A NOTE ON ERROR BOUNDS FOR APPROXIMATION IN INNER PRODUCT SPACES

Ajit Dingankar and Irwin W. Sandberg

Department of Electrical and Computer Engineering

The University of Texas at Austin

Austin, Texas 78712

Abstract

In a recent paper a method is described for constructing certain approximations to a general element in the closure of the convex hull of a subset of an inner product space. This is of interest in connection with neural networks. Here we give an algorithm that generates simpler approximants with somewhat less computational cost.

I. INTRODUCTION

One of the earliest interesting results in the area of neural networks is the proposition that any continuous real function defined on a compact subset of $\mathbb{R}^k$ (k an arbitrary positive integer) can be approximated arbitrarily well using a single-hidden-layer network with sigmoidal nonlinearities [1]. There are several other results in the literature concerning arbitrarily good approximation that concern more general types of "target" functionals, different network structures, other nonlinearities, and various criteria by which the quality of approximation is gauged.[1] The earliest proofs of these results are nonconstructive and do not address the question of how one determines the number of nonlinearities needed for a given degree of approximation.

In an interesting recent paper, Barron [3] shows that a lemma attributed to B. Maurey can be used in some cases to obtain a bound on the number of nonlinearities. To discuss that lemma, let G be a subset of a real or complex inner product space H with norm $||\cdot||$, such that the elements of G are bounded in norm by some positive constant b, and let $\overline{co}(G)$ denote the closure of the convex hull of G. Maurey's lemma, which concerns the error in approximating an element of $\overline{co}(G)$ with a convex combination of n points in G, is the following:

Lemma 1. Let f be an element of $\overline{co}(G)$ and let $c > (b^2 - ||f||^2)$. Then for each positive integer n there is an f_n in the convex hull of n points of G such that

$$||f - f_n||^2 \leq \frac{c}{n}. \tag{1}$$

As mentioned in [3], this lemma is attributed in [4] to B. Maurey. In the main application given in [3] assumptions are made on an f so that $f(x)$ can be written as an integral with respect to a probability measure. This yields a setting in which f is an element of the closure of the convex hull of functions generated by the integrand. For a related later study concerning approximations using translates of a given function see [5]. An algorithm for achieving the bounds in Lemma 1 is also given in [3]. (The

[1] An example of another result is the proposition [2] that any continuous real functional on a compact subset of a real normed linear space can be approximated arbitrarily well using a single-hidden-layer neural network with a linear functional input layer and exponential (or polynomial or sigmoidal) nonlinearities.

algorithm is an improved version of a result due to Jones [6].) In this algorithm, for each $n \geq 1$ one chooses an f_n of the form $\alpha_n f_{(n-1)} + (1-\alpha_n)g_n$ such that

$$||f_n - f||^2 \leq \inf_{0\leq\alpha\leq 1} \inf_{g\,\in\, G} ||\alpha f_{n-1} + (1-\alpha)g - f||^2 + \epsilon_n$$

in which ϵ_n is a positive number that satisfies $\epsilon_n \leq \mathrm{O}(n^{-2})$. In particular, for large n one chooses g as well as α so that $||\alpha f_{n-1} + (1-\alpha)g - f||$ is nearly minimal.

The purpose of this note is to direct attention to an observation concerning the proof of the lemma described above, and to give a corresponding simpler algorithm for achieving the bounds (1). The observation is just that the proof of Lemma 1 given in [3] shows that each f_n in Lemma 1 can in fact be taken to an *arithmetic average* of n points in G.[2] In the remainder of this note, we show how one can compute such n-point averages.

Let f be an element of $\overline{co}(G)$, and define

$$\gamma = \inf_{v\,\in\, H} \sup_{g\,\in\, G} \{||g - v||^2 - ||f - v||^2\}.$$

Choose $\delta > \gamma$ (it suffices to take $\delta > b^2 - ||f||^2$, because $\gamma \leq b^2 - ||f||^2$ since $0 \in H$).

Set $\epsilon_k = (\delta - \gamma)/k^2$ for $k = 1, 2, \ldots$.

