ASYMPTOTIC EXPANSIONS AND VORONOVSKAJA TYPE THEOREMS FOR THE MULTIVARIATE NEURAL NETWORK OPERATORS

DANilo costarelli∗ and Gianluca vinti

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

(Communicated by Ding-Xuan Zhou)

Abstract. In this paper, an asymptotic formula for the so-called multivariate neural network (NN) operators has been established. As a direct consequence, a first and a second order pointwise Voronovskaja type theorem has been reached. At the end, the particular case of the NN operators activated by the logistic function has been treated in details.

1. Introduction. The theory of neural network (NN) operators activated by sigmoidal functions has been largely studied in the last years, see, e.g., [8, 6, 7]. In particular, they arise as an extension of the family of approximation operators originally introduced in [8] by Cardaliaguet and Evrard, where only bell shaped activation function, with compact support were considered. Clearly, the latter assumption is restrictive; in the applications the main instances of activation function have unbounded support, as happens, e.g., for the logistic or the hyperbolic tangent sigmoidal functions.

Indeed, in [6, 7] the theory of NN operators has been directly approached for these very useful examples of sigmoidal functions, while, in [12, 13] the above results have been extended to a more general class of sigmoidal activation functions satisfying suitable (not restrictive) assumptions.

The usefulness of the NN operators is related to the theory of artificial neural networks (ANNs) introduced in twentieth century with the aim to obtain a simple model for the human brain. The ANNs revealed to be very suitable for the applications in various fields, such as biology, computer science, engineering and mathematics ([25, 27]). Nowadays, the theory of ANNs is very current in view of

2010 Mathematics Subject Classification. Primary: 41A25, 41A05; Secondary: 41A30, 47A58.

Key words and phrases. Sigmoidal functions, neural network operators, asymptotic formula, Voronovskaja type theorem, logistic function.

The first author has been partially supported within the 2019 GNAMPA-INdAM Project “Metodi di analisi reale per l’approssimazione attraverso operatori discreti e applicazioni”, while the second author within the projects: (1) Ricerca di Base 2018 dell’Università degli Studi di Perugia - “Metodi di Teoria dell’Approssimazione, Analisi Reale, Analisi Nonlineare e loro Applicazioni”, (2) Ricerca di Base 2019 dell’Università degli Studi di Perugia - “Integrazione, Approssimazione, Analisi Nonlineare e loro Applicazioni”, (3) “Metodi e processi innovativi per lo sviluppo di una banca di immagini mediche per fini diagnostici” funded by the Fondazione Cassa di Risparmio di Perugia, 2018.

∗ Corresponding author: Danilo Costarelli.
their connections with the artificial intelligence (AI). For what concerns applications to mathematics, more precisely to Approximation Theory and Learning Theory, several papers have been published, see e.g. [18, 26, 17]. The main results proved for the above subject are characterized by a non-constructive approach, where for any given function $f$, the analytic expression of the various elements which are composing a neural network which approximates $f$ in some sense, such as coefficients, weights, and thresholds, can not be explicitly determined. The non-constructive nature of the above results become them, from the mathematical point of view, of not-easy applicability. Actually, some constructive results have been proved by a quite difficult approach based on convolution, and on the theory of ridge functions, see [10]. For the above reasons, we studied the NN operators, which provide a constructive approximation processes of the neural network type based on the approach of Operator Theory.

In the present paper, an asymptotic formula for the multivariate version of the NN operators has been established. Here, we expand the above approximation operators, when they are evaluated for sufficiently smooth functions, in terms of their partial derivatives, computed at suitable points, and with suitable truncated algebraic moments of the multivariate density functions $Ψ_σ$, defined as the product of $d$ one-dimensional density functions $φ_σ$, generated by a certain finite linear combination of sigmoidal function $σ$. The function $Ψ_σ$ plays the role of the kernel for the above operators.

As a direct consequence of the above asymptotic theorem, a first and a second order pointwise Voronovskaja type theorem has been reached. In order to achieve the latter pointwise approximation result, some additional assumptions on the truncated algebraic moments of the function $Ψ_σ$ must be assumed.

At the end of the paper, the particular case of the NN operators activated by the logistic function has been treated in details. In this case, we obtain that the order of pointwise approximation is, at least, $O(n^{-2})$, as $n \to +\infty$, when $C^2$-functions are approximated, according to the well-known Korovkin’s theory for positive linear operators [24].

