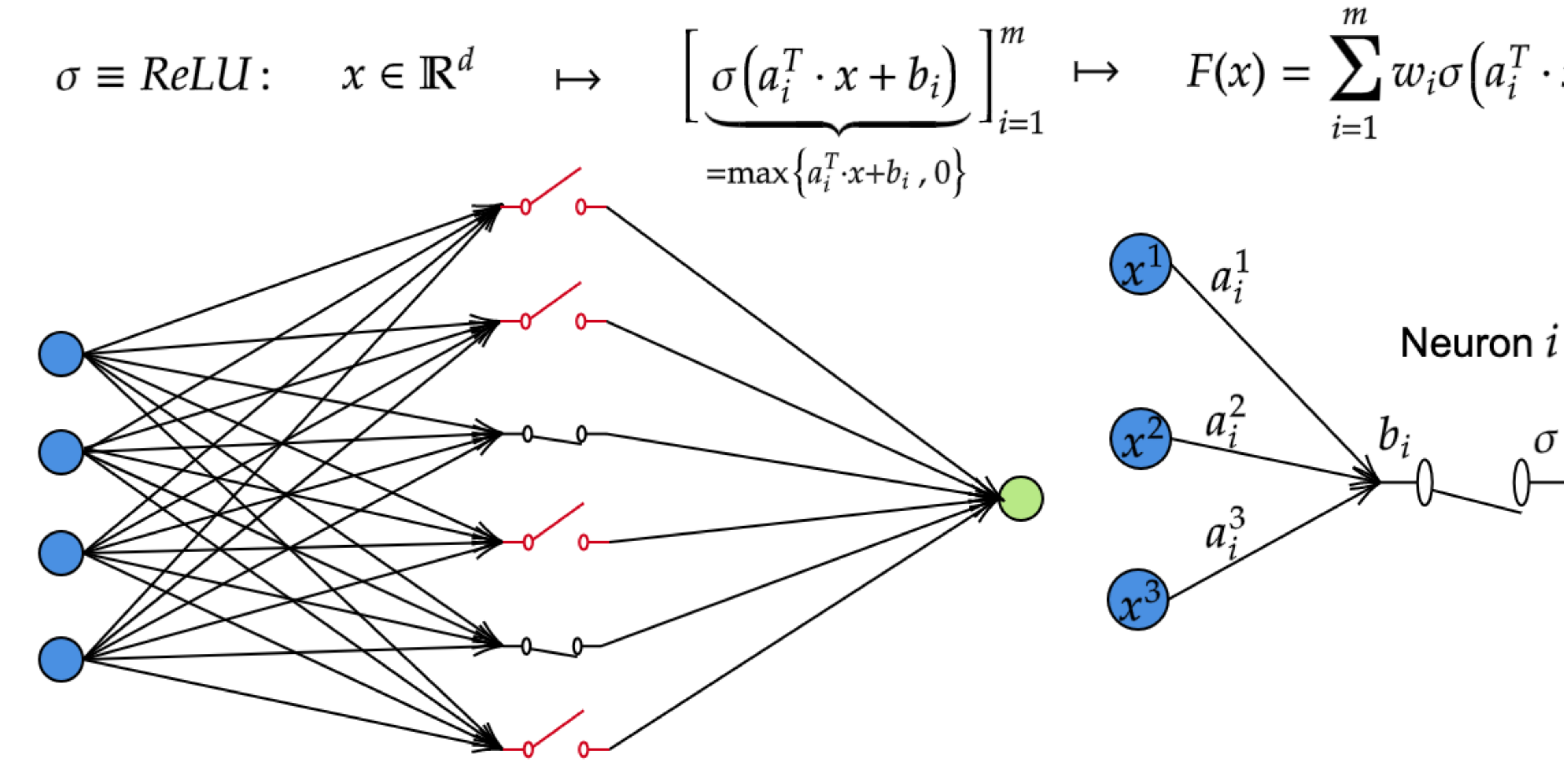
Friedrich-Alexander-Universität
DYNAMICS, CONTROL,
MACHINE LEARNING
AND NUMERICS

