Deep Neural Networks and the Analysis of the Approximation Error

Anish Ray

University of Muenster

01 June 2022

Introduction

An artificial neural network combines biological principles with advanced statistics to solve problems in domains such as pattern recognition and game-play.
In this talk we are going to learn the mathematics behind the deep neural networks, the vectorized and the structured description and how they are connected, some examples and properties of DNNs.
Then in the last section we will try to analyse the approximation error.
Note. We wil see soon that d denotes the no. of neurons, L denotes the no. of layers, l_i the no. of neurons in the *ith* layer, and the vector $\theta \in \mathbb{R}^d$ stores the real parameters (the weights and the biases) for the considered neural network $\mathcal{N}^{\theta,\delta,l_0}_{\Psi_1,\Psi_2,\ldots,\Psi_L}$.

Deep Neural Networks

Definition 1

(Affine functions) Let $d, r, s \in \mathbb{N}$, $\delta \in \mathbb{N} \cup \{0\}$, $\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \mathbb{R}^d$ satisfy $d \geq \delta + rs + r$. Then we denote by $\mathcal{A}_{r,s}^{\theta,\delta} : \mathbb{R}^s \to \mathbb{R}^r$ the function which satisfies for allthe function which satisfies for all $x = (x_1, x_2, \cdots, x_s) \in \mathbb{R}^s$ that

$$\mathcal{A}_{r,s}^{\theta,\delta}(x) = \begin{pmatrix} \theta_{\delta+1} & \theta_{\delta+2} & \cdots & \theta_{\delta+s} \\ \theta_{\delta+s+1} & \theta_{\delta+s+2} & \cdots & \theta_{\delta+2s} \\ \theta_{\delta+2s+1} & \theta_{\delta+2s+2} & \cdots & \theta_{\delta+3s} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{\delta+(r-1)s+1} & \theta_{\delta+(r-1)s+2} & \cdots & \theta_{\delta+rs} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_s \end{pmatrix} + \begin{pmatrix} \theta_{\delta+rs+1} \\ \theta_{\delta+rs+2} \\ \theta_{\delta+rs+3} \\ \vdots \\ \theta_{\delta+rs+r} \end{pmatrix} \tag{1}$$

$$= \left(\left[\sum_{k=1}^{s} x_k \theta_{\delta+k} \right] + \theta_{\delta+rs+1}, \ldots, \left[\sum_{k=1}^{s} x_k \theta_{\delta+(r-1)s+k} \right] + \theta_{\delta+rs+r} \right)$$

Vectorized description of DNNs

Definition 2

Let $d, L \in \mathbb{N}$, $l_o, l_1, \ldots, l_L \in \mathbb{N}$, $\delta \in \mathbb{N} \cup \{0\}$, $\theta \in \mathbb{R}^d$ satisfy $d \geq \delta + \sum_{k=1}^{L} l_k(l_{k-1} + 1)$ and let $\Psi : \mathbb{R}^{l_k} \to \mathbb{R}^{l_k}$, $k \in \{1, 2, \ldots, L\}$ be functions. Then we denote by $\mathcal{N}^{\theta,\delta,l_0}_{\psi_1,\Psi_2,\ldots,\Psi_L} : \mathbb{R}^{l_0} \to \mathbb{R}^{l_L}$ the function which satisfies for all $x \in \mathbb{R}^{l_0}$ that

$$\mathcal{N}^{\theta,\delta,l_0}_{\Psi_1,\Psi_2,\ldots,\Psi_L}(x) = \left(\Psi_L \circ \mathcal{A}^{\theta,\delta+\sum_{k=1}^{L-1} l_k(l_{k-1}+1)}_{l_L,l_{L-1}} \circ \cdots \circ \Psi_1 \circ \mathcal{A}^{\theta,\delta}_{l_1,l_0} \right)(x). \tag{2}$$

Definition 3

(Activation functions) Let $d \in \mathbb{N}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be a function. Then we denote by $\mathfrak{M}_{\psi,d} : \mathbb{R}^d \to \mathbb{R}^d$ the function which satisies for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that

$$\mathfrak{M}_{\psi,d}(x) = (\psi(x_1), \psi(x_2), \ldots, \psi(x_d)). \tag{3}$$

