Embeddings for Barron spaces with higher-order activation functions

T.J. Heeringa, L. Spek, F.L. Schwenninger, C. Brune

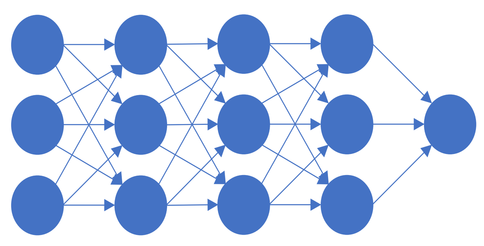
Department of Applied Mathematics, University of Twente, Enschede, The Netherlands

Mathematics of Imaging & AI

Introduction

Neural networks are universal function approximators. They are functions of the form

$$z^0(x) = W^0 x + b^0$$
$$z^{\ell+1}(x) = W^\ell \sigma\left(z^\ell(x)\right) + b^\ell$$
$$f(x) = z^L(x)$$



with $W^\ell \in \mathbb{R}^{d^{\ell+1}\times d^\ell}$, $b^\ell \in \mathbb{R}^{d^\ell}$ for $\ell \in \{0,\dots,L\}$ with $d^\ell \in \mathbb{R}$.

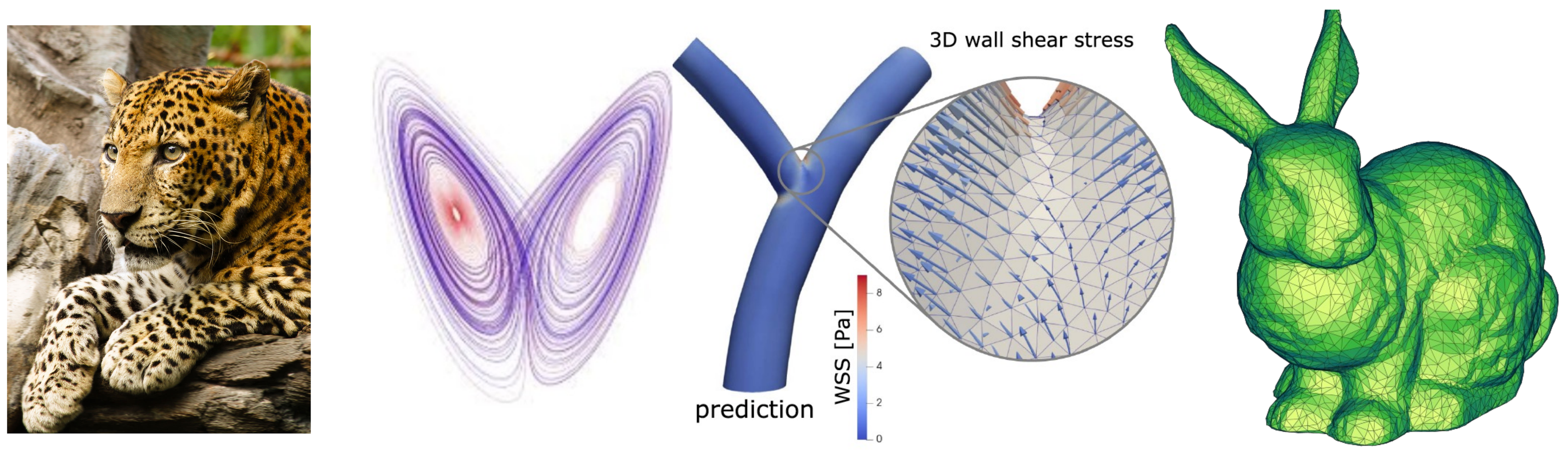


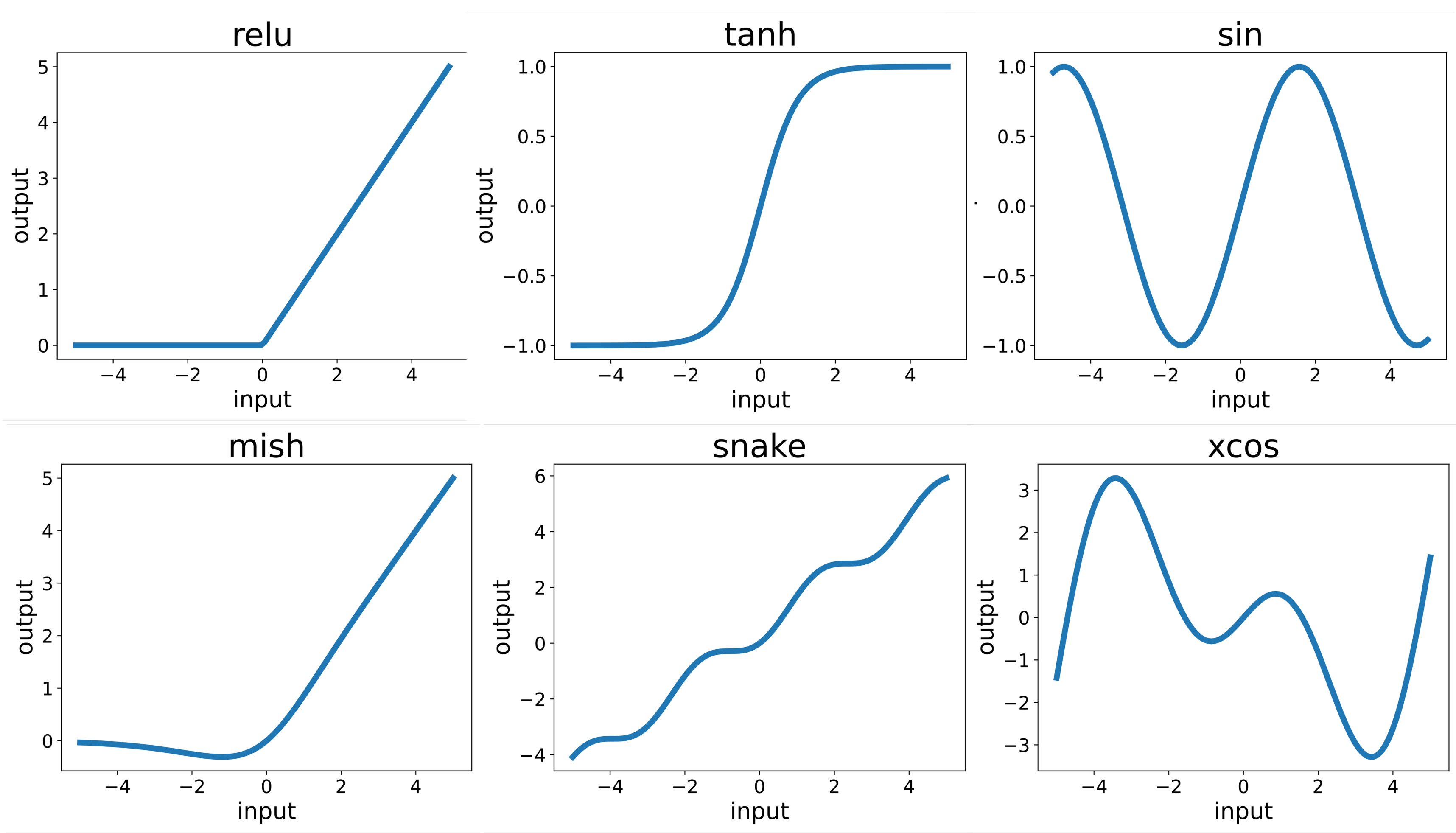
Image recognition[1] PDE discovery[2] Model reduction[3] Shape interpolation[4]

The question

The choice of activation functions σ^ℓ impacts

- expressive power
- convergence rate
- generalization performance

What happens to the underlying function space when we change the activation function σ?



Examples of activation functions[5,6,7]

