Feed-Forward Neural Networks and Splines for Approximation of Functions

Hlaváčková, Kateřina
1995

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Synthesis of feed-forward neural networks using splines for approximation of functions

Kateřina Hlaváčková   Michel Verleysen

Technical report No. 651

1995

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
phone: (+4202) 6605 3461   fax: (+4202) 8585789
e-mail: katka@uivt.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Synthesis of feed-forward neural networks using splines for approximation of functions

Kateřina Hlaváčková[1] Michel Verleysen[2]

Technical report No. 651
1995

Abstract

When using neural networks with spline activation functions, the quality of approximation depends on the placement of knots of spline functions. This paper shows how to choose the number of equidistant knots in each subdivision of the space when an arbitrary initial division is given, in order to keep the approximation error under a predefined limit.

Keywords
Cubic spline, activation function, approximation error.

[1]Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, Czech Republic

[2]Microelectronics Laboratory, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
email: verleysen@dice.ucl.ac.be

1 Introduction

Approximation of functions is a classical problem in engineering tasks being solved by many mathematical approaches. Recently, a resurgence of interest in this area came from researchers in the field of neural networks, who developed new approaches or rediscovered conventional ones; advantages of approximation of functions by neural networks mainly consist in their simplified descriptions, usually in terms of algorithms or diagrams instead of complex mathematical equations. This simple description helps to work more easily in high dimensional spaces.

Among various "neural networks" algorithms used in approximation of functions there are methods based on "local" activation functions. Radial-Basis Functions (RBF) networks are widely used; methods based on splines seem to be less studied by the researches dealing with neural networks, while their approximation properties are justified by numerous convergence theorems from 60s and 70s.

This paper deals with approximation of functions by cubic spline networks. Bounds on the approximation of functions by cubic splines are used to evaluate the number of knots necessary for partition of the space to keep the approximation under a predefined limit. The results are derived in the one and two-dimensional cases, where spline approximation is mostly used.

2 Neural Network with cubic spline activation functions

2.1 One-dimensional case

We start our presentation of cubic splines interpolation by the one-dimensional case, according to [1].

Consider $I = [0, 1]$ and a real function $f : I \to R$. Let Δ be a division of the $I = [0, 1]$ given by $\Delta \equiv \{0 = \gamma_0 < \gamma_1 < ... < \gamma_N = 1\}$.

Given Δ, let the *space of cubic splines with respect to* Δ, $S(\Delta)$, be the vector space of all twice continuously differentiable, piecewise cubic polynomials on I with respect to Δ, i.e.

$$S(\Delta) \equiv \{p(x) \in C^2(I)\} \tag{2.1}$$

so that $p(x)$ is a cubic polynomial on each subinterval $[x_i, x_{i+1}], 0 \leq i \leq N$, defined by Δ.

In this paper we will use words cubic spline and spline interchangably.

Every cubic spline may be represented in terms of the *basis functions* $\{h_i(x), h_i^1(x)\}_{i=0}^{N+1}$:

$$s(x) = \sum_{i=0}^{N+1} (s(x_i)h_i(x) + s'(x_i)h_i^1(x)), \tag{2.2}$$

where $h_i(x)$ is the unique piecewise cubic polynomial on each subinterval defined by Δ in $C^1(I)$ such that $h_i(x_j) = \delta_{ij}, 0 \leq i, j \leq N + 1$, and $h_i'(x_j) = 0, 0 \leq i, j \leq N + 1$, and $h_i^1(x)$ is the unique piecewise cubic polynomial on each subinterval defined by Δ in $C^1(I)$ such that $h_i^1(x_j) = 0, 0 \le i, j \le N+1$, and $(h_i^1)'(x_j) = \delta_{ij}, 0 \le i, j \le N+1$.

This mapping $s(x)$ is "local" in the sense that if $x \in [x_i, x_{i+1}]$, $0 \le i \le N$, then $s(x)$ depends only on $s(x_i), s'(x_i), s(x_{i+1}), s'(x_{i+1})$.

