Express Letters

Notes on Weighted Norms and Network Approximation of Functionals

Irwin W. Sandberg

One of the earliest 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 (see, for example, [1]). Among other results in the literature concerning arbitrarily good approximation that concern more general types of "target" functionals, different network structures, other nonlinearities, and various measures of approximation errors is the proposition in [2], [3] (see also [4]) that any continuous real nonlinear 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 or radial basis function) nonlinearities. This has applications concerning, for example, the theory of classification of signals (see [4]).

In interesting papers [5], [6] by Chen and Chen related results are given concerning the approximation of nonlinear functionals. In [6] they show, for instance, that for the case in which the nonlinear functional's domain is a compact subset of a Banach space with a basis, the linear functionals can be taken to be the coefficient maps associated with the basis (see their Theorem 3). Here we observe that this type of result follows from either [4, Theorem 1], or the variant of part of [4, Theorem 1] proved in the Appendix,[1] and the observation we state in the form of a proposition in the Appendix.

We also give a tool theorem in the Appendix that is useful in focusing attention on the range of applications of results concerning the approximation of nonlinear functionals. This theorem shows that certain sets of functions defined on unbounded domains are relatively compact[2] in spaces with weighted norms. The theorem is useful because while the usual criteria for compactness concern signals defined on a compact subset of $\mathbb{R}^n$ (where typically $n = 1$, 2, or 3 in engineering and scientific applications), signals defined on infinite n-dimensional intervals are often of interest. A simple corollary of Theorem 2 is that for each positive α and β, the set of all real-valued continuous functions x defined on $\mathbb{R}^n$ such that $\sup_\gamma |x(\gamma)| \leq \alpha$, and such that x satisfies a Lipschitz condition with Lipshitz constant β, is compact with respect to the metric $\rho(x, y) = \sup_\gamma |w(\gamma)[x(\gamma) - y(\gamma)]|$, where w is any continuous positive function on $\mathbb{R}^n$ such that $w(\gamma) \to 0$ as $\|\gamma\| \to 0$.[3] This establishes, for example, a setting in which it is possible to use certain simple network structures (see [4]) to classify patterns represented by real-valued functions defined on $\mathbb{R}^n$. Theorem 2 also provides the key to extending the continuous-time fading-memory approach in [8] so that it addresses the important problem of approximating input-output maps of systems with inputs that need not be continuous. Further details and other types of applications of Theorem 2 are described in [9] and [10].

Manuscript received November 22, 1995. This paper was recommended by Associate Editor B. Sheu.

The author is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA.

Publisher Item Identifier S 1057-7122(96)04867-2.

[1] Theorem 1 in the Appendix is stated without proof in [7] which establishes a connection between [4, Theorem 1] and one of the main results in [5] involving linear functionals that have the special form of a finite sum of integrals of a certain type.

[2] A subset S of a normed space is relatively compact if the closure of S is compact.

Appendix

A. Approximation Theorem

Let C be a nonempty compact subset of a real normed linear space X, and let X^* be the set of bounded linear functionals on X (i.e., the set of bounded linear maps from X to the reals $\mathbb{R}$). For each $\rho > 0$, let Y_ρ be any set of maps from X to $\mathbb{R}$ that is ρ-dense on C in the closed unit ball of X^* in the sense that given $\phi \in X^*$ with $\|\phi\| \leq 1$ there is a $y \in Y_\rho$ such that $|\phi(x) - y(x)| < \rho, x \in C$.

Let U be any set of maps $u: \mathbb{R} \to \mathbb{R}$ such that given $\alpha \geq 1$ and $\sigma > 0$ and any bounded interval $(\beta_1, \beta_2) \subset \mathbb{R}$ there exists a finite number of elements $u_1, \cdots, u_\ell$ of U for which $|\exp(\alpha\beta) - \sum_j u_j(\beta)| < \sigma$ for $\beta \in (\beta_1, \beta_2)$.[4]

Theorem 1: Let f be a continuous map of C into $\mathbb{R}$. Then given $\epsilon > 0$ there are a positive integer k, real numbers $c_1, \cdots, c_k$, elements $u_1, \cdots, u_k$ of U, a positive number ρ and elements $y_1, \cdots, y_k$ of Y_ρ such that $|f(x) - \sum_j c_j u_j[y_j(x)]| < \epsilon$ for $x \in C$.

Proof: Let f be given, and notice that the set V of all functions $v: C \to \mathbb{R}$ of the form $v(x) = \sum_j a_j \exp[\phi_j(x)]$, in which the sum is finite and the a_j and the ϕ_j belong to $\mathbb{R}$ and X^*, respectively, is an algebra under the natural definition of addition and multiplication. By a consequence [11, p. 198] of the Hahn-Banach theorem, given distinct x_a and x_b in C there is a ϕ in X^* such that $\exp[\phi(x_a)] \neq \exp[\phi(x_b)]$, showing that V separates the points of C. It is clear that $v(x) \neq 0$ for some $v \in V$ for each x. Thus, by a version of the Stone-Weierstrass Theorem [12, p. 162], given $\epsilon > 0$ there are a positive integer p, real numbers $d_1, \cdots, d_p$, and elements $w_1, \cdots, w_p$ of X^* such that

$$\left| f(x) - \sum_j d_j \exp[w_j(x)] \right| < \epsilon/3$$

for $x \in C$.[5] Select $\alpha \geq 1$ so that each $z_j := w_j/\alpha$ has norm at most unity.