2. The general framework. First of all, we introduce the following notation that we will use in the paper: for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, and $\alpha \in \mathbb{R}$, we will use the usual notation for products and quotients, i.e., $\alpha x = (\alpha x_1, \ldots, \alpha x_d)$ and, for $\alpha \neq 0$, $\frac{z}{\alpha} = \left(\frac{z_1}{\alpha}, \ldots, \frac{z_d}{\alpha}\right)$. Finally, $|x|$ will denote the usual Euclidean norm of $\mathbb{R}^d$, and $x^\ell = \prod_{i=1}^d x_i^\ell_i$, $x, y \in \mathbb{R}^d$, when the power is well-defined.

Let $I := [-1, 1]^d \subset \mathbb{R}^d$ be a fixed multidimensional interval, and in what follows we define by $C(I)$ the space of all functions $f : I \to \mathbb{R}$, which are continuous on $I$, while by $C^r(I)$ we denote the functions of $C(I)$ which have continuous $s$-order partial derivatives:

$$D^h f := \frac{\partial |h|}{\partial x_h^{|h|}} f = \frac{\partial |h|}{\partial x_{h_1}^{h_1} \cdots \partial x_{h_d}^{h_d}} f,$$

with $|h| = h_1 + \cdots + h_d = s$, $h \in \mathbb{N}_0^d$ on $I$, for every $1 \leq s \leq r$, $s \in \mathbb{N}^+$. The above spaces can be considered endowed by the usual sup-norm $\|\cdot\|_{\infty}$.

For any function $Φ : \mathbb{R}^d \to \mathbb{R}$, and $ν \in \mathbb{N}$, we can define the (multivariate) truncated algebraic moment of order $ν$ by:

$$m_{ν,h}(Φ, u) := \sum_{-n \leq k \leq n} Φ(u - k) (k - u)^h, \quad u \in \mathbb{R}^d,$$
for every \( n \in \mathbb{N}^+ \), with \( \nu = |h| \), where by the notation \(-n \leq k \leq n\) we denotes the set of all the indexes \( k \in \mathbb{Z}^d \) such that \( k_i = -n, \ldots, n, i = 1, \ldots, d \). Moreover, we can also define the discrete absolute moments of \( \Phi \), as follows:

\[
M_{\nu}(\Phi) := \sup_{u \in \mathbb{R}^d} \sum_{k \in \mathbb{Z}^d} |\Phi(u - k)| \|u - k\|^{\nu},
\]

with \( \nu \geq 0 \). It is well-known that, if \( \Phi \) is bounded in a neighborhood of the origin and \( \Phi(u) = O(\|u\|^{-r-1-\varepsilon}) \), as \( \|u\| \to +\infty \), for some \( r > 0 \), \( \varepsilon > 0 \), it turns out that

\[
M_{\nu}(\Phi) < +\infty, \quad 0 \leq \nu \leq r,
\]

(1)

see e.g., [15].

Now, we recall the definition of the density functions used in order to define the neural network operators.

We say that a measurable function \( \sigma : \mathbb{R} \to \mathbb{R} \) is called a sigmoidal function if it satisfies the following conditions:

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

see e.g., [18, 12]. From now on, we consider non-decreasing sigmoidal functions \( \sigma \) which satisfy the following assumptions:

(\( \Sigma_1 \)) \( \sigma(x) - 1/2 \) is an odd function;
(\( \Sigma_2 \)) \( \sigma \in C^2(\mathbb{R}) \) is concave for \( x \geq 0 \);
(\( \Sigma_3 \)) \( \sigma(x) = O(|x|^{-\alpha - 1}) \) as \( x \to -\infty \), for some \( \alpha > 0 \).

Thus, the (univariate) density function generated by \( \sigma \) can now be defined by:

\[
\phi_\sigma(x) := \frac{1}{2} [\sigma(x + 1) - \sigma(x - 1)], \quad x \in \mathbb{R},
\]

and the corresponding multivariate version is given by:

\[
\Psi_\sigma(x) := \phi_\sigma(x_1) \cdot \phi_\sigma(x_2) \cdots \phi_\sigma(x_d), \quad x \in \mathbb{R}^d,
\]

(2)

i.e., \( \Psi_\sigma \) is the tensor product of \( d \) scalar functions. In [13] the following lemma has been established.

