Some Estimates of the Approximation Error of Continuous Functions by RBF and KBF networks

Hlaváčková, Kateřina
1994

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

SOME ESTIMATES OF THE APPROXIMATION ERROR OF CONTINUOUS FUNCTIONS BY RBF AND KBF NETWORKS

Kateřina Hlaváčková

Technical report No. 602

December 2, 1994

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
phone: (+422) 66414244 fax: (+422) 8585789
e-mail: katka@uivt.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

SOME ESTIMATES OF THE APPROXIMATION ERROR OF CONTINUOUS FUNCTIONS BY RBF AND KBF NETWORKS

Kateřina Hlaváčková[1]

Technical report No. 602
December 2, 1994

Abstract

This paper deals with approximation of continuous functions by networks with radial basis function (RBF) units and kernel basis function (KBF) units based on classical convolution kernels. We derive some estimates of the approximation error as a function of the number of hidden units.

Keywords

approximation error, kernel basis function, radial basis function

[1]This work is based on the grant GACR 201/93/0427

1 Introduction

Radial basis function (RBF) networks were introduced in the neural nettworks theory as an alternative architecture to hierarchies of perceptrons (Broomhead and Lowe [1]). They have been successfully applied to problems such as e.g. timeseries prediction (Moody and Darken [7]). Theoretically approximation properties of RBF networks with Gaussian radial function were studied by Girosi and Poggio [2] and by Hartman et al. [3] and for more general radial functions by Park and Sandberg [8].

In [4], we showed how classical methods of the approximation of functions by convolutions with kernel functions imply universal approximation properties of RBF networks with non-zero integrable radial functions and introduced kernel basis function (KBF) units. We showed that these networks have the universal approximation property and extended learning algorithms to KBF networks.

In this paper, we present some estimates of rates of approximation. We show that for any of a number of classical kernel functions rate of approximation is bounded above by terms depending on moduli of continuity and convolution approximation error. Using Jackson's estimate, we give an upper bound on approximation error for KBF networks with Jackson convolution kernel. Further, we present an upper bound on approximation error for Lipschitz continuous functions.

Chapter 2 recalls some definitions used in the paper. RBF networks are discussed in chapter 3. The approach of approximation by convolutions is presented in chapter 4. Chapter 5 discusses KBF networks and finally chapter 6 presents our estimates of the error of approximation.

2 Preliminaries

By $\mathcal{R}$ and $\mathcal{N}$ we denote the set of real numbers and positive integers, respectively; also, $I = [0,1]$ and $\mathcal{R}_+ = [0,\infty)$. For a bounded function $f : \mathcal{R}^d \to \mathcal{R}$ the uniform norm is defined by

$$\| f \|_\infty = \sup_{x \in \mathcal{R}^d} |f(x)|.$$

As usual, for a compact subset A of $\mathcal{R}^d, \mathcal{C}(\mathcal{A})$ denotes the set of all real-valued continuous functions on A with the uniform norm and corresponding topology. A *convolution* of two functions $f, g : \mathcal{R}^d \to \mathcal{R}$ is $f * g = \int\limits_{\mathcal{R}^d} f(x)g(x-y)dy$. Let $f : \mathcal{R} \to \mathcal{R}$ be a continuous function, $A \subseteq R$, put $\|f\|_A = \sup_{x \in A} |f(x)|$.

$\omega_A(f,h) = \sup\limits_{\substack{|x_1 - x_2| \leq h, \\ x_1, x_2 \in A}} |f(x_1) - f(x_2)|$ is *modulus of continuity of* f *on* A.

3 RBF networks

Recall here the definition of RBF networks.

A *radial basis function (RBF) unit* with d inputs is a computational unit that computes a function from $\mathcal{R}^d$ to $\mathcal{R}$ of the form $\phi(\| x - c \| / b)$, where $\phi : \mathcal{R} \to \mathcal{R}$ is an even (radial) function, $\| \, . \, \|$ is a norm on $\mathcal{R}^d$, and $c \in \mathcal{R}^d$, $b \in \mathcal{R}$, $b > 0$ are parameters called *center* and *width*, resp.

A *radial basis function (RBF) network* is a neural network with a single linear output unit, one hidden layer with RBF units that have the same radial function ϕ and the same norm $\| \, . \, \|$ on $\mathcal{R}^d$, and d inputs.