Definition 4

(Rectifier function) We denote by $\mathfrak{r} : \mathbb{R} \to \mathbb{R}$ the function which satisfies for all $x \in \mathbb{R}$ that

$$\mathfrak{r}(x) = \max\{x, 0\}. \tag{4}$$

Definition 5

(Multidimensional rectifier function) Let $d \in \mathbb{N}$. Then we denote by $\mathfrak{R}_d : \mathbb{R}^d \to \mathbb{R}^d$ the function given by

$$\mathfrak{R}_d = \mathfrak{M}_{\mathfrak{r},d}. \tag{5}$$

Definition 6

(Clipping function) Let $u \in [-\infty, \infty)$, $v \in (u, \infty]$. Then we denote by $\mathfrak{c}_{u,v} : \mathbb{R} \to \mathbb{R}$ the function which satisifes for all $x \in \mathbb{R}$ that

$$\mathfrak{c}_{u,v}(x) = \max\{u, \min\{x, v\}\}. \tag{6}$$

Definition 7

(Multidimensional clipping function) Let $d \in \mathbb{N}$, $u \in [-\infty, \infty)$, and $v \in (u, \infty]$. Then we denote by $\mathfrak{C}_{u,v,d} : \mathbb{R}^d \to \mathbb{R}^d$ the function given by

$$\mathfrak{C}_{u,v,d} = \mathfrak{M}_{\mathfrak{c}_{u,v},d}. \tag{7}$$

Definition 8

(Rectified Clipped DNNs) Let $d \in \mathbb{N}$, $u \in [-\infty, \infty)$, $v \in (u, \infty]$, $\mathbf{l} = (l_0, l_1, \ldots, l_L) \in \mathbb{N}^{L+1}$, and $\theta \in \mathbb{R}^d$, satisfy $d \geq \sum_{k=1}^{L} l_k(l_{k-1} + 1)$. Then we denote by $\mathcal{N}_{u,v}^{\theta,\mathbf{l}} : \mathbb{R}^{l_0} \to \mathbb{R}^{l_L}$ the function which satisfies for all $x \in \mathbb{R}^{l_0}$ that

$$\mathcal{N}_{u,v}^{\theta,\mathbf{l}}(x) = \begin{cases} \mathcal{N}_{\mathfrak{C}_{u,v,l_L}}^{\theta,0,l_0}(x) & : L = 1 \\ \mathcal{N}_{\mathfrak{R}_{l_1},\mathfrak{R}_{l_2},\ldots,\mathfrak{R}_{l_{L-1}},\mathfrak{C}_{u,v,l_L}}^{\theta,0,l_0}(x) & : L > 1 \end{cases}. \tag{8}$$

Structured description of DNNs

Definition 9

We denote by $\mathbf{N}$ the set given by

$$\mathbf{N} = \bigcup_{L\in\mathbb{N}} \bigcup_{(l_0,l_1,\ldots,l_L)\in\mathbb{N}^{L+1}} \left(\bigtimes_{k=1}^{L} \left(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right) \right) \tag{9}$$

and we denote by $\mathcal{P}, \mathcal{L}, \mathcal{I}, \mathcal{O} : \mathbf{N} \to \mathbb{N}$, $\mathcal{H} : \mathbf{N} \to \mathbb{N} \cup \{0\}$, and $\mathcal{D} : \mathbf{N} \to \left(\bigcup_{L=2}^{\infty} \mathbb{N}^L \right)$ the functions which satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi \in \left(\bigtimes_{k=1}^{L} \left(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right) \right)$ that $\mathcal{P}(\Phi) = \sum_{k=1}^{L} l_k(l_{k-1}+1)$, $\mathcal{L}(\Phi) = L$, $\mathcal{I}(\Phi) = l_0$, $\mathcal{O}(\Phi) = l_L$, $\mathcal{H}(\Phi) = L - 1$, and $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$.

Note.$\mathcal{L}(\Phi)$ denotes the number of layers, $\mathcal{I}(\Phi)$ denotes the number of neurons in the input layer, $\mathcal{O}(\Phi)$ denotes the numbers of neurons in the output layer, $\mathcal{P}(\Phi)$ the total number of neurons.

