, then for $\alpha \in [0, 1]$,

$$\begin{aligned}
&\inf_{s\in S} \Phi[\alpha h + (1-\alpha)s - f] \\
&= \inf_{s\in S} \Phi[\alpha(h-f) + (1-\alpha)(s-f)] \\
&\leq \Phi[\alpha(h-f)] + (1-\alpha) \min_{1\leq i\leq n} \Phi(s_i - f) \\
&\leq \alpha\Phi(h-f) + (1-\alpha)\sum_{i=1}^{n} \beta_i \Phi(s_i - f) \\
&= \alpha\Phi(h-f) + (1-\alpha)\Phi(\sum_{i=1}^{n} \beta_i s_i - f) \\
&= \alpha\Phi(h-f) + (1-\alpha)\Phi(p-f) \\
&\leq \alpha\|\Phi\| \cdot \|h-f\| + (1-\alpha)\epsilon \\
&\leq \alpha\|h-f\| + \epsilon.
\end{aligned}$$

Similarly, we have

$$\inf_{s\in S} \Phi[f - \alpha h - (1-\alpha)s] \leq \alpha\|h-f\| + \epsilon$$

and hence

$$|\inf_{s\in S} \Phi[\alpha h + (1-\alpha)s - f]| \leq \alpha\|h-f\| + \epsilon.$$

Since $\inf_{s\in S} |\Phi(s)| \leq |\inf_{s\in S} \Phi(s)|$ and $\epsilon > 0$ is arbitrary, we complete the proof. □

Comparing to Lemma 2.2, the approximation error bound in Theorem 3.1 is very much improved because $(1-\alpha)^2(b^2 - \|f\|^2)$ is not needed any more. This enables us to get a constructive sequence to approximate a function $f \in co(S)$.

THEOREM 3.2. *Let S be a bounded subset of a real normed linear space and let Φ be a continuous linear functional of norm 1. Then for all n, there is $s_n \in co(S)$ such that*

$$|\Phi(f) - \Phi(s_n)|^2 \leq \frac{C}{n}$$

where C is a positive constant.

In order to prove Theorem 3.2, we construct a sequence from the greedy algorithm and this sequence is constructive.

Let C be a positive constant and ϵ_n be a number such that $\epsilon_n \le C/n^2$ for each n. Select $r_1 \in S$ such that

$$|\Phi(f) - \Phi(s_1)|^2 \le \inf_{s \in S} |\Phi(f) - \Phi(s)|^2 + \epsilon_1.$$

If we set $s_1 = r_1$, then $s_1 \in co(S)$. Inductively, we select $r_n \in S$ such that

$$\begin{aligned} &|\Phi(f) - \Phi((1 - \frac{1}{n})s_{n-1} + \frac{1}{n} r_n)|^2 \\ \le\ &\inf_{s \in S} |\Phi(f) - \Phi((1 - \frac{1}{n})s_{n-1} + \frac{1}{n} s)|^2 + \epsilon_n. \end{aligned}$$

Setting $s_n = (1 - \frac{1}{n})s_{n-1} + \frac{1}{n} r_n$ gives

$$\begin{aligned} s_n &= \left(1 - \frac{1}{n}\right) s_{n-1} + \frac{1}{n} r_n \\ &= \left(1 - \frac{1}{n}\right) \left[\left(1 - \frac{1}{n-1}\right) s_{n-2} + \frac{1}{n-1} r_{n-1}\right] + \frac{1}{n} r_n \\ &= \frac{n-2}{n} s_{n-2} + \frac{1}{n} r_{n-1} + \frac{1}{n} r_n. \end{aligned}$$

Therefore, by induction, we get

$$\begin{aligned} s_n &= \frac{1}{n} s_1 + \frac{1}{n} r_2 + \cdots + \frac{1}{n} r_n \\ &= \frac{1}{n} r_1 + \frac{1}{n} r_2 + \cdots + \frac{1}{n} r_n \\ &= \frac{1}{n} \sum_{i=1}^{n} r_i. \end{aligned}$$

Thus s_n is the arithmetic average of $r_1, \ldots, r_n$ and hence $s_n \in co(S)$.