Breaking the curse of dimensionality with Barron spaces


$\sigma \equiv ReLU: \quad x \in \mathbb{R}^d \mapsto \Big[\underbrace{\sigma\big(a_i^T \cdot x + b_i\big)}_{=\max\{a_i^T \cdot x + b_i\,,\,0\}} \Big]_{i=1}^{m} \mapsto F(x) = \sum_{i=1}^{m} w_i \sigma\big(a_i^T \cdot$
x^1
x^2
x^3
a_i^1
a_i^2
a_i^3
Neuron i
b_i
σ

Published December 14, 2022

Breaking the curse of dimensionality with Barron spaces

By Antonio Álvarez López

1 Introduction

Recent advances in computational hardware have enabled the implementation of the set of algorithmic methods known as Deep Learning, whose development nevertheless dates back several decades. In this way, they have emerged in the latest years as the main tool in many practical Machine Learning problems, especially those belonging to the category of Supervised Learning that uses labelled datasets to train algorithms that classify data or predict outcomes accurately. Some of these applications are image analysis, natural language processing or sales forecasting.

However, the unparalleled superior performance of Deep Learning hasn't found a precise explanation yet. From a theoretical point of view, we still have not passed the initial stages of exploration. Broadly speaking, Deep Learning consists of learning complex representations by repeatedly passing inputs through the layers of a neural network to disentangle their higher-level features. One of the first steps in understanding the source of its power is to characterize the set of problems that a neural network model can learn efficiently.

ability to identify the class to which each element of a set belongs from a finite number of possibilities (classification problem), or to assign each element to a value from a continuous range (regression problem).

From a mathematical point of view, the problem seeks to approximate an unknown *objective function* F^* that can label any element (input vector) x of an *input space* $X \subset \mathbb{R}^d$ to its corresponding label y from the *output space* $Y \subset \mathbb{R}^m$. Because of this ideal characterization as an absolute knowledge, it is sometimes called the oracle function.

Remark 1.1
The dimension d is typically very large, while m might take lower values. In fact, we will often consider $m = 1$ throughout this text.

For this purpose, we can only use the information contained in a finite dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N \subset X \times Y$, which verifies that $y_i = F^*(x_i)$ for all $i = 1, \ldots, N$, and the goal is to obtain a good representation of F^* that can accurately predict the unknown label of any new input. But, how can we construct such a predictive model?

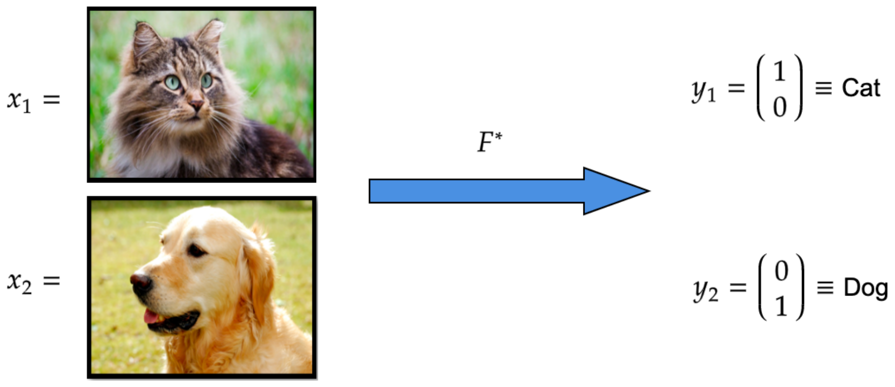

$x_1 =$
$x_2 =$
F^*
$y_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \equiv$ Cat
$y_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \equiv$ Dog

FIGURE 1. Example of an objective function F^* that recognizes whether a given picture from the input space is containing a dog or a cat. Observe that in this classification problem we have $m = 2$ while d takes an extremely large value: the input has one component for each of the pixels of the photograph, and there are three extra dimensions for each pixel to represent its color.

1.2 Learning procedure

The process to construct the predictive model $\hat{F}$ is called the *learning procedure*. However, it is not clear where we can even look for it. Consequently, this motivates the first step of the method: limiting the search to a fixed parameterized class of functions called the *hypothesis space* $\mathcal{H}_\theta$ and choosing inside it the function $\hat{F} \in \mathcal{H}_\theta$ that "best approximates" F^* as our learned model.