By $\mathcal{F}(\phi, \| \, . \, \|)$ we denote the set of real-valued functions on I^d computable by RBF networks with the radial function ϕ and the norm $\| \, . \, \|$ with any number of hidden units:

$$\begin{aligned}\mathcal{F}(\phi, \| \, . \, \|) &= \{ f : I^d \to \mathcal{R} : f(x) = \sum_{i=1}^{n} w_i \phi(\| x - c_i \| / b_i) : \\ & \qquad n \in \mathcal{N}, \; c_i \in \mathcal{R}^d, \; b_i, \; w_i \in \mathcal{R}, \; b_i > 0 \}.\end{aligned}$$

The most popular radial function currently used in applications is the Gaussian $\gamma(t) = \exp(-t^2)$ (see [3], [7]). By $\mathcal{F}_u(\phi, \| \, . \, \|)$ we denote the set of functions computable by RBF networks with a uniform width, i.e.

$$\begin{aligned}\mathcal{F}_u(\phi, \| \, . \, \|) &= \{ f : I^d \to \mathcal{R} : f(x) = \sum_{i=1}^{m} w_i \phi(\| x - c_i \| / b) : \\ & \qquad m \in \mathcal{N}, \; c_i \in \mathcal{R}^d, \; b, \; w_i \in \mathcal{R}, \; b > 0 \}.\end{aligned}$$

The property of a class of feedforward networks to approximate general functions arbitrarily well can be described succintly using topology. Let U be a class of functions, T its subset, and ρ a metrics on U. The class T is said to have the *universal approximation property* with respect to (U, ρ) if it is dense in U with respect to the topology induced by ρ.

4 Approximation by Convolutions

The approximation of functions by convolutions with various kernel functions with a "peak" is a classical method. Weierstrass in 1885 used convolutions with Gaussian functions $\gamma_\delta(x) = \exp(-x^2/\delta)/\delta$ for the proof of his famous theorem on uniform approximation by polynomials. He approximated an arbitrary continuous function f uniformly on compact subsets of $\mathcal{R}$ by

$$f(x) = \lim_{\delta \to 0} f * \gamma_\delta / \sqrt{\pi} \tag{1}$$

By the standard technique generalizing Weierstrass' formula (1), one can approximate continuous functions by sequences of convolutions $f * \phi_n$, where functions $\{\phi_n,\ n \in \mathcal{N}\}$ are constructed from a non-zero integrable function ϕ by normalizing and "sharpening", i.e. putting $\phi_n(t) = n^d \phi(nt)$. Approximating a convolution by an appropriate Riemann sum, we proved in [4] the following theorem.

Theorem 1 *For every positive integer d and for every continuous function $\phi : \mathcal{R} \rightarrow \mathcal{R}_+$ with finite non-zero integral and for every norm $\| \, . \, \|$ on $\mathcal{R}^d$, $\mathcal{F}_u(\phi, \| \, . \, \|)$ is dense in $\mathcal{C}(I^d)$.*

In other words, the class of single hidden layer RBF networks with uniform width has the universal approximation property.

5 KBF networks

There are many classical sequences of kernel functions (like Dirichlet's kernel, see below) that are not derived from one function by dilation (multiplying the argument by n, as in case of RBF). To introduce general kernel functions into neural networks, in [4], we defined *kernel basis function* (*KBF*) *units.*

A KBF unit with d inputs computes a function $\mathcal{R}^d \rightarrow \mathcal{R}$ of the form $k_n(\| x - c \|)$, where $\{k_n : \mathcal{R} \rightarrow \mathcal{R}\}$ is a sequence of functions}, $\| \, . \, \|$ is a norm on $\mathcal{R}^d$, and $c \in \mathcal{R}^d$, $n \in \mathcal{N}$ are parameters. We call n *sharpness.*

A *kernel basis function (KBF) network* is a neural network with a single linear output unit, one hidden layer with KBF units with the same sequence of functions $\{k_n, n \in \mathcal{N}\}$ and the same norm $\| \, . \, \|$ on $\mathcal{R}^d$, and d inputs.

