ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 167-174

Approximation of Continuous Functions by RBF and KBF Networks

Věra Kůrková, Kateřina Hlaváčková[1]

Institute of Computer Science,
Academy of Sciences of the Czech Republic,
Prague, Czech Republic
vera@uivt.cas.cz, katka@uivt.cas.cz

Abstract. We study approximation of continuous functions by networks with kernel basis function (KBF) units based on classical convolution kernels. We derive estimates of the error of approximation as a function of the number of hidden units.

1 Introduction

Radial basis function (RBF) networks arouse 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 [8]. Theoretically approximation properties of RBF networks with Gaussian radial function were studied by Girosi and Poggio [3] and by Hartman et al. [4] and for more general radial functions by Park and Sandberg [9, 10].

In contrast to the perceptron-type networks where most proofs of the universal approximation property are based on Stone-Weierstrass' theorem, in the case of RBF networks, we prove the universal approximation property using lemmas related to the properties of convolution. In [6], we showed how 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 build on these results to derive 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.

Section 2 contains preliminaries concerning RBF networks and approximation of functions. In Section 3, we review use of convolutions as a tool for study of approximation capabilities of RBF networks. Section 4 introduces a more

[1]This work was suppported by GACR under grant 201/93/0427.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 167-174

general concept of KBF networks and gives examples of KBF networks based on kernel functions arising in classical analysis. In section 5, we derive estimates of rates of approximation by KBF networks based on approximation by convolutions.

2 RBF networks

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 $\mathcal{R}^d$, $\mathcal{C}(A)$ denotes the set of all real-valued continuous functions on A with the uniform norm and corresponding topology.

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, $\| \cdot \|$ 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 $\| \cdot \|$ on $\mathcal{R}^d$, and d inputs.

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

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

The most popular radial function currently used in applications is the Gaussian $\gamma(t) = \exp(-t^2)$ (see [4], [8]).

By $\mathcal{F}_u(\phi, \| \cdot \|)$ we denote the set of functions computable by RBF networks with a uniform width, i.e.

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

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 167-174

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.

3 Approximation by Convolutions

A **convolution** of two functions $f, g : \mathcal{R}^d \to \mathcal{R}$ is $f * g = \int_{\mathcal{R}^d} f(x)g(x-y)dy$.

The approximation of functions by convolutions with various kernel functions with a "peak" is a classical method. Weierstrass in 1885 used convolutions with Gaussians $\gamma_\delta(x) = \exp(-x^2/\delta)/\delta$ 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

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

To generalize this approach, we need the following lemma which is a straightforward modification of the classical theorem on approximation by convolutions (see, e.g. [11]).

Lemma 3.1 *Let d be a positive integer, $\| \cdot \|$ a norm in $\mathcal{R}^d$ and $\{K_n : \mathcal{R}^{2d} \to \mathcal{R}, n \in \mathcal{N}\}$ be a sequence of functions such that*

(i) for every $n \in \mathcal{N}$ and every $x, y \in \mathcal{R}^d$ $K_n(x, y) \geq 0$;

(ii) for every $n \in \mathcal{N}$ and every $x \in \mathcal{R}^d$ $\int_{\mathcal{R}^d} K_n(x, y)dy = 1$;

(iii) for every $\delta > 0$ and every $x \in \mathcal{R}^d$ $\lim_{n\to\infty} \int_{J_\delta(x)} K_n(x, y)dy = 0$,

where $J_\delta(x) = \{y |\ y \in \mathcal{R}^d, \| x - y \| \geq \delta\}$;

Then for every bounded continuous function $f : \mathcal{R}^d \to \mathcal{R}$ and for every $x \in \mathcal{R}^d$

$$\lim_{n\to\infty} \int_{\mathcal{R}^d} f(y)K_n(x, y)dy = f(x).$$

If all K_n are continuous then the convergence is uniform on compacta.

A special case when $K_n(x, y) = k_n(x - y)$ applies to the approximation by convolutions. 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 such convolution by an appropriate Riemann sum, we obtained in [6] the following.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 167-174

Theorem 3.2 *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 $\| \cdot \|$ on $\mathcal{R}^d$, $\mathcal{F}_u(\phi, \| \cdot \|)$ is dense in $\mathcal{C}(I^d)$.*

Thus the class of single hidden layer RBF networks with uniform width has the universal approximation property.