Definition 10

(Realization associated to a DNN) Let $a \in C(\mathbb{R}, \mathbb{R})$. Then we denote by $\mathcal{R}_a : \mathbf{N} \to \left(\bigcup_{k,l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l) \right)$ the function which satisfies for all $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), \ldots, (W_L, B_L)) \in \left(\times_{k=1}^{L} \left(\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k} \right) \right)$, $x_0 \in \mathbb{R}^{l_0}$, $x_1 \in \mathbb{R}^{l_1}$,..., $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $\forall k \in \mathbb{N} \cap (0, L)$: $x_k = \mathfrak{M}_{a, l_k}(W_k x_{k-1} + B_k)$ that

$$\mathcal{R}_a(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}) \quad \text{and} \quad (\mathcal{R}_a(\Phi))(x_0) = W_L x_{L-1} + B_L. \quad (10)$$

On the connection to the vectorized description of DNNs

Definition 11

We denote by $\mathcal{T} : \mathbf{N} \to \left(\bigcup_{d \in \mathbb{N}} \mathbb{R}^d \right)$ the function which satisfies for all $L, d \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), \ldots, (W_L, B_L)) \in \left(\bigtimes_{m=1}^{L} \left(\mathbb{R}^{l_m \times l_{m-1}} \times \mathbb{R}^{l_m} \right) \right)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \mathbb{R}^d$, $k \in \{1, 2, \ldots, L\}$ with $\mathcal{T}(\Phi) = \theta$ that $d = \mathcal{P}(\Phi)$,

$$B_k = \begin{pmatrix} \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)) + l_k l_{k-1} + 1} \\ \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)) + l_k l_{k-1} + 2} \\ \vdots \\ \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)) + l_k l_{k-1} + l_k} \end{pmatrix},$$ and

$$W_k = \begin{pmatrix} \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)+1} & \dots & \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)+l_{k-1}} \\ \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)+l_{k-1}+1} & \dots & \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)+2l_{k-1}} \\ \vdots & \ddots & \vdots \\ \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)+(l_k-1)l_{k-1}+1} & \dots & \theta_{(\sum_{i=1}^{L} l_i(l_{i-1}+1)+l_k l_{k-1}} \end{pmatrix} \tag{11}$$

Lemma 1.1

Let $a, b \in \mathbb{N}$, $W = (W_{i,j})_{(i,j)\in\{1,2,\dots,a\}\times\{1,2,\dots,b\}} \in \mathbb{R}^{a\times b}$, $B = (B_i)_{i\in\{1,2,\dots,a\}} \in \mathbb{R}^a$. *Then*

$$\mathcal{T}(((W, B))) = (W_{1,1}, \dots, W_{1,b}, \dots, W_{a,1}, \dots, W_{a,b}, B_1, \dots, B_a) \tag{12}$$

Lemma 1.2

Let $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$,
$W_k = (W_{k,i,j})_{(i,j) \in \{1,2,\ldots,l_k\} \times \{1,2,\ldots,l_{k-1}\}} \in \mathbb{R}^{l_k \times l_{k-1}}$,
$B_k = (B_{k,i})_{i \in \{1,2,\ldots,l_k\}} \in \mathbb{R}^{l_k}$, $k \in \{1, 2, \ldots, L\}$. *Then it holds for all* k*, that*

(i)
$\mathcal{T}(((W, B))) = (W_{k,1,1}, \ldots, W_{k,1,l_{k-1}}, \ldots, W_{k,l_k,1}, \ldots, W_{k,l_k,l_{k-1}},$
$B_{k,1}, \ldots, B_{k,l_k})$ *and,*

(ii) $\mathcal{T}(((W_1, B_1), \ldots, (W_L, B_L)))) =$
$(W_{1,1,1}, \ldots, W_{1,1,l_0}, \ldots, W_{1,l_1,1}, \ldots, W_{1,l_1,l_0}, B_{1,1}, \ldots, B_{1,l_1},$
$\ldots,$
$W_{L,1,1}, \ldots, W_{L,1,l_{L-1}}, \ldots, W_{L,l_L,1}, \ldots, W_{L,l_L,l_{L-1}}, B_{L,1}, \ldots, B_{L,l_L})$.