By $\mathcal{K}(\{k_n,\ n \in \mathcal{N}\}, \| \, . \, \|)$ we denote the set of functions computable by KBF networks with $\{k_n,\ n \in \mathcal{N}\}$ and $\| \, . \, \|$ with any number of hidden units. So

$$\begin{aligned} \mathcal{K}(\{k_n, n \in \mathcal{N}\}, \| \, . \, \|) \;&=\; \{f : I^d \rightarrow \mathcal{R} : f(x) = \sum_{i=1}^{m} w_i k_{n_i}(\| x - c_i \|), \\ &\qquad m,\ \ n_i \in \mathcal{N},\ \ c_i \in \mathcal{R}^d,\ \ w_i \in \mathcal{R}\}. \end{aligned}$$

By $\mathcal{K}_u(\{k_n,\ n \in \mathcal{N}\}, \| \, . \, \|)$ we denote the set of functions computable by KBF networks with the same ϕ_n for all units in the hidden layer, i.e.

$$\begin{aligned} \mathcal{K}_u(\{k_n\}, \| \, . \, \|) \;&=\; \{f : I^d \rightarrow \mathcal{R} : f(x) = \sum_{i=1}^{m} w_i k_n(\| x - c_i \|) : \\ &\qquad m,\ \ n \in \mathcal{N},\ \ c_i \in \mathcal{R}^d,\ \ w_i \in \mathcal{R}\}. \end{aligned}$$

Similarly, we obtained in [4] universal approximation property for quite general KBF networks.

Theorem 2 *For every positive integer d and for every sequence of continuous functions $\{k_n : \mathcal{R} \to \mathcal{R}_+,\ n \in \mathcal{N}\}$ and for every norm $\| \, . \, \|$ on $\mathcal{R}^d$ satisfying for every $n \in \mathcal{N}$ and every $x \in \mathcal{R}^d$ $\int\limits_{\mathcal{R}^d} k_n(\| x - y \|)dy = 1$ and for every $\delta > 0$ and every $x \in \mathcal{R}^d$ $\lim\limits_{n\to\infty} \int\limits_{J_\delta(x)} k_n(\| x - y \|)dy = 0$, where $J_\delta(x) = \{y |\ y \in \mathcal{R}^d,\ \| x - y \| \geq \ \delta\}$; the class $\mathcal{K}_u(\{k_n,\ n \in \mathcal{N}\},\ \| \, . \, \|)$ is dense in $\mathcal{C}(I^d)$.*

Note that all of the following classical kernels satisfy the assumptions of Theorem 2 and so KBF networks with any of these kernels are powerful enough to approximate continuous functions (of course, to achieve arbitrary accuracy, one must increase the number of hidden units).

Féjer kernel	$k_n(x) = [\sin nx/(n \cdot \sin x)]^2$
Dirichlet kernel	$k_n(x) = [\sin(n - 1/2)x/(2n \sin(x/2)]$
Jackson kernel	$k_n(x) = [\sin nx/(n \cdot \sin x)]^4$
Abel-Poisson kernel	$k_n(x) = 1/[1 + (nx)^2]$
Weierstrass kernel	$k_n(x) = e^{-nx^2}$
Landau kernel	$k_n(x) = (1 - x^2)^n$

6 Some Estimates of the Error of Approximation

For some of the above mentioned convolution kernels upper bounds on convolution approximation are known. The following theorem derives estimate of the rate of approximation by KBF networks depending on the error of approximation $E(f, k_n) = \| f - f * k_n \|$ and modulus of continuity of f and k_n.

Theorem 3 *Let $a \in \mathcal{R}$, $A = [-a, a]$, $A^* = [-2a, 2a]$, $f : A \to \mathcal{R}$ be a continuous function, $E(f, k_n) = \|f(x) - \int_A f(t)k_n\|x - t\|dt\|_A$. Then for every $m \in \mathcal{N}$ there exists a KBF network with m hidden units computing a function $g \in \mathcal{K}_u(\{k_n\}, \|.\|)$ such that for every $x \in A$*

$$|f(x) - g(x)| \leq E(f, k_n) + 2a\|f\|_A \ \omega_{A^*}(k_n, \frac{2a}{m}) + \|k_n\|_{A^*} \ \omega_A(f, \frac{2a}{m}).$$

Proof: The proof can be found in [5].

We use this theorem to estimate the approximation error for the KBF networks based on Jackson kernel with inputs in the interval $[-\pi, \pi]$. Consider the following operator:

$$\int_{-\pi}^{\pi} f(t)L_n(x-t)dt = \int_{-\pi}^{\pi} f(x+t)L_n(t)dt, \tag{2}$$

where L_n is the *Jackson kernel*

$$L_n(t) = \lambda_n^{-1}\left(\frac{\sin(nt/2)}{\sin(t/2)}\right)^4, \quad \int_{-\pi}^{\pi} L_n(t)dt = 1,$$

where the last relation defines λ_n. It is proved in [6], p. 55 that $\lambda_n \approx n^3$.