Given $f \equiv (f_0, ..., f_{N+1}, f'_0, f'_{N+1}) \in R^{N+3}$, let $\hat{f}(x)$, the $S(\Delta)$*–interpolate of f* be the unique spline $s(x)$ in $S(\Delta)$ such that $s(x_i) = f_i$, $0 \le i \le N+1$ and $\hat{f}'(x_i) = f'_i$, $i = 0$ and $N+1$.

It can be also proven (see [1]) that, given the set $\Omega = \{f_0, \ldots, f_{N+1}, f'_0, f'_{N+1}\}$ of $N+3$ real values corresponding respectively to the values of the function f at $x_i, 0 \le i \le N+1$, and to the value of its derivative f' at x_0 and at x_{N+1}, $\hat{f}$, the $S(\Delta)$ interpolate of f such that $\hat{f}(x_i) = f_i, 0 \le i \le N+1$ and $\hat{f}'(x_i) = f'_i, i = 0$ and $N+1$, is uniquely determined, and the above definition is correct. For a standard procedure to calculate coefficients $s(x_i)$ and $s'(x_i)$, see [1].

Moreover, if parameters f'_0 and f'_{N+1} are not known, they can be approximated by the local cubic Lagrange interpolating polynomials at both ends of the interval I, before determining the approximation $\hat{f}(x)$.

Finally, we must mention that it is possible to determine a basis for $S(\Delta)$, namely the "cardinal splines" $\{C_i(x)\}_{i=0}^{N+3}$, defined by

$C_j(x_i) = \delta_{ij}, C'_j(0) = C'_j(1) = 0$, for $0 \le i, j \le N+1$, and by $C_{N+2}(x_i) = C_{N+3}(x_i) = 0, 0 \le i, j \le N+1$, $C'_{N+2}(0) = C'_{N+3}(1) = 1$ and $C'_{N+2}(1) = C_{N+3}{}'(0) = 0$.

In this case, the approximation $\hat{f}(x)$ can be written as:

$$\hat{f}(x) = \sum_{i=0}^{N+1} f_i C_i(x) + f'_0 C_{N+2}(x) + f'_{N+1} C_{N+3}(x). \tag{2.3}$$

Note that in this case the interpolation mapping $\hat{f}$ is no more local, i.e. $\hat{f}(x)$ depends on all parameters $f_i, 0 \le i \le N+1, f'_0$ and f'_{N+1}.

We can construct a corresponding spline neural network having cardinal splines as activation functions. *** Michel, could you draw it here? The figure of the network

with one-dimensional input, $N+3$ units in the hidden layer and network weights f_i and f_0 and f_{N+1}.

2.2 Two-dimensional case

Consider $U \equiv [0,1] \times [0,1]$ and a real function $f : U \to R$. In the agreement with [1], let $\rho = \Delta \bigotimes \Delta_y$ be a rectangular grid given by $\Delta \equiv \{0 = \gamma_0 < \gamma_1 < ... < \gamma_N \equiv 1\}$ and $\Delta_y = \{0 = \mu_0 < \mu_1 < ... < \mu_M = 1\}$ are some given divisions of I.

Again, according to [1] let $S(\rho)$ be the $(N+4)(M+4)$-dimensional vector space of all functions of the form

$$s(x,y) = \sum_{i=0}^{N+3} \sum_{j=0}^{M+3} \beta_{ij} C_i(x) C_j(y), \tag{2.4}$$

where $C_i(x), C_j(x)$ are cardinal splines of $S(\Delta)$. $S(\rho)$ is the space of piecewise bicubic polynomials in U with respect to ρ.