How can we take $\hat{F}$ among all the functions in the hypothesis space? This is the second step of the learning procedure, and consists in the minimization of a functional that measures the closeness between F^* and the possible models $F_\theta \in \mathcal{H}_\theta$.

Ideally, the best possible predictor in $\mathcal{H}$ would be obtained through *population risk minimization*:

$$\operatorname{argmin}_{F_\theta \in \mathcal{H}_\theta} \mathbb{E}_{(x,y)\sim\mu^*} L(F_\theta(x), y),$$

where μ^* is the unknown assumed input-output distribution in $X \times Y$, and $L(\cdot, \cdot)$ is a suitable *loss function*.

In practice, there is no prior knowledge of μ^*, so the best possible predictor $\hat{F}$ is constructed using $\mathcal{D}$ through the so-

Typically, a regularization term depending on the norm of the parameters is added to the empirical loss to prevent overfitting, this is, the model performing perfectly in the dataset but predicting very poorly in unseen examples.

The third step of the learning procedure is choosing the algorithm to solve the minimizing problem, being the Stochastic Gradient Descent the most common choice.

Having described the learning procedure, it is clear that there are three theoretical paradigms corresponding to the three steps:

- **Approximation**: Is the chosen hypothesis space $\mathcal{H}$ large enough to get as close as desired to any given target function F^*? The distance separating F^* and the set $\mathcal{H}$ is called the *approximation error.*
- **Optimization**: How can we ensure that our optimization algorithm converges to $\hat{F}$? The *optimization error* measures the distance from $\hat{F}$ to the predictor that we obtain when we stop running the algorithm.
- **Generalization**: How can we estimate the performance of the constructed model $\hat{F}$ on unseen data if only the limited training data is available? This gives rise to the *generalization error*, mainly rooted in the finiteness of $\mathcal{D}$, which causes the difference between the population and empirical risk minimization.

