Neural Networks 67 (2015) 28–36

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Neural network operators: Constructive interpolation of multivariate functions

Danilo Costarelli

Department of Mathematics and Computer Science, University of Perugia, 1, Via Vanvitelli, 06123 Perugia, Italy

ARTICLE INFO

Article history:
Received 10 November 2014
Received in revised form 28 January 2015
Accepted 1 February 2015
Available online 8 February 2015

Keywords:
Sigmoidal functions
Neural networks operators
Multivariate interpolation
Multivariate approximation
Order of approximation
Irregular sampling scheme

ABSTRACT

In this paper, the interpolation of multivariate data by operators of the neural network type is proved. These operators can also be used to approximate continuous functions defined on a box-domain of $\mathbb{R}^d$. In order to show this fact, a uniform approximation theorem with order is proved. The rate of approximation is expressed in terms of the modulus of continuity of the functions being approximated. The above interpolation neural network operators are activated by suitable linear combinations of sigmoidal functions constructed by a procedure involving the well-known central B-spline. The implications of the present theory with the classical theories of neural networks and sampling operators are analyzed. Finally, several examples with graphical representations are provided.

1. Introduction

The function $h(\underline{x})$ implementing a neural network (NN) can be represented by:

$$h(\underline{x}) = \sum_{j=0}^{n} c_j \, \sigma(\underline{w}_j \cdot \underline{x} + \theta_j), \quad \underline{x} \in \mathbb{R}^d, \; d \in \mathbb{N}^+, \tag{I}$$

where $c_j \in \mathbb{R}$ are the coefficients, $\underline{w}_j \in \mathbb{R}^d$ are the weights and $\theta_j \in \mathbb{R}$ are thresholds of the NN, for every $j = 0, 1, \ldots, n$. The terms $\underline{w}_j \cdot \underline{x}$ denote the inner product in $\mathbb{R}^d$ between the two vectors $\underline{w}_j$ and $\underline{x}$, while the function $\sigma : \mathbb{R} \to \mathbb{R}$ is the *activation function* of the NN, see, e.g., Pinkus (1999). Typically, $\sigma(x)$ is a *sigmoidal function*, i.e., a measurable function satisfying the properties:

$$\lim_{x \to -\infty} \sigma(x) = 0, \quad \text{and} \quad \lim_{x \to +\infty} \sigma(x) = 1.$$

Examples of sigmoidal functions are $\sigma_\ell(x) := (1 + e^{-x})^{-1}$ and $\sigma_h(x) := (1/2)(\tanh x - 1)$, $x \in \mathbb{R}$, i.e., the well-known logistic and hyperbolic tangent function, see, e.g., Cybenko (1989).

In the last thirty years, NNs of the form in (I), activated by sigmoidal functions, have been successfully applied in Approximation Theory, in order to approximate functions of one or several variables, see, e.g., Cybenko (1989), Barron (1993), Hahm and Hong (2002), Lewicki and Marino (2003), Costarelli and Spigler (2013a) and Costarelli (2014a). The most common approach used to study approximation by NNs was the non-constructive one, see, e.g., Cybenko (1989). Recently, constructive approximation results have been proved in both univariate and multivariate settings, see, e.g., Costarelli (2014a), where some of them are summarized.

For instance, in Cheney et al. (1992, 1993) the idea of convolution kernel from a sigmoidal function is used. This approach is quite difficult due to the nature of the problem and is based on some results related to the theory of *ridge functions.* Another possibility to obtain constructive NNs activated by sigmoidal functions was described in Lenze (1992), where a convolution approach is considered again, but for Lebesgue–Stieltjes integrals. Moreover, we mentioned the approach proposed by Barron (1993), where multivariate functions satisfying a suitable condition involving the Fourier transform of f were approximated in L^2-norm. Further, in the paper (Costarelli and Spigler, 2013a) a constructive L^p-version of Cybenko's approximation theorem (Cybenko, 1989) is provided, in both univariate and multivariate settings. Finally, in Costarelli and Spigler (2015) the exponential convergence of certain NNs constructed by sigmoidal function is proved by an approach based on multiresolution approximation and the corresponding wavelet scaling functions. For other results concerning NNs and their applications to approximation problems, see, e.g., Costarelli and Spigler (2013b, 2014a), Gripenberg (2003), Ismailov (2014), Kainen and Kurková (2009), Kurková (2012), Makovoz (1996, 1998) and Maiorov (2006).