Lemma 1.3

Let $a \in C(\mathbb{R}, \mathbb{R})$, $\Phi \in \mathbf{N}$, $L \in \mathbb{N}$, $l_0, l_1, \ldots, l_L \in \mathbb{N}$ *satisfy* $\mathcal{D}(\Phi) = (l_0, l_1, \ldots, l_L)$. *Then it holds for all* $x \in \mathbb{R}^{l_0}$ *that*

$$(\mathcal{R}_a(\Phi))(x) = \begin{cases} \mathcal{N}_{\mathrm{id}_{\mathbb{R}^{l_L}}}^{\mathcal{T}(\Phi),0,l_0}(x) & : L = 1 \\ \mathcal{N}_{\mathfrak{M}_{a,l_1},\mathfrak{M}_{a,l_2},\ldots,\mathfrak{M}_{a,l_{L-1}},\mathrm{id}_{\mathbb{R}^{l_L}}}^{\mathcal{T}(\Phi),0,l_0}(x) & : L > 1 \end{cases} \tag{13}$$

Proof.

Let $W_k \in \mathbb{R}^{l_k \times l_{k-1}}$, $k \in \{1, 2, \ldots, L\}$ satisfy $\Phi = ((W_1, B_1), \ldots, (W_L, B_L))$. Then from (11) for $x \in \mathbb{R}^{l_{k-1}}$ it holds that

$$W_k x + B_k = \left(\mathcal{A}_{l_k, l_{k-1}}^{\mathcal{T}(\Phi), \sum_{i=1}^{k-1} l_i(l_{i-1}+1)} \right)(x). \tag{14}$$

Then for $x_k \in \mathbb{R}^{l_{k-1}}$ for all $k \in \{1, 2, \ldots, L-1\}$ with

(Cont.)

$x_k = \mathfrak{M}_{a,l_k}(W_k x_{k-1} + B_k)$ it holds that $x_{L-1} =$

$$\begin{cases} x_0 & : L = 1 \\ \left(\mathfrak{M}_{a,l_{L-1}} \circ \mathcal{A}_{l_{L-1},l_{L-2}}^{\mathcal{T}(\Phi),\sum_{i=1}^{L-2} l_i(l_{i-1}+1)} \circ \cdots \circ \mathfrak{M}_{a,l_1} \circ \mathcal{A}_{l_1,l_0}^{\mathcal{T}(\Phi),0} \right)(x_0) & : L > 1 \end{cases}$$

Then the result follows from (2) and (10).

Corollary 1.4. Let $\Phi \in \mathbf{N}$. Then it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that $\left(\mathcal{N}_{-\infty,\infty}^{\mathcal{T}(\Phi),\mathcal{D}(\Phi)}\right)(x) = (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)$. We will need this particular result later in the proof of Corollary 2.6.

Definition 12

(Parallelization of DNNs) Let $n \in \mathbb{N}$. Then we denote by $\mathbf{P}_n : \{(\Phi_1, \Phi_2, \ldots, \Phi_n) \in \mathbf{N}^n : \mathcal{L}(\Phi_1) = \cdots = \mathcal{L}(\Phi_n)\} \to \mathbf{N}$ the function which satisfies for all $L \in \mathbb{N}$, $(l_{k,0}, \ldots, l_{k,L}) \in \mathbb{N}^{L+1}$, $k \in \{1, 2, \ldots, n\}$,

$\Phi_k = ((W_{k,1}, B_{k,1}), \dots, (W_{k,L}, B_{k,L})) \in \left(\bigtimes_{m=1}^{L} (\mathbb{R}^{l_{k,m} \times l_{k,m-1}} \times \mathbb{R}^{l_{k,m}}) \right)$, $k \in \{1, 2, \dots, n\}$ that

$$\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n) = ((\begin{pmatrix} W_{1,1} & 0 & \dots & 0 \\ 0 & W_{2,1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & W_{n,1} \end{pmatrix} \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ \vdots \\ B_{n,1} \end{pmatrix}),$$

$$(\begin{pmatrix} W_{1,2} & 0 & \dots & 0 \\ 0 & W_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & W_{n,2} \end{pmatrix} \begin{pmatrix} B_{1,2} \\ B_{2,2} \\ \vdots \\ B_{n,2} \end{pmatrix}), \dots,$$

$$(\begin{pmatrix} W_{1,L} & 0 & \dots & 0 \\ 0 & W_{2,L} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & W_{n,L} \end{pmatrix} \begin{pmatrix} B_{1,L} \\ B_{2,L} \\ \vdots \\ B_{n,L} \end{pmatrix})).$$