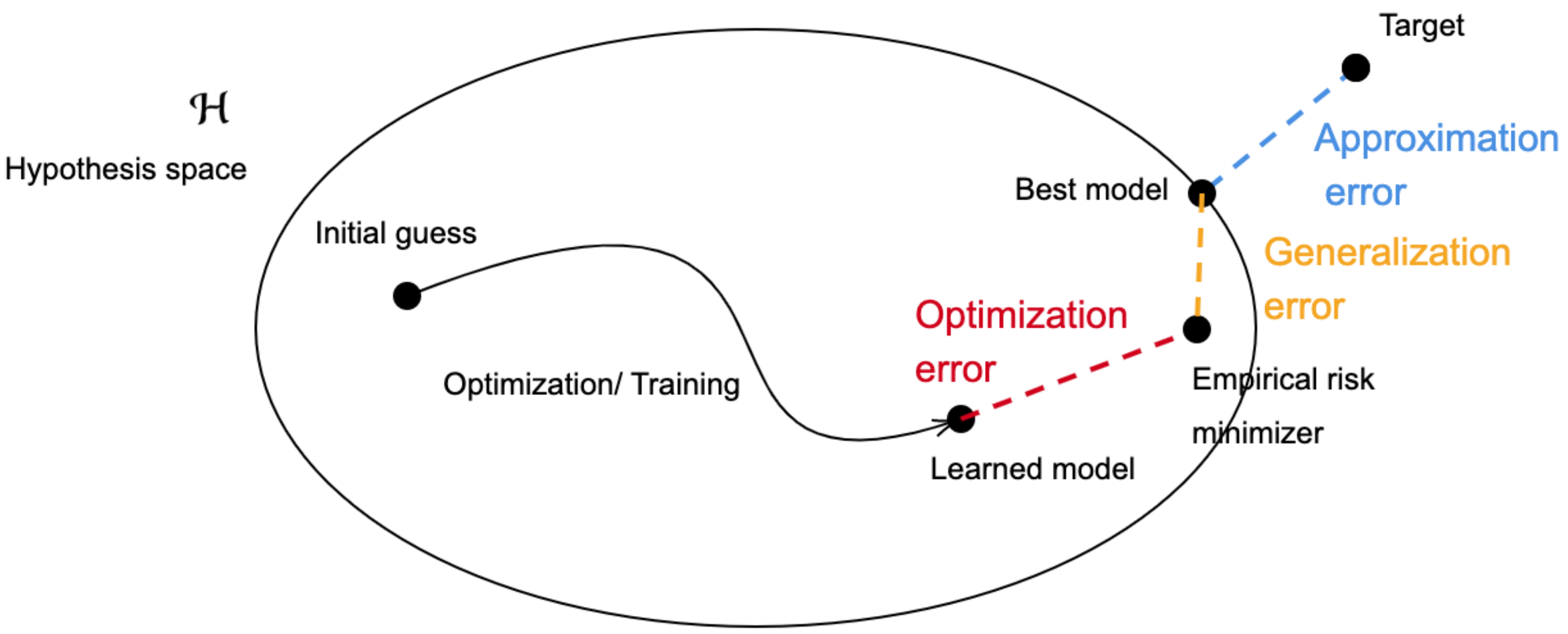


FIGURE 2. Diagram representing the learning procedure, the three main paradigms and their corresponding errors.

2 Shallow Neural Networks

We shall focus from now on on the first paradigm: the approximation of an objective function using a suitable parametric space. We introduce *neural networks* (NNs) as a class of hypothesis spaces whose name and structure are inspired by the human brain. They where first developed as an attempt to mathematically model the way in which biological neurons interact with each other through electrical nerve signals. It is clear that such models are oversimplified with respect to the complex human physiology, but they still form a class of powerful machine learning models showing unique properties.

The simplest form of neural networks are called *Shallow Neural Networks (SNNs)*, which are defined as the following parametric hypothesis space:

$$\text{ReLU: } \sigma(z) = \max(0, z), \qquad \sigma(z) = \tanh(z), \qquad \sigma(z) = \frac{1}{1+\exp -z}).$$

The ReLU activation function is generally the most common choice, and we will focus on it throughout this post. The parameters w_i are referred to as weights, while the other parameters a_i and b_i are respectively called the coefficients and the biases. Finally, the index M is the *width* of the SNN and controls the complexity of the model. To mention,