The approaches used to obtain the results quoted and described above present some difficulties and are not obvious. The theory of

E-mail address: danilo.costarelli@unipg.it.

http://dx.doi.org/10.1016/j.neunet.2015.02.002
0893-6080/

NN operators has been introduced in order to study a constructive approximation process by NNs, which was more intuitive than those proposed in previous papers. Moreover, results for NN operators can be proved by using techniques typically used in Operator Theory.

The NN operators N_n^σ were introduced by G.A. Anastassiou in Anastassiou (1997), where the results originally proved in Cardaliaguet and Euvrard (1992) by P. Cardalignet and G. Euvrard have been extended. The above NN operators were defined by:

$$N_n^\sigma(f, \underline{x}) := \frac{\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor} f\left(\underline{k}/n\right) \Psi_\sigma(n\underline{x} - \underline{k})}{\sum_{k_1=\lceil na_1\rceil}^{\lfloor nb_1\rfloor} \cdots \sum_{k_s=\lceil na_s\rceil}^{\lfloor nb_s\rfloor} \Psi_\sigma(n\underline{x} - \underline{k})}, \quad \underline{x} \in \mathcal{R} \qquad \text{(II)}$$

where $\mathcal{R} := [a_1, b_1] \times \cdots \times [a_d, b_d] \subset \mathbb{R}^d$, $n \in \mathbb{N}^+$, $f : \mathcal{R} \to \mathbb{R}$ be a given bounded function, and $\Psi_\sigma(\underline{x}) := \phi_\sigma(x_1) \cdots \phi_\sigma(x_d)$, $\underline{x} \in \mathbb{R}^d$, is the multivariate density functions defined through the product of d one-dimensional density functions $\phi_\sigma(x) := \frac{1}{2}[\sigma(x+1) - \sigma(x-1)]$, $x \in \mathbb{R}$.

In Anastassiou (2011a,b,c, 2012), Anastassiou studied neural network operators N_n^σ both in univariate and multivariate settings, for the special cases of logistic and hyperbolic tangent sigmoidal activation functions, i.e., with $\sigma(x) = \sigma_\ell(x)$ and $\sigma(x) = \sigma_h(x)$. Approximation results involving continuous functions defined on bounded domains were proved therein for the family $(N_n^\sigma)_{n\in\mathbb{N}^+}$, together with estimates concerning the order of approximation.

Subsequently, other results concerning the NN operators N_n^σ have been obtained in Cao and Chen (2009, 2012) and Costarelli and Spigler (2013c,d, 2014b). In particular, in Costarelli and Spigler (2013c,d) the approximation results proved in Anastassiou (2011a,b,c, 2012) have been extended, in order to consider NN operators activated by any sigmoidal function $\sigma(x)$ belonging to a suitable class which contains also $\sigma_\ell(x)$ and $\sigma_h(x)$. Moreover, the results concerning the order of approximation have been improved therein.

Further, in Costarelli and Spigler (2014b) NN operators of the Kantorovich type have been introduced in order to study the problem of approximating L^p functions, for $1 \le p < +\infty$.

An important task for NNs is the capability to interpolate any given data. This problem is strictly related with the theory of training neural networks. Indeed, NNs which are able to interpolate data belonging to a suitable training set can be used to reproduce exactly certain values, without errors.

The above problem has been already studied by many authors (see e.g. Dasgupta and Shristava, 1990; Llanas and Sainz, 2006; Sontag, 1992) by means of analytical or algebraic approaches. By the word *analytical* we refer to results proved by non-constructive arguments, while by the word *algebraic* we refer to proofs in which the coefficients of interpolating NNs are obtained by solving suitable linear algebraic systems.