Examples of DNNs

1. Let $m, n \in \mathbb{N}$, $W \in \mathbb{R}^{m\times n}$. Then we denote by $\mathfrak{N}_W \in \mathbb{R}^{m\times n} \times R^m$ the pair given by $\mathfrak{N}_W = (W, 0)$. In this case, $L = 1$.
2. We denote by $\mathfrak{J} = (\mathfrak{J}_d)_{d\in\mathbb{N}} : \mathbb{N} \to \mathbf{N}$ the function which satisfies for all $d \in \mathbb{N}$ that $\mathfrak{J}_1 = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & -1 \end{pmatrix}, 0 \right) \right)$ and $\mathfrak{J}_d = \mathbf{P}_d(\mathfrak{J}_1, \ldots, \mathfrak{J}_1)$. In this case, $L = 2$.

Definition 13

(Composition of DNNs) We denote by $(.) \bullet (.) : \{(\Phi_1, \Phi_2) \in \mathbf{N} \times \mathbf{N} : \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)\} \to \mathbf{N}$ be the function which satisfies for all $L, \mathfrak{L} \in \mathbb{N}$, $l_0, l_1, \ldots, l_L, \mathfrak{l}_0, \ldots, \mathfrak{l}_L \in \mathbb{N}$,
$\Phi_1 = ((W_1, B_1), \ldots, (W_L, B_L)) \in \left(\times_{k=1}^{L} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}) \right)$,
$\Phi_2 = ((\mathfrak{W}_1, \mathfrak{B}_1), \ldots, (\mathfrak{W}_{\mathfrak{L}}, \mathfrak{B}_{\mathfrak{L}})) \in \left(\times_{k=1}^{\mathfrak{L}} (\mathbb{R}^{\mathfrak{l}_k \times \mathfrak{l}_{k-1}} \times \mathbb{R}^{\mathfrak{l}_k}) \right)$

with $l_0 = \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2) = \mathfrak{l}_{\mathfrak{L}}$ that if $L > 1 < \mathfrak{L}$ then
$\Phi_1 \bullet \Phi_2 = ((\mathfrak{W}_1, \mathfrak{B}_1), \ldots, (\mathfrak{W}_{\mathfrak{L}-1}, \mathfrak{B}_{\mathfrak{L}-1}), (W_1\mathfrak{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1), (W_2, B_2), \ldots, (W_L, B_L))$, else if $L > 1 = \mathfrak{L}$ then
$\Phi_1 \bullet \Phi_2 = ((W_1\mathfrak{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1), (W_2, B_2), \ldots, (W_L, B_L)(W_1\mathfrak{B}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1), (W_2, B_2), \ldots, (W_L, B_L))$, else if $L = 1 < \mathfrak{L}$ then
$\Phi_1 \bullet \Phi_2 = ((\mathfrak{W}_1, \mathfrak{B}_1), \ldots, (\mathfrak{W}_{\mathfrak{L}-1}, \mathfrak{B}_{\mathfrak{L}-1}), (W_1\mathfrak{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1))$, else if $L = 1 = \mathfrak{L}$ then, $\Phi_1 \bullet \Phi_2 = ((W_1\mathfrak{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}} + B_1)$.

Definition 14

(Maximum norm) We denote by $|||.||| : \left(\bigcup_{d=1}^{\infty} \mathbb{R}^d\right) \to [0, \infty)$ the function which satisfies for all $d \in \mathbb{N}$, $\theta = (\theta_1, \theta_2, \ldots, \theta_d) \in \mathbb{R}^d$ that

$$|||\theta||| = \max_{i \in \{1,2,\ldots,d\}} |\theta_i|$$

.

Lemma 1.4

Let Φ_1, Φ_2 *and all the other variables be the same as described in definition* 13. *Then*

$$|||\mathcal{T}(\Phi_1 \bullet \Phi_2)||| \leq \max\{|||\mathcal{T}(\Phi_1)|||, |||\mathcal{T}(\Phi_2)|||, |||\mathcal{T}(W_1\mathfrak{W}_{\mathfrak{L}}, W_1\mathfrak{B}_{\mathfrak{L}}+B_1)||| \tag{15}$$

.

Analysis of the approximation error

In this section we study the approximation error. In particular, the main result in in this section is Proposition 2.4 below, establishes an upper bound for the error in the approximation of a Lipschitz continuous function by DNNs. This approximation result is obtained by combining the essentially well-known approximation result in Lemma 2.1 with the DNN calculus in the previous section. The elementary result in Lemma 2.2 is basically well-known in the scientific literature.