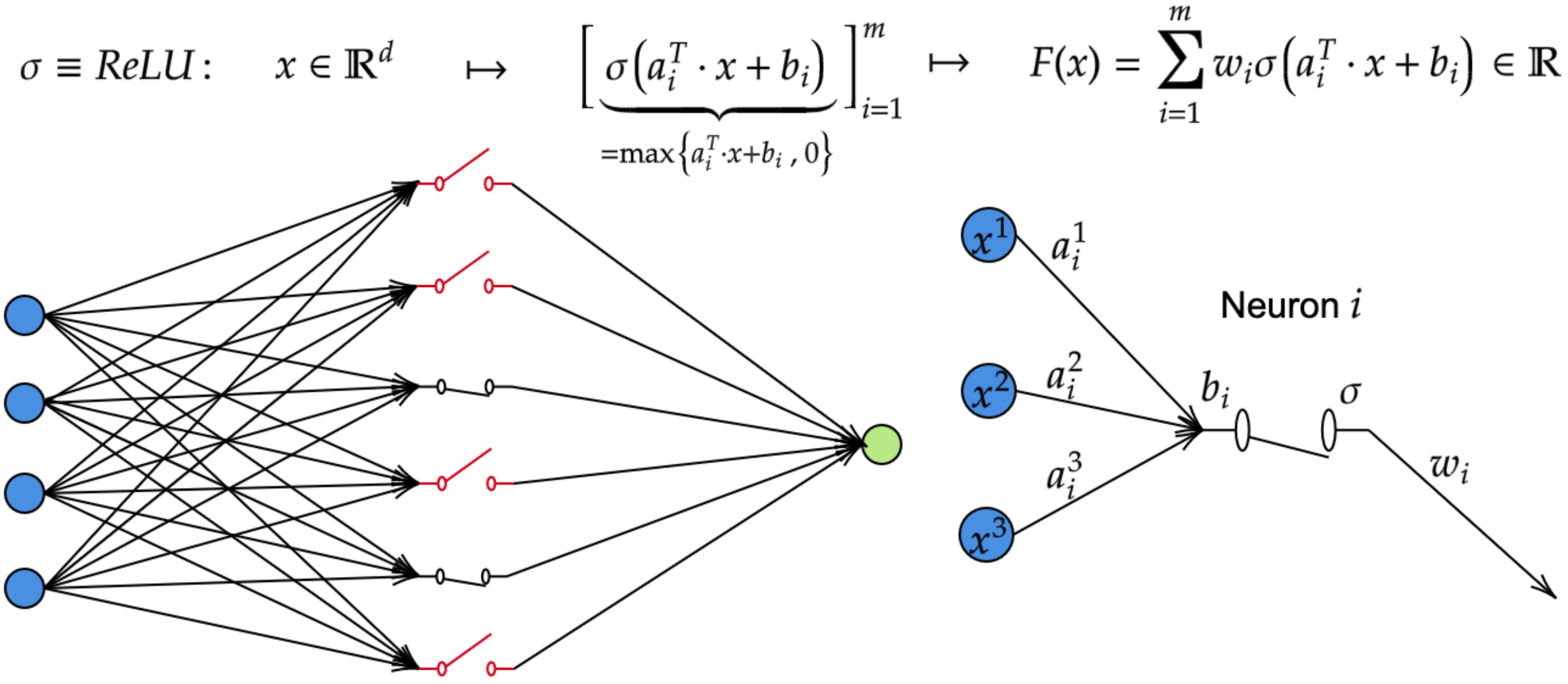

$\sigma \equiv ReLU: \quad x \in \mathbb{R}^d \quad \mapsto \quad \Big[\underbrace{\sigma\big(a_i^T \cdot x + b_i\big)}_{=\max\{a_i^T \cdot x + b_i\,,\,0\}}\Big]_{i=1}^{m} \quad \mapsto \quad F(x) = \sum_{i=1}^{m} w_i \sigma\big(a_i^T \cdot x + b_i\big) \in \mathbb{R}$
x^1
a_i^1
Neuron i
x^2
a_i^2
b_i
σ
a_i^3
w_i
x^3

FIGURE 3. On the left, representation of a SNN as a network composed of several interconnected simple units called *neurons* distributed in three layers: (from left to right) the *input layer*, the *hidden layer* and the *output layer* (from left to right). On the right, diagram showing the operations that take place on each particular neuron.

There are more complex classes of neural networks. Deep Neural Networks (DNNs) refer to those which are constructed as the composition of several simple architectures, like SNNs, by increasing their number of hidden layers, that we call the depth of the model. Two of the most popular choices are:

- *Multilayer Perceptron* (MLP) of *depth* K:

$$\mathcal{H}_K = \left\{F_K : F_K(\mathbf{x}) = \mathbf{w}^{\mathbf{T}}\mathbf{x}(K),\ \mathbf{w} \in \mathbb{R}^{d_K}\right\}$$

being x(K) the final point of the discrete dynamical system given by

$$\mathbf{x}(k+1) = \mathbf{w}(k)\sigma\left(\mathbf{a}(k)^T\mathbf{x}(k) + b(k)\right), x(0) = x, fork = 0, ..., K-1, where \qquad \mathbf{w}(k) \in \mathbb{R}^{d_{k+1}}, \qquad \mathbf{a}(k) \in \mathbb{R}^{d_k}, \qquad b(k) \in \mathbb{R},$$

and $d_0 = d, d_k \in \mathbb{N}$ for all

- *Residual Neural Networks*, obtained by redefining the dynamical system of the MLP as

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{w}(\mathbf{k})\sigma\left(\mathbf{a}(k)^T\mathbf{x}(k) + b(k)\right).$$

2.1 Approximation properties

compact set of $\mathbb{R}^d$ with the topology of uniform convergence.

Theorem 2.1
Let $\Omega \subset \mathbb{R}^d$ be a compact set, and $F^* \in C(\Omega)$. Assume that the activation function σ is continuous and *sigmoidal*, i.e. $\lim_{z\to\infty} \sigma(z) = 1$, $\lim_{z\to-\infty} \sigma(z) = 0$. Then, for every $\epsilon > 0$ there exists $F_M \in \mathcal{H}$ such that

$$\|F_M - F^*\|_{C(\Omega)} = \max_{\mathbf{x}\in\Omega} |F_M(\mathbf{x}) - F^*(\mathbf{x})| < \epsilon. \qquad (1)$$

The proof follows from an application of the Hahn-Banach theorem and the Riesz-Markov representation theorem, together with an argument based on the non-degenerate properties of σ.