Proof of Theorem 3.2. Let C be a positive constant with $\epsilon_n < C/n^2$ for each n. Then

$$\begin{aligned} |\Phi(f) - \Phi(s_1)|^2 &\le \inf_{s \in S} |\Phi(f) - \Phi(s)|^2 + \epsilon_1 \\ &\le \epsilon_1 < C. \end{aligned}$$

Assume that for $n \geq 2$,

$$|\Phi(f) - \Phi(s_{n-1})|^2 \leq \frac{C}{n-1}.$$

Then

$$\begin{aligned}
&|\Phi(f) - \Phi(s_n)|^2 \\
\leq &\inf_{s \in S} |\Phi(f) - \Phi((1 - \frac{1}{n})s_{n-1} + \frac{1}{n}s)|^2 + \epsilon_n \\
\leq &\Big(\frac{n-1}{n}\Big)^2 |\Phi(f) - \Phi(s_{n-1})|^2 + \epsilon_n \\
\leq &\Big(\frac{n-1}{n}\Big)^2 \cdot \frac{C}{n-1} + \epsilon_n \\
< &\frac{C(n-1)}{n^2} + \frac{C}{n^2} \\
= &\frac{C}{n}.
\end{aligned}$$

Thus we complete the proof. □

References

[1] A. R. Barron, *Universal Approximation Bounds for Superposition of a Sigmoidal Function*, IEEE Trans. Information Theory **39** (1993), 930–945.

[2] S. V. Chakravarthy and J. Ghosh, *Functional Emulation Using Radial Basis Function Networks*, Neural Networks **10** (1997), 459–478.

[3] E. W. Cheney and W. Light, *A Course in Approximation Theory*, Brooks/Cole, New York, 1999.

[4] C. K. Chui, X. Li, and H. N. Mhaskar, *Limitations of the Approximation Capabilities of Neural Networks with One Hidden Layer*, Adv. Comput. Math. **5** (1996), 233–243.

[5] J. Ghosh and K. Tumer, *Structural Adaptation and Generalization in Supervised Feedforward Networks*, J. Artificial Neural Networks with Applications in Speech and Vision **1** (1994), 431–458.

[6] G. Cybenko, *Approximation by Superposition of Sigmoidal Functions*, Mathematics of Control, Signal and Systems **2** (1989), 303–314.

[7] N. Hahm and B. I. Hong, *Extension of localised approximation by neural networks*, Bull. Austral. Math. Soc. **59** (1999), 121–131.

[8] L. K. Jones, *A Simple Lemma on Greedy Approximation in Hilbert Space and Convergence Rates for Projection Pursuit Regression and Neural Network Training*, Annals of Statistics **20** (1992), 608–613.

[9] M. Leshno, V. Lin, A. Pinkus, and S. Schocken, *Multilayer Feedforward Networks with a Nonpolynomial Activation Function Can Approximate Any Function*, Neural Networks **6** (1993), 861–867.

[10] H. N. Mhaskar and C. A. Micchelli, *Dimension-Independent Bounds on the Degree of Approximation by Neural Networks*, IBM J. Research and Development **38** (1994), 277–284.

[11] H. N. Mhaskar and N. Hahm, *Neural Networks for Functional Approximation and System Identification*, Neural Computation **9** (1997), 143–159.

Nahmwoo Hahm, Institute of Natural Sciences, Kyung Hee University, Yongin 449-701, Korea
E-mail: bihong@nms.kyunghee.ac.kr

Bum Il Hong, Dept. of Mathematics, Kyung Hee University, Yongin 449-701, Korea
E-mail: nhahm@lion.inchon.ac.kr