Approximations for Lipschitz continuous functions

Lemma 2.1

Let (E, δ) *be a metric space, let* $M \subset E$ *satisfy* $M \neq \phi$*,* $L \in [0, \infty)$ *and* $f : E \to \mathbb{R}$ *satisfy for all* $x \in E$ *and* $y \in M$*, that* $|f(x) - (y)| \leq L\delta(x, y)$ *and let* $F : E \to \mathbb{R} \cup \infty$ *satisfy for all* $x \in E$ *that* $F(x) = \sup_{y \in M} [f(y) - L\delta(x, y)]$*. Then it holds for all*

(i) $x \in E$ *that* $F(x) \leq f(x)$*,*

(ii) $x \in M$ *that* $F(x) = f(x)$*,*

(iii) $x, y \in E$ *that* $|F(x) - (y)| \leq L\delta(x, y)$*,*

(iv) $x \in E$ *that* $|F(x) - f(x)| \leq 2L[\inf_{y \in M} \delta(x, y)]$*.*

DNN representations for maxima

Lemma 2.2

Let $\Phi \in \mathbf{N}$ *satisfy*

$$\Phi = \left(\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, 0 \right) \right) \in$$

$((\mathbb{R}^{3\times 2} \times \mathbb{R}^3) \times (\mathbb{R}^{1\times 3} \times \mathbb{R}))$. *Then*

(i) it holds for all $k \in \mathbb{N}$ *that* $\mathcal{L}(\mathfrak{I}_k) = 2$,

(ii) there exist unique $\phi_k \in \mathbf{N}$, $k \in \{2, 3, ...\}$, *which satisfy for all* $k \in \{2, 3, \dots\}$ *that* $\phi_2 = \Phi$, $\mathcal{I}(\Phi_k) = \mathcal{O}(\mathbf{P}_2(\Phi, \mathfrak{I}_{k-1}))$, *and* $\phi_{k+1} = \phi_k \bullet (\mathbf{P}_2(\Phi, \mathfrak{I}_{k-1}))$,

(iii) it holds for $k \in \{2, 3, \dots\}$ *that* $\mathcal{L}(\phi_k) = k$

(iv) it holds for $k \in \{2, 3, \dots\}$ *that* $\mathcal{D}(\phi_k) = (k, 2k-1, 2k-3, \dots, 3, 1) \in \mathbb{N}^{k+1}$,

(v) it holds for $k \in \{2, 3, \dots\}$, $x = (x_1, x_2, \dots, x_k) \in \mathbb{R}^k$ *that* $(\mathcal{R}_{\mathfrak{r}}(\phi_k))(x) = \max\{x_1, x_2, \dots, x_k\}$.

Interpolations through DNNs

Lemma 2.3

Let $\phi_k \in \mathbf{N}$, $k \in \{2, 3, \ldots\}$ *satisfy* $\mathcal{I}(\Phi_k) = \mathcal{O}(\mathbf{P}_2(\phi_2, \mathfrak{J}_{k-1}))$, $\phi_{k+1} = \phi_k \bullet (\mathbf{P}_2(\phi_2, \mathfrak{J}_{k-1}))$, *and* $\phi_2 \in \mathbf{N}$ *satisfy*

$$\Phi = \left(\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, 0 \right) \right) \in$$

$((\mathbb{R}^{3\times 2} \times \mathbb{R}^3) \times (\mathbb{R}^{1\times 3} \times \mathbb{R}))$, *let* $d \in \mathbb{N}$, $L \in [0, \infty)$ *and let* $M \subset \mathbb{R}^d$ *satisfy* $|M| \in \{2, 3, \ldots\}$, *let* $m : \{1, 2, \ldots, |M|\} \to M$ *be bijective, let* $f : M \to \mathbb{R}$ *and* $F : \mathbb{R}^d \to \mathbb{R}$ *satisfy for all* $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ *that*

$$F(x) = \max_{y=(y_1, y_2, \ldots, y_d) \in M} [f(y) - L \left(\sum_{i=1}^{d} |x_i - y_i| \right)],$$ *let*