Now we can uniquely define the interpolation $\hat{f}(x,y)$ of f in $S(\rho)$ by:

$$\begin{aligned}
\hat{f}(x,y) &= \sum_{i=0}^{N+1}\sum_{j=0}^{M+1} f_{i,j}C_i(x)C_j(y) \\
&+ \sum_{j=0}^{M+1}(f_{0,j}^{1,0}C_{N+2}(x) + f_{N+1,j}^{1,0}C_{N+3}(x))C_j(y) \\
&+ \sum_{i=0}^{N+1}(f_{i,0}^{0,1}C_{M+2}(y) + f_{i,M+1}^{0,1}C_{M+3}(y))C_i(x) \\
&+ f_{0,0}^{1,1}C_{N+2}(x)C_{M+2}(y) + f_{0,M+1}^{1,1}C_{N+2}(x)C_{M+3}(y) \\
&+ f_{N+1,0}^{1,1}C_{N+3}(x)C_{M+2}(y) + f_{N+1,M+1}^{1,1}C_{N+3}(x)C_{M+3}(y), \qquad (2.5)
\end{aligned}$$

where $f_{i,j} \equiv f(x_i, y_j)$, $f_{0,j}^{1,0} \equiv \partial f(0, y_j)/\partial x$,
$f_{N+1,j}^{1,0} \equiv \partial f(1, y_j)/\partial x$, $f_{i,0}^{0,1} \equiv \partial f(x_i, 0)/\partial y$,
$f_{i,M+1}^{0,1} \equiv \partial f(x_i, 1)/\partial y$, $f_{0,0}^{1,1} \equiv \partial^2 f(0,0)/\partial x\partial y$,
$f_{0,M+1}^{1,1} \equiv \partial^2 f(0,1)/\partial x\partial y$, $f_{N+1,0}^{1,1} \equiv \partial^2 f(1,0)/\partial x\partial y$,
$f_{N+1,M+1}^{1,1} \equiv \partial^2 f(1,1)/\partial x\partial y$, for all $0 \leq i \leq N+1$ and $0 \leq j \leq M+1$.

*** Michel, could you draw a figure of 2-dim spline network with cardinal splines as activation functions?

3 Approximation with uniform-bounded error

In this section, we deal with the problem of finding an adequate number of knots, i.e. the number of points x_i and y_j, to ensure the approximation error under a predefined bound. We synthesize our results on the base of error bound estimations known from [1]. Both in one and two-dimensional spaces, the approximation error is bounded by a function depending on the maximum value of the fourth derivative of function f in the considered interval. We realized that the bounds on corresponding fourth derivative may vary from one location of the space to another, and that it is convenient to split the space into subdivisions, among which the distances between knots can thus be different. We deal only with equally-spaced knots into each subdivision of the space.

3.1 One-dimensional case

Consider the interval $I = [0,1]$ of function f divided into R intervals $[\gamma_i, \gamma_{i+1}], 0 \leq i \leq N, \gamma_0 = 0, \gamma_{N+1} = 1$ given by Δ.

Our concern is to find a distance h_i between equally-spaced knots in each interval $[\gamma_i, \gamma_{i+1}]$, such that the approximation error between f and $\hat{f}$ is under a predefined limit ϵ for all points in $[\gamma_i, \gamma_{i+1}]$:

$$\|f - \hat{f}\|_\infty \leq \epsilon, \tag{3.1}$$

where $\|f - \hat{f}\|_\infty = \sup_{x \in I} \|f(x) - \hat{f}(x)\|$ is the uniform norm of $f - \hat{f}$.

Consider $PC^{4,\infty}(I)$, the space of all functions on I such that:

1. f is three times continuously differentiable,
2. there exist γ_i, $i = 0 \leq i \leq N$ with $0 = \gamma_0 < \gamma_1 < \ldots < \gamma_N = 1$ such that on each open subinterval (γ_i, γ_{i+1}), $0 \leq i \leq N-1$, $f^{(3)}$ is continuously differentiable,
3. $\| f^{(4)}(x) \|_\infty = \max_{0 \leq i \leq N} \sup_{x \in (\gamma_i, \gamma_{i+1})} |f^{(4)}(x)| < \infty$ for $x \in I$.