It can be easily seen that the activation function ReLU does not verify the hypotheses of the theorem as it is not bounded. Fortunately, different versions and generalizations of the above result have been proven. Here we show the following universal approximation theorem of Pinkus ([6]), which is striking for the equivalence it establishes:

Theorem 2.2
Let $\Omega \subset \mathbb{R}^d$ be a compact set and σ a continuous activation function. Then, $\mathcal{H}$ is dense in $C(\Omega)$ in the topology of uniform convergence if and only if σ is non-polynomial.

2.2 Curse of dimensionality

Now that we know that a continuous function defined on a compact set can be approximated by an SNN with an arbitrarily small error (and thus, any measurable function too, due to Lusin's Theorem), one may mistakenly think that this is already sufficient to ensure the existence of methods that converge to a continuous objective function in a feasible way.

It turns out that this is not true: there is a type of difficulties that arise when working with high-dimensional data which people refer to as the "curse of dimensionality". They usually express the difficulties and enormous increase in the computational efforts required for the processing and analysis of a dataset due to mathematical phenomena like the concentration of distances, but also caused by the explosive tendencies of certain functions or parameters when the dimension of the input space increases.

In our case, the curse of dimensionality is exhibited on the complexity of our SNN (i.e., its width M), which depends exponentially dependence on the dimension of the input space in the expression of the approximation error bound in L^p norm:

$$\|F_M - F^*\|_{L^p(\Omega)} \leq C \frac{\|F^*\|_{W^{s,p}(\Omega)}}{M^{\alpha(s)/d}}, \qquad (2)$$

being $C > 0$ a constant that depends on Ω and the Sobolev parameters (s, p). From (2), it follows that we need a width of $M \gtrsim \epsilon^{d\alpha}$ to obtain an approximation error of ϵ, meaning that the complexity of the necessary approximation SNN becomes enormous in high dimensions.

Several approaches can be taken to overcome this difficulty, but the most basic one is to consider a smaller set of problems until we identify the functions that require a complexity whose dependence with d for fixed ϵ is at most polynomial: it is a matter of finding the "right" space of functions to approximate.

This correspondence of an approximation method with a space of functions where the algorithm is well suited a clear analog in classical numerical analysis. Finite element methods and splines are two schemes that use piecewise convergence rate if and only if it lies in a certain Sobolev or Besov space.

Barron spaces

Barron spaces are constructed by trying to mimic the mentioned approach and find the functions that can be approximated by "well-behaved" two-layer neural networks.

Let $\Omega \in \mathbb{R}^d$ be a compact set and σ $\equiv$ReLU. We consider the functions $f : \Omega \to \mathbb{R}$ that admit the representation

$$f(\mathbf{x}) = \int_\Theta w\sigma(\mathbf{a}^T\mathbf{x} + c)\rho(dw, \mathbf{da}, db) = \mathbb{E}_\rho[w\sigma(\mathbf{a}^T\mathbf{x} + b)], \qquad \mathbf{x} \in \Omega, \qquad (3)$$

where $\Theta = \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$ space of parameters and ρ is a probability distribution on (Θ, Σ_Θ), being Σ_Θ a Borel σ-algebra on Θ. We can intuitively think of ρ as an hypothetical unknown probability distribution of the parameters that is the limit to which converges the sum of atomic measures defined by the specific finite set of parameters of a DNN when its width (or, equivalently, the number of parameter samples) tends to infinity. With the expression (3), approximating F^* with a SNN becomes a Montecarlo integration problem.

This interpretation is depicted in the picture below.