$W_1 \in \mathbb{R}^{2d\times d}$, $W_2 \in \mathbb{R}^{1\times 2d}$ *and* $B_z \in \mathbb{R}^{2d}$, $z \in M$ *satisfy for all* $z = (z_1, z_2, \ldots, z_d) \in M$

$$W_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & -1 \end{pmatrix},\ B_z = \begin{pmatrix} -z_1 \\ z_1 \\ -z_2 \\ z_2 \\ \vdots \\ -z_d \\ z_d \end{pmatrix} \text{ and}$$

$W_2 = \begin{pmatrix} -L & -L & \cdots - L \end{pmatrix}$, let $\mathcal{W}_1 \in \mathbb{R}^{2d|M| \times d}$, $\mathcal{B}_1 \in \mathbb{R}^{2d|M|}$,

$\mathcal{W}_2 \in \mathbb{R}^{|M| \times 2d|M|}$ and $\mathcal{B}_1 \in \mathbb{R}^{|M|}$ satisfy $\mathcal{W}_1 = \begin{pmatrix} W_1 \\ W_1 \\ \vdots \\ W_1 \end{pmatrix}$,

$$\mathcal{B}_1 = \begin{pmatrix} B_{m(1)} \\ B_{m(2)} \\ \vdots \\ B_{m(|M|)} \end{pmatrix},\ \mathcal{W}_2 = \begin{pmatrix} W_2 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_2 \end{pmatrix},$$

$\mathcal{B}_2 = \begin{pmatrix} f(m(1)) \\ f(m(2)) \\ \vdots \\ f(m(|M|)) \end{pmatrix}$, and let $\Phi \in \mathbf{N}$ satisfy

$\Phi = \phi_{|M|} \bullet ((\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2))$. Then it holds that

(i) $\mathcal{D}(\Phi) = (d, 2d|M|, 2|M| - 1, 2|M| - 3, \ldots, 3, 1) \in \mathbb{N}^{|M|+2}$,

(ii) $\mathcal{L}(\Phi) = |M| + 1$,

(iii) $|||\mathcal{T}(\Phi)||| \leq \max\{1, L, \sup_{z \in M} |||z|||, 2[\sup_{z \in M} |f(z)|]\}$,

(iv) $F = \mathcal{R}_{\mathfrak{r}}(\Phi)$.

Explicit approximations through DNNs

Proposition 2.4. Let $\phi_k \in \mathbf{N}$, $k \in \{2, 3, \dots\}$ satisfy $\mathcal{I}(\Phi_k) = \mathcal{O}(\mathbf{P}_2(\phi_2, \mathfrak{I}_{k-1}))$, $\phi_{k+1} = \phi_k \bullet (\mathbf{P}_2(\phi_2, \mathfrak{I}_{k-1}))$, and $\phi_2 \in \mathbf{N}$ satisfy

$$\Phi = \left(\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, 0 \right) \right) \in$$

$((\mathbb{R}^{3\times 2} \times \mathbb{R}^3) \times (\mathbb{R}^{1\times 3} \times \mathbb{R}))$, let $d \in \mathbb{N}$, $L \in \mathbb{R}$ and let $D \subset \mathbb{R}^d$ be a set $f : D \to \mathbb{R}$ satisfy for all $x = (x_1, x_2, \dots, x_d) \in D$ and $y = (y_1, y_2, \dots, y_d) \in D$ that $|f(x) - f(y)| \leq L[\sum_{i=1}^{d} |x_i - y_i|]$, let $M \subset D$ satisfy $|M| \in \{2, 3, \dots\}$, let $m : \{1, 2, \dots, |M|\} \to M$ be bijective, let $W_1 \in \mathbb{R}^{2d\times d}, W_2 \in \mathbb{R}^{1\times 2d}$ and $B_z \in \mathbb{R}^{2d}$, $z \in M$ satisfy for all $z = (z_1, z_2, \dots, z_d) \in M$ that

$$W_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & -1 \end{pmatrix}, \ B_z = \begin{pmatrix} -z_1 \\ z_1 \\ -z_2 \\ z_2 \\ \vdots \\ -z_d \\ z_d \end{pmatrix} \text{ and}$$