Barron spaces

Barron spaces[8] $\mathcal{B}_\sigma$ are the function spaces naturally associated to shallow neural networks when $\sigma \in C^{0,1}(\mathbb{R})$ is a Lipschitz continuous function or when $\sigma(x) = \mathrm{RePU}_s(x) \coloneqq \max(0,x)^s$. For a $X \subseteq \mathbb{R}^d$ and $\Omega \subseteq \mathbb{R}^{d+1}$, the functions in this space take the form

$$f: X \to \mathbb{R}, x \mapsto \int_\Omega \sigma(\langle x|w\rangle + b)\, d\mu(w,b)$$

and the the norm of the space is given by

$$\|f\|_{\mathcal{B}_\sigma} = \begin{cases} \inf_{\mu\in\mathbb{G}_{\sigma,f}} \int_\Omega (1 + \|w\|_{\ell_1} + |b|)\, d|\mu|(w,b) & \sigma \in C^{0,1} \\ \inf_{\mu\in\mathbb{G}_{\sigma,f}} \int_\Omega \left(\|w\|_{\ell_1} + |b|\right)^s d|\mu|(w,b) & \sigma = \mathrm{RePU}_s \end{cases}$$

where the set $\mathbb{G}_{\sigma,f}$ contains all the measures μ that can be used to represent f using the activation function σ.

Our contribution

We have studied embeddings between Barron spaces with different activation functions. The main concept that we used is that of push-forwards between measures. These are maps Θ such that

$$f(x) = \int_\Omega \phi(\langle x|w\rangle + b)\, d\mu(w,b) = \int_\Omega \psi(\langle x|w\rangle + b)\, d\Theta_\#\mu(w,b)$$

for certain pairs of activation functions ϕ and ψ.

Main theorem:

Let $s \in \mathbb{N}$. If ϕ and ψ are Lipschitz activation functions such that

$$\phi(x) = \int_{\mathbb{R}^2} \psi(\langle x|w\rangle + b)\, d\gamma(w,b)$$

for all $x \in \mathbb{R}$ and for some $\gamma \in \mathcal{M}(\Omega)$ that satisfies

$$\int_{\mathbb{R}^2} \left(1 + \|w\|_{\ell_1} + |b|\right) d|\gamma|(w,b) < \infty,$$