We may assume that $\sum_j |d_j| \neq 0$. Choose $\gamma > 0$ such that $\gamma \sum_j |d_j| < \epsilon/3$. Let $[a', b']$ be an interval in $\mathbb{R}$ that contains all of the sets $w_j(C)$, and let real a and b be such that $a < a', b > b'$. Select $\eta > 0$ such that $|\exp(\beta_1) - \exp(\beta_2)| < \gamma$ for $\beta_1, \beta_2 \in [a, b]$

[3] A result similar to this for $n = 1$ is proved in [8].

[4] Of course we can take U to be the set $\{\exp(\alpha\cdot), \alpha \geq 1\}$, or the set $\{u: u(\beta) = (\alpha\beta)^n/n!, \alpha \geq 1, n \in \{0, 1, \cdots,\}\}$. Another acceptable choice is $\{u: u(\beta) = cs(w\beta + \rho), c, w, \rho \in \mathbb{R}\}$, where s is a continuous function with $\lim_{\beta \to \infty} s(\beta) = 1$ and $\lim_{\beta \to -\infty} s(\beta) = 0$.

[5] Here we view C as a metric space with the metric derived in the usual way from the norm in X.

with $|\beta_1 - \beta_2| < \eta$. With $\rho = \alpha^{-1}\min(\eta, a' - a, b - b')$, choose $y_j \in Y_\rho$ such that $|z_j(x) - y_j(x)| < \rho, x \in C$ for all j. This gives $|\exp[\alpha z_j(x)] - \exp[\alpha y_j(x)]| < \gamma, x \in C$ for each j (because we have $\alpha y_j(C) \in [a, b]$ and $|\alpha z_j(x) - \alpha y_j(x)| < \eta$ for each j and x), and thus

$$\begin{aligned}\left|f(x) - \sum_j d_j \exp[\alpha y_j(x)]\right| &\le \left|f(x) - \sum_j d_j \exp[\alpha z_j(x)]\right| \\ &\quad + \left|\sum_j d_j \exp[\alpha z_j(x)] - \sum_j d_j \exp[\alpha y_j(x)]\right| \\ &\le \epsilon/3 + \sum_j |d_j| \cdot |\exp[\alpha z_j(x)] - \exp[\alpha y_j(x)]| \\ &\le (2\epsilon)/3, x \in C.\end{aligned}$$

Now let $[c, d] \subset I\!R$ be such that $\alpha y_j(C) \subset [c, d]$ for each j. Pick $u_1, \cdots, u_\ell \in U$ so that $|\exp(\alpha\beta) - \sum_i u_i(\beta)| \le \gamma_1, \beta \in [c, d]$ where $\gamma_1 \sum_j |d_j| < \epsilon/3$. Then

$$\begin{aligned}\left|f(x) - \sum_j \sum_i d_j u_i[y_j(x)]\right| &\le \left|f(x) - \sum_j d_j \exp[\alpha y_j(x)]\right| \\ &\quad + \left|\sum_j d_j \exp[\alpha y_j(x)] - \sum_j \sum_i d_j u_i[y_j(x)]\right| \\ &\le (2\epsilon)/3 + \sum_j \left|d_j \exp[\alpha y_j(x)] - d_j \sum_i u_i[y_j(x)]\right| \\ &\le (2\epsilon)/3 + \sum_j |d_j| \cdot \left|\exp[\alpha y_j(x)] - \sum_i u_i[y_j(x)]\right| \\ &\le (2\epsilon)/3 + \gamma_1 \sum_j |d_j| < \epsilon.\end{aligned}$$

Since $\sum_j \sum_i d_j u_i[y_j(x)]$ can be written in the form $\sum_j c_j u_j[y_j(x)]$, with the c_j, u_j, and y_j in $I\!R$, U, and Y_ρ, respectively, we have proved the theorem.[6] □

B. Approximation of Linear Functionals on Banach Spaces with a Basis

In the proposition below we consider the case in which X and C are as described earlier, and X is an infinite-dimensional Banach space with a (Schauder) basis $e_1, e_2, \cdots$. We use $g_j, j = 1, 2, \cdots$ to denote the functionals with the property that $x = \sum_{j=1}^\infty g_j(x)e_j$ for $x \in X$.