4 KBF networks

To take advantage of Lemma 3.1 for deriving the universal approximation property for classes of feedforward networks, we do not need to restrict our attention to RBF networks only. 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). To introduce general kernel functions into neural networks, in [6], 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}, $\| \cdot \|$ 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 $\{\phi_n, n \in \mathcal{N}\}$ and the same norm $\| \cdot \|$ on $\mathcal{R}^d$, and d inputs.
By $\mathcal{K}(\{\phi_n, n \in \mathcal{N}\}, \| \cdot \|)$ we denote the set of functions computable by KBF networks with $\{\phi_n, n \in \mathcal{N}\}$ and $\| \cdot \|$ with any number of hidden units. So

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

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

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

As in Theorem 3.2, we obtained in [6] universal approximation property for quite general KBF networks.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 167-174

Theorem 4.1 *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 $\| \cdot \|$ on $\mathcal{R}^d$ satisfying for every $n \in \mathcal{N}$ and every $x \in \mathcal{R}^d$ $\int_{\mathcal{R}^d} k_n(\| x - y \|)dy = 1$ and for every $\delta > 0$ and every $x \in \mathcal{R}^d$ $\lim_{n\to\infty} \int_{J_\delta(x)} k_n(\| x - y \|)dy = 0$, where $J_\delta(x) = \{y |\ y \in \mathcal{R}^d,\ \| x - y \| \geq \delta\}$; the class $\mathcal{K}_u(\{k_n,\ n \in \mathcal{N}\}, \| \cdot \|)$ is dense in $\mathcal{C}(I^d)$.*

All of the following classical kernels satisfy the assumptions of Theorem 4.1 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	$k_n(x) = (1 - x^2)^n$

5 An Estimate of the Error of Approximation

Let $f : \mathcal{R} \to \mathcal{R}$ be a continuous function, $A \subseteq R$, put $\|f\|_A = \sup_{x\in A} |f(x)|$. $\omega_A(f, h) = \sup_{\substack{|x_1 - x_2| \leq h, \\ x_1, x_2 \in A}} |f(x_1) - f(x_2)|$ **is modulus of continuity of f on A.**

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 $E(f, k_n) = |f \doteq f * k_n|$ and modulus of continuity of f and k_n.

Theorem 5.1 *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: By assumption, $|f(x) - \int_A f(x)k_n(x - t)dt| \leq E(f, k_n)$ for every $x \in A$. We estimate $\int_A f(t)k_n(x - t)dt$ by a Riemann sum s_m.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 167-174

$$|\int_A f(t)k_n(x-t)dt - s_m(x)| \leq 2a\omega_A(h_x(t), \frac{2a}{m}),$$

where $s_m(x) = \frac{1}{m}\sum_{i=1}^{m} h_x(t_i) = \frac{1}{m}\sum_{i=1}^{m} f(t_i)k_n(x-t_i)$, $t_i = -a+\frac{2ai}{m}$, $i = 0, 1, \ldots m$.
From the properties of modules of continuity we have

$$\omega_A(f(t)k_n(x-t), \frac{2a}{m}) \leq \|f\|_A\omega_A(x-t), \frac{2a}{m}) + \|k_n(x-t)\|_A\omega_A(f(t), \frac{2a}{m}),$$

where $x, t \in A$.

$$\begin{aligned}|\int_A f(x) - s_m(x)| &\leq \frac{E(f,k_n)}{2} + 2a\|f\|_A\omega_A(k_n(x-t), \frac{2a}{m}) + \\ &+ 2a\|k_n(x-t)\|_A \cdot \omega_A(f(t), \frac{2a}{m}).\end{aligned}$$

Since $\omega_A(k_n(x-t), \delta) \leq \omega_{A^*}(k_n(t), \delta)$ for every $\delta > 0$ and $x, t \in A$, we have

$$\begin{aligned}|\int_A f(x) - s_m(x)| &\leq \frac{E(f,k_n)}{2} + 2a\|f\|_A\omega_{A^*}(k_n, \frac{2a}{m}) + \\ &+ 2a\|k_n\|_{A^*}\omega_A(f, \frac{2a}{m}).\end{aligned}$$

Putting $g = s_m(x)$, we have $g \in K_u(\{k_n\}, \|.\|)$ and on A

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

We use this theorem to estimate the error of approximation 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 [7], p. 55 that $\lambda_n \approx n^3$.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 167-174

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

The proof can be found for example in [7], p.56.

Theorem 5.3 *There exists a constant M such that for every $f \in \mathcal{C}(A)$, $A = [-\pi, \pi]$, 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 $x \in A$*

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

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

Proof. From Theorems 5.1 and 5.2, where $E(f, k_n) = M\omega_P(f, \frac{1}{n})$. □

References

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

[2] R. Engelking: General topology. Warszawa: PWN (1977).

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

[4] 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).

[5] E. Hewitt, K.A. Ross: Abstract harmonic analysis. Moscow, Nauka (1975).

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

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

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

[9] J. Park, I.W. Sandberg: Universal approximation using radial-basis-function networks. Neural Computation 3, 246-257 (1991).

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

[11] J.R. Rice: The Approximation of Functions. Linear Theory, Volume 1, Addison-Wesley Publishing Company (1964).