We will need the following result (see [1]):

Lemma 1 *If $f \in PC^{4,\infty}(I)$, then*

$$\|f - \hat{f}\|_\infty \leq \frac{5}{384}\, h^4 \,\|f^{(4)}\|_\infty, \tag{3.2}$$

where $h \equiv \max_{0 \leq i \leq N}(\gamma_{i+1} - \gamma_i)$.

For a proof see [1] p.57.

Now let us present our main theorem of this chapter.

Theorem 2 *Let $\Delta = \{0 = \gamma_0 < \ldots < \gamma_N = 1\}$ be a given division of I and let $L_i = \sup_{x \in (x_i, x_{i+1})} |f^{(4)}(x)|$ for all $i = 0, \ldots, N$, $C = \frac{5}{384}$. Then for every $\epsilon > 0$ there exists a neural network with spline activation functions given on the subdivisions of $[\gamma_i, \gamma_{i+1}]$, $i = 0, \ldots, N$ consisting of $M_i + 1$ equidistant knots $\gamma_{i_0}, \ldots, \gamma_{i_{M_i}}$ on each $[\gamma_i, \gamma_{i+1}]$realizing a function $\hat{f}$ so that for every $x \in I$*

$$|f(x) - \hat{f}(x)| \leq \epsilon. \tag{3.3}$$

This M_i can be found in the way

$$M_i = \lceil \frac{\gamma_{i+1} - \gamma_i}{h_i} \rceil, \tag{3.4}$$

where $h_i = \sqrt[4]{\frac{\epsilon}{CL_i}}$. If $L_i \neq L_j$ then $M_i \neq M_j$. ($\lceil a \rceil$ means the smallest integer greater than a).

Proof: If the right side of the equation in the lemma 1 is less than ϵ, the approximation error will be under the predefined limit.

Applying this condition to each interval $[\gamma_i, \gamma_{i+1}]$, we have:

$$\frac{5}{384} h_i^4, \|f^{(4)}(x)\|_\infty \leq \epsilon, \tag{3.5}$$

or

$$h_i \leq \sqrt[4]{\frac{\epsilon}{CL_i}}, \tag{3.6}$$

where

$$C = \frac{5}{384} \text{ and } L_i = \|f^{(4)}(x)\|_\infty \tag{3.7}$$

on the interval $[\gamma_i, \gamma_{i+1}]$, for each $0 \leq i \leq N$.

Under the condition (11), the best choice for h_i is obviously the largest possible value, in order to decrease the number of necessary knots. However, to keep limits γ_i and γ_{i+1} of the interval as knots too and to have an integer value for $(\gamma_{i+1} - \gamma_i)/h_i$, we will choose the number of interior knots M_i (not including γ_i and γ_{i+1}) to be equal to:

$$M_i = \lceil \frac{\gamma_{i+1} - \gamma_i}{\sqrt[4]{\frac{\epsilon}{CL_i}}} \rceil, \tag{3.8}$$

and consequently

$$h_i = \frac{\gamma_{i+1} - \gamma_i}{M_i + 1}. \tag{3.9}$$

3.2 Two-dimensional case

A similar result can be obtained in the two-dimensional case. We now divide the space $U = [0,1] \times [0,1]$ (the domain) of function f by a rectangular grid ρ, where $\rho = \Delta \bigotimes \Delta_y$ be a rectangular grid given by $\Delta \equiv \{0 = \gamma_0 < \gamma_1 < ... < \gamma_N = 1\}$ and $\Delta_y \equiv \{0 = \mu_0 < \mu_1 < ... < \mu_M = 1\}$, which are given divisions of I. In each rectangle $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$, we want to find distances h_{ij} between equally spaced knots for both coordinates so that the approximation error is under the limit ϵ for all points in U in the rectangle.