Concerning the theory of NN operators, in general the N_n^σ *does not interpolate*, i.e., $N_n^\sigma(f, \underline{k}/n) \neq f(\underline{k}/n)$, for any given bounded function $f : \mathcal{R} \to \mathbb{R}$, $n \in \mathbb{N}^+$ and $\underline{k} \in \mathbb{Z}^d$. In Costarelli (2014b), *interpolation NN operators* have been introduced in *one-dimensional setting*, by a substantial modification in the definition of N_n^σ, when $d = 1$. The changes that have been done in the one-dimensional frame in order to introduce interpolating NN operators, focused on the univariate density functions, the nodes where the sample values (i.e., the coefficients) of the NN are computed, and other important elements such as the weights and the threshold values.

It is well-known that the theory of NNs is mainly multivariate since applications to neurocomputing processes usually involve high dimensional data, then a multivariate extension of results proved in Costarelli (2014b) is needed.

In this paper, the interpolation of functions of several variables, defined on box-domain of $\mathbb{R}^d$, by means of multivariate operators F_n^s (introduced in Section 2) of the NN type is proved. For the sake of simplicity, the points where a given function is interpolated are in general taken on a uniform spaced grid. However, even if the grid is not uniformly spaced, or more in general the points are not disposed over a grid, NN operators which interpolate a given function at such nodes can be constructed (see Section 2.1).

In order to obtain such results, as happens in one-dimensional case, the definition of the operators N_n^σ must be strongly modified. Here for instance, the multivariate density functions $\Psi_\sigma(\underline{x})$ must be replaced by $\Psi_s(\underline{x})$, which are defined by sigmoidal functions constructed by a certain procedure involving the well-known central B-spline, see, e.g., Costarelli and Spigler (2015). In this way, we are able to consider a general family of activation functions, including for instance some known examples, such as the *ramp function*, see, e.g., Cao and Chen (2012), Cheang (2010) and Costarelli (2014a).

In order to describe the behavior of the operators F_n^s at points of $\mathcal{R}$ where continuous functions f are in general not interpolated, a uniform approximation theorem with order is also obtained. The rate of approximation is expressed in terms of the modulus of continuity of the function being approximated. Both interpolation and approximation results (see Theorems 2.5 and 2.6 in Section 2) proved in this paper are the multivariate versions of theorems first proved in Costarelli (2014b) in one-dimensional setting. In Section 3 some concrete examples of approximations and interpolations are presented, in both one and two space dimensions.

Finally, in Section 4 the main results of this paper are discussed in relation to the theory of NNs, with particular attention to the field of applications, such as, applications to the training of NNs. Moreover, a detailed comparison between N_n^σ and F_n^s is made, together with a discussion among the results here proved and those already existing concerning interpolation by NNs. In addition, the relations between NNs and sampling operators are pointed out.

2. The main results

We first introduce some notations and preliminary concepts. In this paper, we will denote by $M_s(x)$, the well-known one-dimensional central B-spline of order $s \in \mathbb{N}^+$ (see, e.g., Bardaro et al., 2003; Butzer and Nessel, 1971), defined as follows:

$$M_s(x) := \frac{1}{(s-1)!} \sum_{i=0}^{s} (-1)^i \binom{s}{i} \left(\frac{s}{2} + x - i\right)_+^{s-1}, \quad x \in \mathbb{R},$$

where the function $(x)_+ := \max\{x, 0\}$ denotes the positive part of $x \in \mathbb{R}$.

In Costarelli and Spigler (2015), a procedure to construct sigmoidal functions by using the central B-spline of order s has been described. More in detail, for any given positive integer s, we will denote by $\sigma_s(x)$ the sigmoidal function:

$$\sigma_s(x) := \int_{-\infty}^{x} M_s(t)\, dt, \quad x \in \mathbb{R}. \qquad (1)$$

Note that, $\sigma_s(x)$ are non-decreasing and $0 \le \sigma_s(x) \le 1$, for every $x \in \mathbb{R}$ and $s \in \mathbb{N}^+$. We are now able to introduce the non-negative one-dimensional *density functions* by the following finite linear combination of σ_s:

$$\phi_s(x) := \sigma_s(x + 1/2) - \sigma_s(x - 1/2), \quad x \in \mathbb{R}. \qquad (2)$$

We will use the density functions defined above as activation functions of the neural network operators studied in this paper.

It is easy to see that, the functions of the form $\phi_s(x)$ satisfy the following useful properties: