Algorithm of incremental approximation using variation of a function with respect to a subset

Kateřina Hlaváčková * Marcello Sanguineti **

* Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 2, 182 07, Praha 8, Czech Republic
E-mail: katka@uivt.cas.cz
** Department of Communications, Computer, and System Sciences
DIST–University of Genoa, Via Opera Pia 13, 16145 Genova, Italy
E-mail: marce@dist.unige.it

Abstract

We present an algorithm of incremental approximation by feedforward neural networks for functions in the closure of finite subsets of a Hilbert space, by using the concept of variation of a function with respect to a set. Since the number of hidden units and their parameters are determined incrementally, the algorithm optimizes one node at a time. Thus, at each step it is required to solve an optimization problem with a small number of parameters.

1 Introduction

The rate of approximation of real functions by feedforward neural networks has been recently studied by various authors. Jones [3] introduced a recursive construction of approximants with "dimension-independent" rate of convergence. Together with Barron [1], he proposed to apply this construction to the functions computable by one-hidden-layer neural networks. Several authors (e.g., Barron [1], Kůrková et al. [5], Mhaskar and Micchelli [9]) characterized sets of functions with d real variables which can be approximated by networks with n hidden units of various types, within an error $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$.

The standard approach to the construction of such networks sets the number of hidden units in advance and then determines the parameters of the entire network. In this case, it is necessary to solve a nonlinear optimization problem in a high-dimensional space. Moreover, in advance it is difficult to estimate the number of units able to guarantee a given degree of accuracy in the approximation. A possible procedure to optimize the network size for a class of networks characterized by a given structure consists in training a (supposed) larger–than–needed network and then removing the units not actively used, i.e. pruning of the unnecessary units, according to a reduction cost (e.g. [10]). In practice, the choice of the pruning criterion is non-trivial and, moreover learning of initial huge networks is cumbersome and time-consuming. This motivates the interest in incremental algorithms, where the number of hidden units and their parameters are determined incrementally. Such algorithms require solving a nonlinear optimization problem in a lower dimensional space at each step, since the search task consists in optimizing one node at a time ([1], [4], [7]).

Barron [1] presents an algorithm for incremental approximation in Hilbert spaces, which can be reformulated in terms of the norm "variation of a function with respect to a set of functions". Such a norm, defined by Kůrková [6], generalizes a concept introduced by Barron [2] and provides a deeper insight into the approximation capabilities of neural networks. In this paper, we propose an incremental algorithm for approximation of functions in the closure of finite subsets of a Hilbert space using the variation norm. The algorithm does not require knowledge of the variation of the function f to be approximated, which could be difficult to compute. Instead, it uses the variation of the n-th approximation f_n of f, which is easier to obtain.

The paper is organized as follows. Section 2 reviews some definitions and results on incremental approximation. Our incremental algorithm is presented in section 3, and section 4 is devoted to its discussion.

2 Preliminaries

Let $(\mathcal{X}, \|\cdot\|)$ be a real vector space. *cl conv* $\mathcal{G}$ denotes the closure of the convex hull of $\mathcal{G}$, where $\mathcal{G}$ is a subset of $\mathcal{X}$; the closure is taken with respect to the topology generated by the norm $\|.\|$. $\mathcal{N}_+$

the elements of the tolerance sequence relative to the case in which $V(f,\mathcal{G})$ is known. The price for using a sequence of approximants $\{V(f_n,\mathcal{G}), n \in \mathcal{N}_+\}$ instead of $V(f,\mathcal{G})$ itself is the fact that minimization in step n of the algorithm must be made within a smaller error that the error acceptable for the case in which $V(f,\mathcal{G})$ is known. This drawback is not so high in comparison to other alternative approaches to incremental learning based on pruning algorithms. Pruning algorithms are computationally burdensome as they require to perform the learning of big networks. Moreover, it is not trivial to choose a pruning criterion: a small weight value need not be a reliable indication of redundancy.

Acknowledgements

We are grateful to Věra Kůrková for valuable discussions and remarks. K. Hlaváčková was partly supported by GA AV grants A2030602 and A2075606. M. Sanguineti was partly supported by grant CNR 96.02472.CT07.

References

[1] Barron, A.R.: Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory* 39, 930-945, 1993.

[2] Barron, A. R.: Neural net approximation. *Proc. 7th Yale Workshop on Adaptive and Learning Systems.* K. Narendra Ed., Yale University Press, 1992.

[3] Jones, L.K.: A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training. *Annals of Statistics* 20, 608-613, 1992.

[4] Kůrková, V. and Beliczynski, B.: Incremental approximation by one-hidden-layer neural networks. *Proceedings of ICANN 1995 Conference,* 505-510, 1995.

[5] Kůrková, V., Kainen, P.C., Kreinovich, V.: Estimates of the number of hidden units and variation with respect to half-spaces. *Neural Networks,* 10, 1061 -1068, 1997.

[6] Kůrková, V.: Dimension-independent rates of approximation by neural networks. *Computer-intensive methods in Control and Signal Processing: Curse of Dimensionality* (Eds. K. Warwick, M. Kárný). Birkhäuser, Boston, 261-270, 1997.

[7] Kůrková, V.: Incremental approximation by neural networks. *Computer-intensive methods in Control and Signal Processing: A Neural Networks Approach* (Eds. M. Kárný, K. Warwick, and V. Kůrková). Springer, London, 177-188, 1997.

[8] Kůrková, V., Savický, P., Hlaváčková, K.: Representations and rates of approximation of real valued Boolean functions by neural networks. *Neural Networks*, in print.

[9] Mhaskar, H.N., Micchelli, C.A.: Dimension-independent bounds on the degree and of approximation by neural networks. *IBM Journal of Research and Development*, 38, 277-284, 1994.

[10] Reed, R.: Pruning algorithms–A survey. *IEEE Trans. on Neural Networks* 4, 740-747, 1994.