Similarly as in the section for one-dimensional problem, denote

$PC^{4,\infty}(U)$, the space of all functions on I such that:

1. $f(x, y)$ is three times continuously differentiable, i.e. $\frac{\partial^l \partial^k f(x,y)}{\partial x^l \partial y^k}$ exist and are continuous for all $0 \leq l + k \leq 3$,

2. there exist γ_i, $0 \leq i \leq R$ and μ_j $0 \leq j \leq S$ with $0 = \gamma_0 < \gamma_1 < \ldots < \gamma_R = 1$ and $0 = \mu_0 < \ldots < \mu_s = 1$ such that on each open subrectangle $(\gamma_i, \gamma_{i+1}) \times (\mu_j, \mu_{j+1})$, $0 \leq i \leq R - 1$, $0 \leq j \leq S - 1$ $\frac{\partial^l \partial^k f(x,y)}{\partial x^l \partial y^k}$ are continuously differentiable for $0 \leq l + k \leq 3$, and

3. for all $0 \leq l+k \leq 3$

$$\| \frac{\partial^l \partial^k f(x,y)}{\partial x^l \partial y^k} \|_\infty = \max_{\substack{0\leq i\leq R\\ 0\leq j\leq S}} \sup_{(x,y)\in(\gamma_i,\gamma_{i+1})\times(\mu_j,\mu_{j+1})} |\frac{\partial^l \partial^k f(x,y)}{\partial x^l \partial y^k}| < \infty$$

for $(x,y) \in U$.

Under the similar conditions on the function f as in the one-dimensional case, we have (see [1]):

Lemma 3 *If $f \in PC^{4,\infty}(U)$, then*

$$\|f - \hat{f}\|_\infty \leq \hat{\rho}^4 \left(\frac{5}{384} \|\frac{\partial^4 f}{\partial x^4}\|_\infty + \frac{4}{9} \|\frac{\partial^4 f}{\partial x^2 \partial y^2}\|_\infty + \frac{5}{384} \|\frac{\partial^4 f}{\partial y^4}\|_\infty\right), \tag{3.10}$$

where $\hat{\rho} = max\{h,k\}$ and h,k are the distances between knots in the x and y-coordinates respectively.

For a proof see [1] p.60.
Applying this condition to each subrectangle of the space $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$, we have our main theorem of this chapter.

Theorem 4 *Let the space $U = [0,1] \times [0,1]$ (the domain) of function f be divided by a rectangular grid ρ, where $\rho = \Delta \bigotimes \Delta_y$ be a rectangular grid given by $\Delta \equiv \{0 = \gamma_0 < \gamma_1 < ... < \gamma_N \equiv 1\}$ and $\Delta_y \equiv \{0 = \mu_0 < \mu_1 < ... < \mu_M = 1\}$ are given divisions of I. Let*

$$L_{ij} = \sup\left(\frac{5}{384} \|\frac{\partial^4 f}{\partial x^4}\|_\infty + \frac{4}{9} \|\frac{\partial^4 f}{\partial x^2 \partial y^2}\|_\infty + \frac{5}{384} \|\frac{\partial^4 f}{\partial y^4}\|_\infty\right) \tag{3.11}$$

on the rectangle $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$.

Then for every $\epsilon > 0$ there exists a neural network with spline activation functions given on the subdivisions of equidistant knots on $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$ $i = 0, \ldots R, j = 0, \ldots, S$ consisting of $M_{ij} = (R_i + 1) \times (S_j + 1)$ equidistant knots realizing a function $\hat{f}$ so that for every $(x,y) \in U$

$$|f(x,y) - \hat{f}(x,y)| \leq \epsilon. \tag{3.12}$$

These M_{ij} can be found as:

$$R_i = \frac{|\gamma_{i+1} - \gamma_i|}{h_i}, where h_i = \gamma_{i+1} - \gamma_i \tag{3.13}$$

$$S_j = \frac{|\mu_{j+1} - \mu_j|}{k_j}, where k_j = \mu_{j+1} - \mu_j \tag{3.14}$$

and where $\hat{\rho}_{i,j} = \sqrt[4]{\frac{\epsilon}{L_{i,j}}}$ and $h_{ij} = \frac{h_i}{R_{ij}+1}$, $k_{ij} = \frac{k_j}{S_{ij}+1}$ and h_i, $k_j \leq \hat{\rho}_{i,j}$. The best solution (in terms of knots $M_{i,j}$): $h_i = k_j = \rho_{i,j}$.