**Lemma 2.1.** (i) \( \phi_\sigma(x) \geq 0 \) for every \( x \in \mathbb{R} \), with \( \phi_\sigma(1) > 0 \), and moreover \( \lim_{x \to \pm\infty} \phi_\sigma(x) = 0 \);

(ii) \( \Psi_\sigma(x) = O(\|x\|^{-\alpha - 1}) \) as \( x \to \pm\infty \), where \( \alpha \) is the positive constant of condition (\( \Sigma_3 \)).

(iii) Let \( x \in I \) and \( n \in \mathbb{N}^+ \). Then:

\[
m_{0,0}^n(\Psi_\sigma, nx) = \sum_{-n \leq k \leq n} \Psi_\sigma(nx - k) \geq [\phi_\sigma(1)]^d > 0.
\]

Now we are able to recall the definition of the multivariate NN operators, see e.g., [13].

Let \( f : I \to \mathbb{R} \) be a bounded function, and \( n \in \mathbb{N}^+ \). The NN operators \( F_n \), activated by a sigmoidal function \( \sigma \) satisfying assumptions (\( \Sigma_i \)), \( i = 1, 2, 3 \), are
defined by:

\[
F_n(f, x) := \frac{\sum_{-n \leq k \leq n} f \left( \frac{k}{n} \right) \cdot \Psi_\sigma(nx - k)}{\sum_{-n \leq k \leq n} \Psi_\sigma(nx - k)}, \quad x \in I. \tag{3}
\]

Obviously, for every \( n \in \mathbb{N}^+ \), and \( k \in \mathbb{Z} \) such that \( -n \leq k \leq n \), it turns out that \( -1 \leq \frac{k}{n} \leq 1 \), for every \( i = 1, \ldots, d \). Further, in view of Lemma 2.1 (iii), the operators \( F_n \) are well-defined, and it is easy to see that \(|F_n(f, x)| \leq \|f\|_\infty\), for all \( x \in I \).

**Remark 1.** We observe that, since \( \Psi_\sigma(x) \geq 0 \), \( x \in \mathbb{R}^d \), the neural network operators \( F_n \) are positive operators, i.e., they satisfy the property \( F_n(f, \cdot) \geq 0 \), provided \( f(\cdot) \geq 0 \).

We now recall a pointwise and uniform convergence theorem for the above positive operators.

**Theorem 2.2 ([13]).** Let \( f : I \to \mathbb{R} \) be bounded. Then,

\[
\lim_{n \to +\infty} F_n(f, x) = f(x),
\]

at each point \( x \in I \) where \( f \) is continuous. Moreover, if \( f \in C(I) \) we have:

\[
\lim_{n \to +\infty} \|F_n(f, \cdot) - f(\cdot)\|_\infty = 0.
\]

**Remark 2.** We can observe that, all the theory of the neural network operators can be reformulated for sigmoidal activation functions \( \sigma \) which not-necessarily satisfy assumption \((\Sigma 2)\). In the latter case, we can assume that the corresponding \( \phi_\sigma \) satisfies the following conditions:

- \( \phi_\sigma(x) \) is non-decreasing for \( x < 0 \) and non-increasing for \( x \geq 0 \);
- \( \phi_\sigma(1) > 0 \).

As a consequence of Remark 2 we obtain that the theory of NN operators holds also in the case of not-necessarily smooth (or continuous) sigmoidal functions. Examples of activation functions and of the corresponding density functions can be generated, e.g., by the well-known ramp function, see [9].

3. **Asymptotic formulas for the multivariate NN operators.** Here, we establish an asymptotic formula for the multivariate NN operators.

**Theorem 3.1.** Let \( \sigma \) be a sigmoidal function such that, for every \( \gamma > 0 \),

\[
\lim_{n \to +\infty} \sum_{\|nu-k\|_{\gamma n} > \gamma n} \Psi_\sigma(nu-k) \cdot \|nu-k\|^r = 0, \tag{4}
\]

uniformly with respect to \( u \in \mathbb{R}^d \), for a certain \( r \in \mathbb{N} \). Moreover, we also assume that \( M_r(\Psi_\sigma) < +\infty \). Then, for any \( f \in C^r(I) \) there holds:

\[
F_n(f, x) = f(x) + \sum_{\nu=1}^{r} \sum_{|\alpha|=\nu} \frac{D^\nu f(x)}{\nu! n^\nu} \frac{m_\nu(\Psi_\sigma, nx)}{m_0(\Psi_\sigma, nx)} + o(n^{-r}),
\]

as \( n \to +\infty \), for every \( x \in \mathbb{R}^d \).
Proof. First of all, we recall that for every \( x \in I \) the following multivariate version of the Taylor formula for \( f \in C^r(I) \), \( r \in \mathbb{N} \):

\[
f(u) = f(x) + \sum_{\nu=1}^{r} \sum_{|\nu|=\nu} \frac{D^\nu f(x)}{\nu!} (u - x)^\nu + R_r(u; x), \quad u \in I, \tag{5}
\]

holds, where the term \( R_r(u, x) \) denotes a suitable remainder, and

\[
h! := h_1! h_2! \ldots h_N!.
\]

Now, expanding the function \( f \) at the nodes \( u = k/n \) by the formula (5), we obtain:

\[
f\left(\frac{k}{n}\right) = f(x) + \sum_{\nu=1}^{r} \sum_{|\nu|=\nu} \frac{D^\nu f(x)}{\nu!} \left(\frac{k}{n} - x\right)^\nu + R_r\left(\frac{k}{n}; x\right),
\]

where now the remainder term is expressed by the following local form:

\[
R_r\left(\frac{k}{n}; x\right) := \lambda \left(\frac{k}{n} - x\right) \left\|\frac{k}{n} - x\right\|^r,
\]

where \( \lambda \) is a bounded function such that \( \lim_{\nu \to 0} \lambda(\nu) = 0 \). Now, we can write what follows:

\[
F_n(f, x) = \frac{1}{m_{0,0}(\Psi, nx)} [f(x) m_{0,0}(\Psi, nx) + \sum_{-n \leq k \leq n} \left\{ \sum_{\nu=1}^{r} \sum_{|\nu|=\nu} \frac{D^\nu f(x)}{\nu!} \left(\frac{k}{n} - x\right)^\nu + R_r\left(\frac{k}{n}; x\right) \right\} \Psi(\nu x - k)]
\]

\[
= f(x) + \frac{1}{m_{0,0}(\Psi, nx)} \left[ \sum_{\nu=1}^{r} \sum_{|\nu|=\nu} \frac{D^\nu f(x)}{\nu!} \left\{ \sum_{-n \leq k \leq n} \left(\frac{k}{n} - x\right)^\nu \Psi(\nu x - k) \right\} \right] + \frac{1}{m_{0,0}(\Psi, nx)} \left[ \sum_{-n \leq k \leq n} R_r\left(\frac{k}{n}; x\right) \Psi(\nu x - k) \right] =: I_1 + I_2 + I_3.
\]

Now we can estimate the remainder term \( I_3 \). Since \( \lim_{\nu \to 0} \lambda(\nu) = 0 \) for a fixed \( \varepsilon > 0 \) there exists \( \gamma > 0 \) such that, for \( \left\|\nu\right\| \leq \gamma \) there holds \( |\lambda(\nu)| < \varepsilon \). Thus, using Lemma 2.1 (iii) we have:

\[
|I_3| \leq \frac{1}{[\phi(1)]^d} \left\{ \sum_{-n \leq k \leq n} \lambda \left(\frac{k}{n} - x\right) \left\|\frac{k}{n} - x\right\|^r \Psi(\nu x - k) \right\}
\]

\[
=: I_{3,1} + I_{3,2}
\]

For what concerns \( I_{3,1} \) we have that, if \( \left\|\nu x - k\right\| \leq \gamma n \) it turns out that \( \left\|x - k/n\right\| \leq \gamma \), hence:

\[
I_{3,1} \leq \varepsilon \sum_{k \in \mathbb{Z}^d} \left\|\frac{k}{n} - x\right\|^r \Psi(\nu x - k) \leq \varepsilon n^{-r} M_r(\Psi) < +\infty.
\]
While, for $I_{3,2}$, using assumption (4), we get:

$$I_{3,1} \leq \|\lambda\|_{\infty} n^{-r} \sum_{\|nx-k\| > \gamma n} \|nx-k\|^{r} \Psi_{\sigma}(nx-k) \leq \|\lambda\|_{\infty} n^{-r} \varepsilon,$$

for $n \in \mathbb{N}^+$ sufficiently large. From the above estimates we finally obtain:

$$I_3 = o(n^{-r}), \quad \text{as } n \to +\infty.$$

This completes the proof.