$W_2 = \begin{pmatrix} -L & -L & \cdots - L \end{pmatrix}$, let $\mathcal{W}_1 \in \mathbb{R}^{2d|M| \times d}$, $\mathcal{B}_1 \in \mathbb{R}^{2d|M|}$, $\mathcal{W}_2 \in \mathbb{R}^{|M| \times 2d|M|}$ and $\mathcal{B}_1 \in \mathbb{R}^{|M|}$ satisfy $\mathcal{W}_1 = \begin{pmatrix} W_1 \\ W_1 \\ \vdots \\ W_1 \end{pmatrix}$,

$$\mathcal{B}_1 = \begin{pmatrix} B_{m(1)} \\ B_{m(2)} \\ \vdots \\ B_{m(|M|)} \end{pmatrix}, \ \mathcal{W}_2 = \begin{pmatrix} W_2 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_2 \end{pmatrix},$$

$\mathcal{B}_2 = \begin{pmatrix} f(m(1)) \\ f(m(2)) \\ \vdots \\ f(m(|M|)) \end{pmatrix}$, and let $\Phi \in \mathbf{N}$ satisfy

$\Phi = \phi_{|M|} \bullet ((\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2))$. Then it holds that

(i) $\mathcal{D}(\Phi) = (d, 2d|M|, 2|M| - 1, 2|M| - 3, \ldots, 3, 1) \in \mathbb{N}^{|M|+2}$,

(ii) $|||\mathcal{T}(\Phi)||| \leq \max\{1, L, \sup_{z \in M} |||z|||, 2[\sup_{z \in M} |f(z)|]\}$,

(iii) $\sup_{x \in D} |f(x) - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x)| \leq$

$2L[\sup_{x=(x_1,x_2,\ldots,x_d) \in D} (\inf_{y=(y_1,y_2,\ldots,y_d) \in M} \sum_{i=1}^{d} |x_i - y_i|)]$.

Proof.

Let $F : \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ that $F(x) = \max_{y=(y_1,y_2,\ldots,y_d) \in M} [f(y) - L\left(\sum_{i=1}^{d} |x_i - y_i|\right)]$. Then lemma 2.3 establishes (i) and (ii). Then, we use lemma 2.1 and lemma 2.3 along with appropriate change of variables according to the notations in lemma 2.1 to establish (iii). □

Implicit approximations through DNNs

Corollary 2.5. Let $d, \mathfrak{d} \in \mathbb{N}$, $L \in \mathbb{R}$ and let $D \subset \mathbb{R}^d$ be a set $f : D \to \mathbb{R}$ satisfy for all $x = (x_1, x_2, \dots, x_d) \in D$ and $y = (y_1, y_2, \dots, y_d) \in D$ that $|f(x) - f(y)| \leq L[\sum_{i=1}^{d} |x_i - y_i|]$, $M \subset D$ satisfy $|M| \in \{2, 3, \dots\}$, and let $l = (l_0, l_1, \dots, l_{|M|+1}) \in \mathbb{N}^{m+2}$ satisfy $l = (d, 2d|M|, 2|M| - 1, 2|M| - 3, \dots, 3, 1)$ and $\sum\limits_{k=1} |M| + 1 l_k(l_{k-1} + 1) \leq \mathfrak{d}$. Then there exist $\theta \in \mathbb{R}^{\mathfrak{d}}$ such that $|||\theta||| \leq \max\{1, L, \sup\limits_{z \in M} |||z|||, 2[\sup\limits_{z \in M} |f(z)|]\}$ and

$$\sup_{x \in D} |f(x) - \left(\mathcal{N}^{\theta, l}_{\infty, \infty}\right)(x)| \leq$$

$$2L[\sup_{x=(x_1,x_2,\dots,x_d) \in D} (\inf_{y=(y_1,y_2,\dots,y_d) \in M} \sum_{i=1}^{d} |x_i - y_i|)].$$

Proof.

From Proposition 2.4 and lemma 2.2 it follows that there exist a $\Phi \in \mathbf{N}$ such that $\sup_{x \in D} \left| f(x) - (\mathcal{R}_{\mathfrak{r}}(\Phi))(x) \right| \leq 2L[\sup_{x=(x_1,x_2,\ldots,x_d) \in D} (\inf_{y=(y_1,y_2,\ldots,y_d) \in M} \sum_{i=1}^{d} |x_i - y_i|)]$ holds. Then using corollary 1.4 the result follows. $\square$

Beck C., Jentzen A., Kuckuck B., Full error analysis for the training of deep neural networks. Infinite Dimensional Analysis, Quantum Probability and Related Topics, Vol. 25, No. 02, 2150020 (2022)