[6] As mentioned earlier, Theorem 1 is a variant stated in [7] of part of [4, Theorem 1]. That part of [4, Theorem 1], which is provable using a direct modification of the proof above, asserts that Theorem 1 above remains true if "a positive number ρ and elements $y_1, \cdots, y_k$ of Y_ρ" is replaced with "and elements $y_1, \cdots, y_k$ of Y," where Y is any set of continuous maps from X to $I\!R$ that is dense in X^* on C, in the sense that for each $\phi \in X^*$ and any $\epsilon > 0$ there is a $y \in Y$ such that $|\phi(x) - y(x)| < \epsilon$, $x \in C$, and U is instead any set of continuous maps $u : I\!R \to I\!R$ such that given $\sigma > 0$ and any bounded interval $(\beta_1, \beta_2) \subset I\!R$ there exists a finite number of elements $u_1, \cdots, u_\ell$ of U for which $|\exp(\beta) - \sum_j u_j(\beta)| < \sigma$ for $\beta \in (\beta_1, \beta_2)$.

Proposition: Given $\rho > 0$ there is a positive integer ℓ such that $|\phi(x) - \sum_{j=1}^{\ell} \phi(e_j)g_j(x)| < \rho, x \in C$ for all $\phi \in X^*$ with $\|\phi\| \le 1$.

Proof: The proposition follows directly from the fact that given ρ there is an ℓ such that $\|\sum_{j=\ell+1}^\infty g_j(x)e_j\| < \rho$, $x \in C$ (see [13, p. 136]). □

C. Weighted-Space Tool

Theorem 2: Let S be a subset of a complete metric space A with metric ρ, and let $T_1, T_2, \cdots$ be maps of A into itself such that

(i) $T_k(S)$ is a relatively compact subset of A for each k, and
(ii) $\rho(s, T_k s) \to 0$ as $k \to \infty$ uniformly for $s \in S$.

Then S is a relatively compact subset of A.

Proof: Let $\epsilon > 0$ be given. Select k so that $\rho(s, T_k s) < \epsilon/2$ for $s \in S$. Since $T_k(S)$ is relatively compact, A contains an $\epsilon/2$-net $a_1, a_2, \cdots, a_p$ for $T_k(S)$ (see p. 200 of [14]).[7] Now let $u \in S$ be given. Choose $j \in \{1, 2, \cdots, p\}$ so that $\rho(T_k u, a_j) < \epsilon/2$. Thus,

$$\rho(u, a_j) \le \rho(u, T_k u) + \rho(T_k u, a_j) < \epsilon/2 + \epsilon/2 = \epsilon$$

showing that $a_1, a_2, \cdots, a_p$ is an ϵ-net for S. Since A is complete and A contains a finite ϵ-net for S for every $\epsilon > 0$, S is relatively compact (by the Theorem on p. 201 of [14]). □

REFERENCES

[1] K. Hornik, M. Stinchcombe, and H. White, "Multi-layer feed forward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.
[2] I. W. Sandberg, "Gaussian basis functions and approximations for nonlinear systems," in *Dig. Ninth Kobe Int. Symp. Electron. Inform. Sciences*, Kobe, Japan, June 1991, pp. 3-1–3-6.
[3] ——, "Structure theorems for nonlinear systems," *Multidimen. Syst. Signal Processing*, vol. 2, no. 3, pp. 267–286, 1991. (See also the Errata in vol. 3, no. 1, p. 101, 1992.)
[4] ——, "General structures for classification," *IEEE Trans. Circuits Syst.—I*, vol. 41, pp. 372–376, May 1994.
[5] T. Chen and H. Chen, "Approximations of continuous functionals by neural networks with application to dynamical systems," *IEEE Trans. Neural Networks*, vol. 4, pp. 910–918, Nov. 1993.
[6] ——, "Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks," *IEEE Trans. Neural Networks*, vol. 6, pp. 904–910, July 1995.
[7] I. W. Sandberg and A. Dingankar, "On approximation of linear functionals on L_p Spaces," *IEEE Trans. Circuits Syst. I*, vol. 42, pp. 402–404, July 1995. (See also the correction in the Dec. 1995 issue.)
[8] S. P. Boyd, "Volterra series: Engineering fundamentals," Ph.D. dissertation, Univ. California, Berkeley, 1985.
[9] A. Dingankar and I. W. Sandberg, "Network approximation of dynamical systems," in *Proc. 1995 Int. Symp. Nonlinear Theory and Its Applicat.*, Las Vegas, NV, Dec. 1995, vol. 2, pp. 357–362.
[10] I. W. Sandberg and L. Xu, "Uniform approximation of multidimensional myopic maps," to be published.
[11] G. Bachman and L. Narici, *Functional Analysis*. New York: Academic, 1966.
[12] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.
[13] L. A. Liusternik and V. J. Sobolev, *Elements of Functional Analysis*, New York: Frederick Ungar, 1961.
[14] I. P. Natanson, *Theory of Functions of a Real Variable*. New York: Frederick Ungar, vol. II, 1960.

[7] In [14], compact means what we call relatively compact.