Proof: Applying the condition of lemma 3 to each subrectangle of the space $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$, we have:

$$\overline{\rho}_{ij}^4 \left(\frac{5}{384} \| \frac{\partial^4 f}{\partial x^4} \|_\infty + \frac{4}{9} \| \frac{\partial^4 f}{\partial x^2 \partial y^2} \|_\infty + \frac{5}{384} \| \frac{\partial^4 f}{\partial y^4} \|_\infty \right) \leq \epsilon, \tag{3.15}$$

or

$$\hat{\rho}_{ij} \leq \sqrt[4]{\frac{\epsilon}{L_{ij}}}, \tag{3.16}$$

where $\hat{\rho}_{ij} = max\{h_{ij}, k_{ij}\}$ and

$$L_{ij} = \frac{5}{384} \| \frac{\partial^4 f}{\partial x^4} \|_\infty + \frac{4}{9} \| \frac{\partial^4 f}{\partial x^2 \partial y^2} \|_\infty + \frac{5}{384} \| \frac{\partial^4 f}{\partial y^4} \|_\infty \tag{3.17}$$

on the rectangle $[\gamma_i, \gamma_{i+1}] \times [\mu_j, \mu_{j+1}]$.

Under the condition (20), the best choice for h_{ij} and k_{ij} is obviously the largest possible values, in order to decrease the number of necessary knots. However, to keep knots on the boundaries of the rectangle and to have integer values for $(\gamma_{i+1} - \gamma_i)/h_{ij}$ and $(\mu_{j+1} - \mu_j)/k_{ij}$, we will choose the number of interior knots in the x-direction R_{ij} (not including γ_i and γ_{i+1}) and in the y-direction S_{ij} (not including μ_j and μ_{j+1}) to be equal to:

$$R_{ij} = \lceil \frac{\gamma_{i+1} - \gamma_i}{\sqrt[4]{\frac{\epsilon}{L_{ij}}}} \rceil, \tag{3.18}$$

$$S_{ij} = \lceil \frac{\mu_{j+1} - \mu_j}{\sqrt[4]{\frac{\epsilon}{L_{ij}}}} \rceil, M_{ij} = (R_{ij} + 1) \times (S_{ij} + 1) \tag{3.19}$$

and consequently

$$h_{ij} = \frac{\gamma_{i+1} - \gamma_i}{R_{ij} + 1}, \tag{3.20}$$

$$k_{ij} = \frac{\mu_{j+1} - \mu_j}{S_{ij} + 1}. \tag{3.21}$$

We see that the lattice of knots is regular inside each subrectangle of the space, but that it may no more be regular over the whole space U.

4 Discussion

Placement of knots in neural network approximation by superposition of local functions (RBF, splines) is a non-trivial problem, as well in one- as in mutil-dimensional spaces. The number of knots to be placed in a defined region of the space has also to be determined. In many situations, this number is fixed a priori, assuming some knowledge on the function to approximate, and some experience from the programmer to guess a number of knots which offers a compromise between the complexity of the algorithm and the accuracy of approximation.

By using results on the error of approximation in one- and two-dimensional spaces, this paper presents a method to determine a lower bound for the number of knots in each region of the partitioned space, assuming that a bound on the fourth derivative of the function to approximate is known.

Further studies could include the extension of the results to spaces whose dimension is greater than two, and the search for approximations of the bound of the fourth derivative of the function to approximate, given a set of data, but without any more information about the function.

Bibliography

[1] M.H. Schultz: Spline Analysis. Prentice–Hall, New Jersey (1973).

[2] L.L. Schumaker: Spline Functions:Basic Theory. John Wiley and Sons, New York (1981).