It is convenient to normalize the operator (2) in such a way as to obtain a trigonometric polynomial of degree n. For this purpose, we put

$$K_n(t) = L_r(t), \qquad r = [\frac{n}{2}] + 1$$

The operator $J_n(x) = J_n(f,x) = \int_{-\pi}^{\pi} f(x+t)K_n(t)dt$ is called the *Jackson operator.*

Theorem 4 (Jackson) *There exists a constant M such that, for each function $f \in \mathcal{C}(A)$, where $A = [-\pi, \pi]$ and for every $n \in \mathcal{N}$, $|f(x) - J_n(x)| \leq M\omega_A(f, \frac{1}{n})$.*

Proof: The proof can be found for example in [6], p.56.

Theorem 5 *There exists a constant M such that for every $f \in \mathcal{C}(\mathcal{A})$, $A = [-\pi, \pi]$, for every n (sharpness of the Jackson kernel) and for every $m \in \mathcal{N}$ and a function g computable by a Jackson KBF network with m hidden units and with sharpness n such that for every $x \in A$*

$$|f(x) - g)x)| \ \leq \ M\omega_A(f, \frac{1}{r}) + 2\pi \ \|f\|_A \omega_{A^*}(L_r, \frac{2\pi}{m}) + \|L_r\|_{A^*} \ \omega_A(f, \frac{2\pi}{m}), \tag{3}$$

where $r = [\frac{n}{2}] + 1$ and $A^* = [-2\pi, 2\pi]$.

Proof: From Theorems 3 and 4, where $E(f, k_n) \leq M\omega_P(f, \frac{1}{n})$. □

Our following estimate of the approximation error is made for classes of Lipschitz functions. Recall here the definition.

We say that $f \in Lip\alpha, 0 < \alpha \leq 1$, if there exists a constant C such that

$$|f(x) - f(y)| \leq C|x-y|^\alpha.$$

Lemma 6 *If* $M = \int_R |x|^\alpha |k_n(x)| dx < \infty, 0 < \alpha \leq 1$ *then* $f \in Lip_\alpha$ *implies*

$$|f * k_n - f| \leq \frac{CM}{n^\alpha}.$$

Proof: The proof can be found in [9], p.21.

Theorem 7 *Let* $a \in \mathcal{R}$, $A = [-a, a]$, $A^* = [-2a, 2a]$, $f : A \to \mathcal{R}$ *be a continuous function. Let* $f \in Lip_\alpha, 0 < \alpha \leq 1$ *and let* $M = \int_R |x|^\alpha |k_n(x)| dx < \infty$ *for a kernel function* k_n. *Then for every* $m \in \mathcal{N}$ *there exists a KBF network with* m *hidden units computing a function* $g \in \mathcal{K}_u(\{k_n\}, \|.\|)$ *such that for every* $x \in A$

$$|f(x) - g(x)| \leq \frac{CM}{n^\alpha} + 2a\|f\|_A \ \omega_{A^*}(k_n, \frac{2a}{m}) + \|k_n\|_{A^*} \ \omega_A(f, \frac{2a}{m}).$$

Proof: Corollary of theorem 5 and lemma 6.

Bibliography

[1] D.S. Broomhead, D. Lowe: Multivariable functional interpolation and adaptive networks. Complex Systems 2, 321-355 (1988).

[2] F. Girosi, T. Poggio: Networks and the best approximation property. Biological Cybernetics 63, 169-176 (1990).

[3] E.J. Hartman, J.D. Keeler, J.M. Kowalski: Layered neural networks with gaussian hidden units as universal approximations. Neural Computation 2, 210-215 (1990).

[4] V. Kůrková, K. Hlaváčková: Uniform approximation by KBF networks, Proceedings of NEURONET'93, Prague, pp. 1-7 (1993).

[5] V. Kůrková, K. Hlaváčková: Approximation of continuous functions by RBF and KBF networks, Proceedings of ESANN'94, Brussels, pp. 167-174 (1994).

[6] G.G. Lorentz: Approximation of functions, Holt, Rinehart and Winston, New York (1966).

[7] J. Moody, Ch.J. Darken: Learning with localized receptive fields. Proceedings of the 1988 Connectionist Models Summer School, San Mateo, CA (1989).

[8] J. Park, I.W. Sandberg: Approximation and radial-basis-function networks. Neural Computation 5, 305-316 (1993).

[9] H.S. Shapiro: Smoothing and approximation of functions, Van Nostrand Reinhold Mathematical Studies, New York (1969).