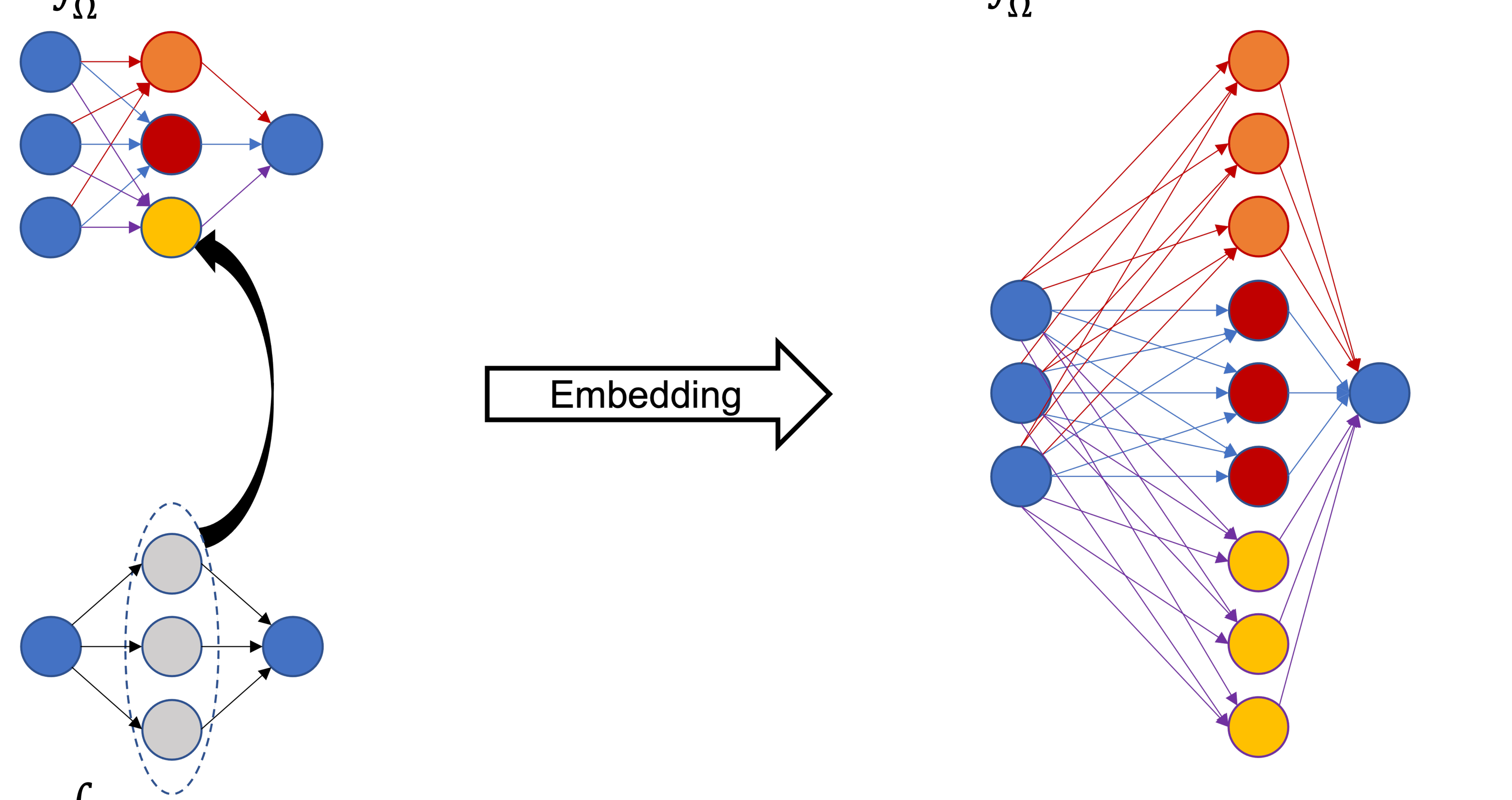
then

1. $\mathcal{B}_\phi \hookrightarrow \mathcal{B}_\psi$,
2. $\mathcal{B}_\psi \hookrightarrow \mathcal{B}_{\mathrm{RePU}_1}$, whenever $\psi \in C^1(\mathbb{R})$ with $D^2\psi \in L^1(\mathbb{R})$,
3. $\mathcal{B}_\psi \hookrightarrow \mathcal{B}_{\mathrm{RePU}_s}$, whenever $\psi \in C^s(\mathbb{R})$ with $D^{s+1}\psi \in L^1(\mathbb{R})$ and Ω bounded,
4. $\mathcal{B}_{\mathrm{RePU}_s} \hookrightarrow \mathcal{B}_{\mathrm{RePU}_t}$ for all $t \in \mathbb{N}$ with $t \leq s$.

Interpretation

Since the functions we consider are shallow neural networks, our theorem tells us we can replace the activation function of a network by replacing each neuron by a suitable subnetwork.

$$f(x) = \int_\Omega \phi(\langle x|w\rangle + b)\, d\mu(w,b) \qquad f(x) = \int_\Omega \psi(\langle x|w\rangle + b)\, d\nu(w,b)$$



$$\phi(x) = \int_{\mathbb{R}^2} \psi(\langle x|w\rangle + b)\, d\gamma(w,b)$$

Future work

Neural networks are able to solve many PDEs fast up to a certain accuracy. Currently, there is no clear notion of regularity for neural networks which can be used to prove existence and uniqueness for PDEs when the initial and boundary conditions can be approximated using neural networks. Our main theorem implies that $\mathcal{B}_{\mathrm{RePU}_1} \hookleftarrow \mathcal{B}_{\mathrm{RePU}_2} \hookleftarrow \mathcal{B}_{\mathrm{RePU}_3} \hookleftarrow \cdots$. This hierarchy is the same as that of the continuous function spaces C^k and the Sobolev spaces H^m. This suggests the order of the RePU could be a suitable notion of reguarity.

Paper

https://arxiv.org/abs/2305.15839

References

1. https://en.wikipedia.org/wiki/Javan_leopard
2. https://nl.m.wikipedia.org/wiki/Bestand:Lorenz_attractor.svg
3. Suk et al., https://doi.org/10.1007/978-3-030-93722-5_11
4. https://www.hiclipart.com/free-transparent-background-png-clipart-pmzin/download
5. Andrei et al., https://doi.org/10.1016/j.neunet.2021.01.026
6. Noel et al., https://doi.org/10.48550/arXiv.2108.12943
7. Ziyin et al., https://doi.org/10.48550/arXiv.2006.08195
8. E et al., https://doi.org/10.1007/s11425-020-1773-8

Contact
Tjeerd Jan Heeringa
t.j.heeringa@utwente.nl

UNIVERSITY OF TWENTE.