**Remark 3.** Note that, if $\sigma$ satisfies assumption $(\Sigma 3)$ with $\alpha > r$, then assumption (4) turns out to be satisfied. More precisely, it is possible to prove that, for every $\gamma > 0$:

$$\lim_{n \to +\infty} \sum_{\|nu-k\| > \gamma n} \Psi_{\sigma}(nu-k) \cdot \|nu-k\|^{r} = 0,$$

uniformly with respect to $u \in \mathbb{R}^d$, for every $0 \leq \nu \leq r$. The proof of the above claim follows easily as in the case of Lemma 2.6 (ii) of [16].

As a consequence of Theorem 3.1 we can establish the following first order Voronovskaja type theorem.

**Theorem 3.2.** Let $\sigma$ be a sigmoidal function assumed as in Theorem 3.1. Suppose in addition that, for every $x \in \mathbb{R}^d$:

$$m_{0,0}^n(\Psi_{\sigma}, x) = m_{0,0} + o(n^{-\theta_0}), \quad n \to +\infty;$$

$$m_{1,i}^n(\Psi_{\sigma}, x) = m_{1,i} + o(n^{-\theta_i}), \quad n \to +\infty,$$  \hspace{1cm} (6)

where $\theta_i := (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^d$, $m_{1,i} \in \mathbb{R}$, $\theta_i > 0$, $i = 1, \ldots, d$, and the “o-term” tends to zero uniformly with respect to $x \in \mathbb{R}^d$. Then, for any $f \in C^1(1)$ we have:

$$\lim_{n \to +\infty} n [F_n(f, x) - f(x)] = \sum_{i=1}^{d} \frac{\partial^1}{\partial x_i} f(x) \frac{m_{1,i}}{m_{0,0}},$$

where $m_{0,0} \neq 0$ in view of Lemma 2.1 (iii).

**Proof.** Applying the asymptotic formula of Theorem 3.1 with $r = 1$, and using assumption (6), we can write what follows:

$$n [F_n(f, x) - f(x)] = \sum_{i=1}^{d} \frac{\partial^1}{\partial x_i} f(x) \frac{m_{1,i}}{m_{0,0}} + o(n^{-\theta_i}) + n o(n^{-1})$$  \hspace{1cm} (7)

for $n \in \mathbb{N}^+$ sufficiently large. Then the proof follows by passing to the limit for $n \to +\infty$. \hfill $\square$

Similarly to above, the following second order Voronovskaja type theorem can be established.

**Theorem 3.3.** Let $\sigma$ be a sigmoidal function assumed as in Theorem 3.2. Suppose in addition that:

$$m_{1,i} = 0, \quad \theta_i \geq 1, \quad i = 1, \ldots, d,$$  \hspace{1cm} (8)

and for every $x \in \mathbb{R}^d$:

$$m_{2,h}^n(\Psi_{\sigma}, x) = m_{2,h} + o(n^{-\theta_2}), \quad n \to +\infty,$$  \hspace{1cm} (9)
where $\theta_{2,h} > 0$, for every $h \in \mathbb{N}_0^d$ with $|h| = 2$, and the “o-term” tends to zero uniformly with respect to $x \in \mathbb{R}^d$. Then, for any $f \in C^2(I)$ we have:

$$
\lim_{n \to +\infty} n^2 [F_n(f, x) - f(x)] = \sum_{|h|=2} \frac{D^h f(x)}{h!} \frac{m_{2,h}}{m_{0,0}} + n^2 o(n^{-2}),
$$

for $n \in \mathbb{N}^+$ sufficiently large. Then the proof follows by passing to the limit for $n \to +\infty$.

**Proof.** Applying the asymptotic formula of Theorem 3.1 with $r = 2$, and using assumptions (8) and (9), we can write what follows:

$$
n^2 [F_n(f, x) - f(x)] = \sum_{|h|=2} \frac{D^h f(x)}{h!} \frac{m_{2,h} + o(n^{-\theta_{2,h}})}{m_{0,0} + o(n^{-\theta_0})} + n^2 o(n^{-2}),
$$

Note that a quantitative version of Theorem 3.2 and Theorem 3.3 can be easily established by repeating the above proof and using some well-known inequalities; for more details see, e.g., [5, 2, 3, 4, 1].