Let $f_1, f_2, \ldots$ be any sequence generated as follows:

Choose $f_1 = g_1 \in G$ such that $||f - g_1||^2 \leq (\inf_{g\,\in\, G} ||f - g||^2) + \epsilon_1$.

For $k = 2, 3, \ldots$ choose $g_k \in G$ such that

$$||f - (1-k^{-1})f_{k-1} - k^{-1}g_k||^2 \leq \inf_{g\,\in\, G} ||f - (1-k^{-1})f_{k-1} - k^{-1}g||^2 + \epsilon_k \qquad (2)$$

and set $f_k = (1-k^{-1})f_{k-1} + k^{-1}g_k$.

Of course f_k is the arithmetic average $k^{-1}\sum_{j=1}^{k} g_j$ of k elements of G. We will prove the following:

Theorem. Under the conditions indicated,

$$||f - f_n||^2 \leq \frac{\delta}{n} \qquad (3)$$

[2] As an aside, the proof is probabilistic and is reminiscent of one used by Shannon in his 1947 random coding argument, in that one observes that if an average is bounded from above by a certain number then at least one of the terms in the expression for the average must also be bounded in this way.

for all n.

Proof

We will use an induction argument based on the following lemma [7, Lemma 3 of Ch. 25] (which is related to a result in [3]).[3]

Lemma 2. Let G, f and γ be as above, and let h belong to the convex hull of G. Then

$$\inf_{g \in G} ||f - \lambda h - (1-\lambda)g||^2 \leq \lambda^2 ||f-h||^2 + (1-\lambda)^2\gamma.$$

for $\lambda \in [0,1]$.

By Lemma 2 and the choice of f_1,

$$\begin{aligned} ||f - f_1||^2 &\leq \inf_{g \in G} ||f-g||^2 + \epsilon_1 \\ &\leq \inf_{g \in G} ||f-g||^2 + (\delta - \gamma) \\ &\leq \gamma + (\delta - \gamma) = \delta, \end{aligned}$$

showing that (3) holds for $n = 1$. Now suppose that $||f - f_{(n-1)}||^2 \leq \delta/(n-1)$ for some $n \geq 2$. By (2) and Lemma 2,

$$\begin{aligned} ||f - f_n||^2 &\leq \inf_{g \in G} ||f - (1-n^{-1})f_{n-1} - n^{-1}g||^2 + \epsilon_n \\ &\leq \left(\frac{n-1}{n}\right)^2 ||f - f_{n-1}||^2 + \frac{1}{n^2}\gamma + \epsilon_n \\ &\leq \frac{\delta(n-1)+\gamma}{n^2} + \epsilon_n \\ &\leq \frac{\delta}{n} - \frac{(\delta-\gamma)}{n^2} + \epsilon_n = \frac{\delta}{n} \end{aligned}$$

which completes the proof.

Besides showing that our algorithm may be used to generate simpler approximants with somewhat less computational cost, the theorem improves Lemma 1 in also the sense that (as simple examples show) δ can be less than $b^2 - ||f||^2$.

[3]The proof of Lemma 2 given in [7] and the proof of the related result in [3] assume that the inner product space is real, but obvious modifications show that both results hold also for complex spaces.

REFERENCES

[1] K. Hornik, M. Stinchcombe, and H. White, "Multi-layer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

[2] I. W. Sandberg, "General structures for classification," to appear in *IEEE Transactions on Circuits and Systems*, May 1994.

[3] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 930-945, May 1993.

[4] G. Pisier, "Remarques sur un resultat non publie de B. Maurey," *Seminaire d' Analyse Fonctionelle*, vol. 1, no. 12, Ecole Polytechnique, Centre de Mathematiques, Palaiseau. 1980–81.

[5] F. Girosi and G. Anzellotti, "Rates of convergence for radial basis functions and neural networks," in *Artificial Neural Networks with Applications in Speech and Vision*, Chapman and Hall, 1993.

[6] L. K. Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training," *Ann. Statist.*, vol. 20, pp. 608-613, March 1992.

[7] E. W. Cheney, "Topics in approximation theory," Course Notes for Mathematics 393D, Department of Mathematics, The University of Texas at Austin, Spring 1993.