4. **Applications.** As first example we consider the case of the well-known logistic activation function ([22, 21]):

$$
\sigma_\ell(x) := (1 + e^{-x})^{-1}, x \in \mathbb{R}.
$$

Obviously, $\sigma_\ell$ is a smooth function and it satisfies all the assumptions $(\Sigma_i), i = 1, 2, 3$. Furthermore, by its exponential decay to zero as $x \to -\infty$, condition $(\Sigma3)$ turns out to be satisfied for every $\alpha > 0$.

Now, in order to apply the results of Section 3 it remains to compute the truncated algebraic moments of the multivariate density function $\Psi_{\sigma_\ell}$ (see Fig. 1) corresponding to $\sigma_\ell$, and generated by $\phi_{\sigma_\ell}$ (see Fig. 2).
Figure 2. The function $\Psi_{\sigma}\ell$ of two variables.

It is well-known (see [16]) that, in the one-dimensional case $d = 1$, i.e., when $\Psi_{\sigma}\ell = \phi_{\sigma}\ell$ it turns out that:

$$m_{0,0}^n(\phi_{\sigma}\ell, x) := \sum_{k=-n}^{n} \phi_{\sigma}\ell(x - k) = 1 + o(n^{-1}),$$
$$m_{1,0}^n(\phi_{\sigma}\ell, x) := \sum_{k=-n}^{n} (k - x) \phi_{\sigma}\ell(x - k) = o(n^{-2}),$$
$$m_{2,0}^n(\phi_{\sigma}\ell, x) := \sum_{k=-n}^{n} (k - x)^2 \phi_{\sigma}\ell(x - k) \approx -3.6232 + o(n^{-3}),$$
as $n \to +\infty$, uniformly with respect to $x \in \mathbb{R}$. The above computations are a consequence of the application of the well-known Poisson’s summation formula, that is based on the usual $L^1$-Fourier transform of the function $\phi_{\sigma}\ell$ (see, e.g., [15]). Thus, from the above relations it is easy to deduce that, in the general multivariate case we have:

$$m_{0,0}^n(\Psi_{\sigma}\ell, x) = \sum_{k_1=-n}^{n} \phi_{\sigma}\ell(x_1 - k_1) \cdots \sum_{k_d=-n}^{n} \phi_{\sigma}\ell(x_d - k_d) = 1 + o(n^{-1}), \quad (10)$$

and similarly we can obtain:

$$m_{1,0}^n(\Psi_{\sigma}\ell, x) = o(n^{-2}), \quad i = 1, \ldots, d, \quad (11)$$
$$m_{2,0}^n(\Psi_{\sigma}\ell, x) = (-3.6232)^d + o(n^{-3}), \quad (12)$$

for every $h \in \mathbb{N}_0^d$ with $|h| = 2$, and where all the above “o-term” tends to zero uniformly with respect to $x \in \mathbb{R}^d$. In view of the above computations, and using both Theorem 3.2 and Theorem 3.3 we immediately obtain the following corollary.

**Corollary 1.** Let $\sigma\ell$ be the logistic function, and $f \in C^1(I)$ be fixed. Then for every $x \in I$, we have:

$$\lim_{n \to +\infty} n \left[ F_{\sigma}\ell_n(f, x) - f(x) \right] = 0.$$
In particular, if $f$ belongs to $C^2(I)$, it turns out that:

$$\lim_{n \to +\infty} n^2 \left[ F_n^\sigma(x) - f(x) \right] = (-3.6232)^d \sum_{|h|=2} \frac{D^h f(x)}{h!},$$

for every $x \in I$.

Applications for other useful examples of sigmoidal activation functions (see, e.g., [23, 14, 19, 20, 11]), such as the hyperbolic tangent sigmoidal function, the ramp activation function, and many other, can be easily obtained proceeding as in the above instance.

Acknowledgments. The authors are members of the Gruppo Nazionale per l’Analisi Matematica, la Probabilitè e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM). This research has been accomplished within RITA (Research ITalian network on Approximation).

REFERENCES

[1] J. A. Adell and D. Cardenas-Morales, Quantitative generalized Voronovskaja’s formulae for Bernstein polynomials, Journal of Approximation Theory, 231 (2018), 41–52.
[2] T. Acar, Asymptotic formulas for generalized Szász-Mirakyan operators, Applied Mathematics and Computation, 263 (2015), 223–239.
[3] T. Acar, A. Aral and I. Rasa, Approximation by k-th order modifications of Szász-Mirakyan operators, Studia Scient. Math. Hungar., 53 (2016), 379–398.
[4] T. Acar, A. Aral and I. Rasa, The New Forms of Voronovskaya’s Theorem in weighted spaces, Positivity, 20 (2016), 25–40.
[5] C. Bardaro and I. Mantellini, On pointwise approximation properties of multivariate semi-discrete sampling type operators, Results in Mathematics, 72 (2017), 1449–1472.
[6] F. Cao and Z. Chen, The approximation operators with sigmoidal functions, Comput. Math. Appl., 58 (2009), 758–765.
[7] F. Cao and Z. Chen, The construction and approximation of a class of neural networks operators with ramp functions, J. Comput. Anal. Appl., 14 (2012), 101–112.
[8] P. Cardaliaguet and G. Euvrard, Approximation of a function and its derivative with a neural network, Neural Networks, 5 (1992), 207–220.
[9] H. L. Cheang Gerald, Approximation with neural networks activated by ramp sigmoids, J. Approx. Theory, 162 (2010), 1450–1465.
[10] E. W. Cheney, W. A. Light and Y. Xu, Constructive methods of approximation by rigde functions and radial functions, Numerical Algorithms, 4 (1993), 205–223.
[11] D. Costarelli and A. R. Sambucini, Approximation results in Orlicz spaces for sequences of Kantorovich max-product neural network operators, Results in Mathematics, 73 (2018), Art. 15, 15 pp.
[12] D. Costarelli and R. Spigler, Approximation results for neural network operators activated by sigmoidal functions, Neural Networks, 44 (2013), 101–106.
[13] D. Costarelli and R. Spigler, Multivariate neural network operators with sigmoidal activation functions, Neural Networks, 48 (2013), 72–77.
[14] D. Costarelli, R. Spigler and G. Vinti, A survey on approximation by means of neural network operators, Journal of NeuroTechnology, 1 (2016).
[15] D. Costarelli and G. Vinti, Inverse results of approximation and the saturation order for the sampling Kantorovich series, Journal of Approximation Theory, 242 (2019), 64–82.
[16] D. Costarelli and G. Vinti, Voronovskaja type theorems and high order convergence neural network operators with sigmoidal functions, in: Mediterranean Journal of Mathematics, (2020).
[17] F. Cucker and D. X. Zhou, Learning Theory: An Approximation Theory Viewpoint, Cambridge University Press, Cambridge, 2007.
[18] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Systems, 2 (1989), 303–314.
[19] N. J. Guliyev and V. E. Ismailov, On the approximation by single hidden layer feedforward neural networks with fixed weights, *Neural Networks*, 98 (2018), 296–304.

[20] N. J. Guliyev and V. E. Ismailov, Approximation capability of two hidden layer feedforward neural networks with fixed weights, *Neurocomputing*, 316 (2018), 262–269.

[21] A. Iliev and N. Kyurkchiev, On the Hausdor distance between the Heaviside function and some transmuted activation functions, *Mathematical Modeling and Applications*, 2 (2016), 1–5.

[22] A. Iliev, N. Kyurkchiev and S. Markov, On the approximation of the cut and step functions by logistic and Gompertz functions, *BIOMATH*, 4 (2015), 201510101, 12pp.

[23] V. E. Ismailov, On the approximation by neural networks with bounded number of neurons in hidden layers, *J. Math. Anal. Appl.*, 417 (2014), 963–969.

[24] P. P. Korovkin, On convergence of linear positive operators in the space of continuous functions, *Dokl. Akad. Nauk SSSR*, 90 (1953), 961–964.

[25] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks*, 61 (2015), 85–117.

[26] S. Smale and D. X. Zhou, Learning theory estimates via integral operators and their approximations, *Constructive Approximation*, 26 (2007), 153–172.

[27] Y. Zhang, J. Wu, Z. Cai, B. Du and P.S. Yu, An unsupervised parameter learning model for RVFL neural network, *Neural Networks*, 112 (2019), 85–97.

Received December 2019; revised February 2020.

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