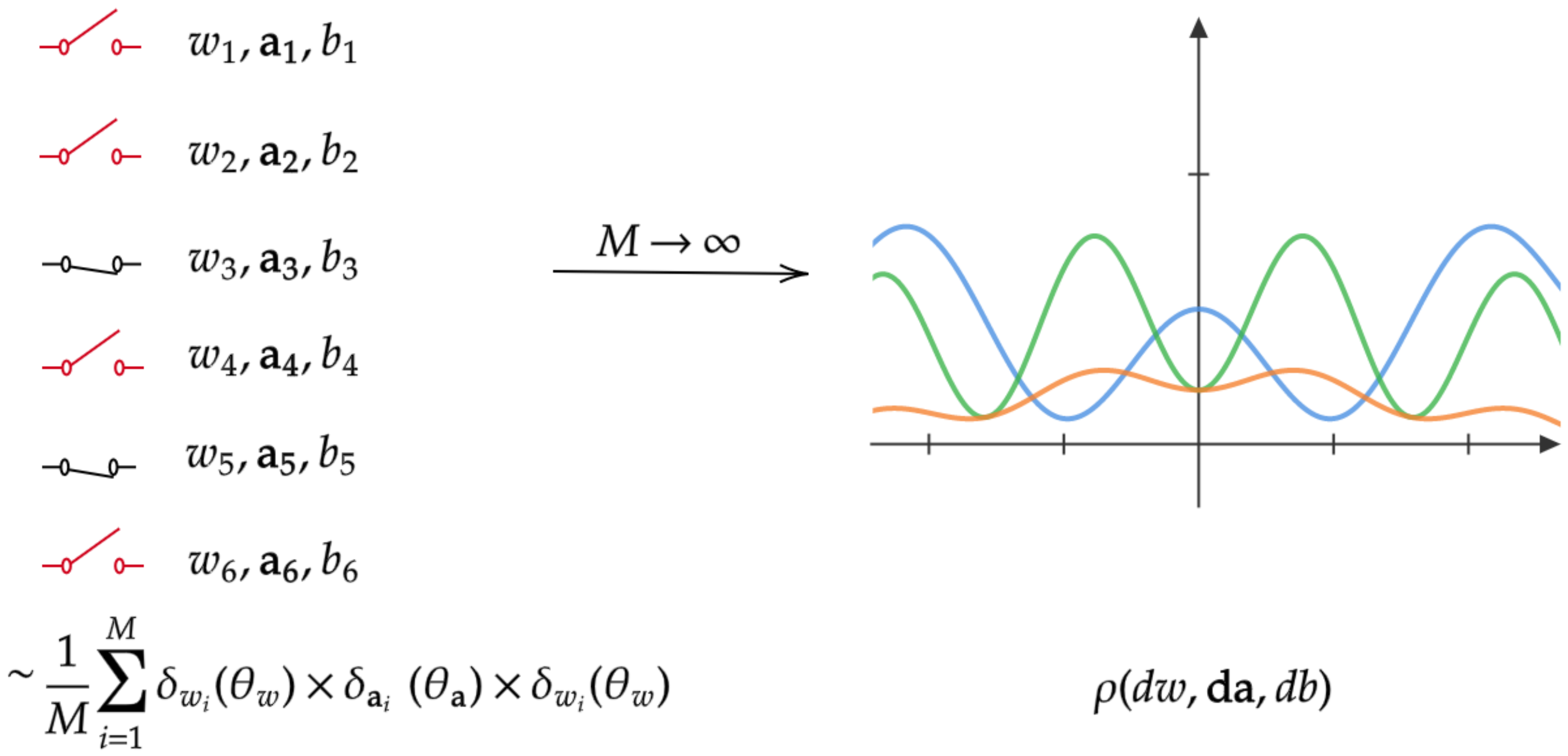

$w_1, \mathbf{a}_1, b_1$
$w_2, \mathbf{a}_2, b_2$
$w_3, \mathbf{a}_3, b_3$
$w_4, \mathbf{a}_4, b_4$
$w_5, \mathbf{a}_5, b_5$
$w_6, \mathbf{a}_6, b_6$
$M \to \infty$
$\sim \frac{1}{M}\sum_{i=1}^{M} \delta_{w_i}(\theta_w) \times \delta_{\mathbf{a}_i}(\theta_\mathbf{a}) \times \delta_{w_i}(\theta_w)$
$\rho(dw, \mathbf{da}, db)$

FIGURE 4. Diagram representing the probability distribution ρ on the space of parameters as the continuum width analog of SNNs.

In general, the measure ρ for which the representation (3) holds is not unique. For a function that admits this representation, we define for $1 \le p \le \infty$ its *p-Barron norm* $\|\cdot\|_{\mathcal{B}_p}$ as

$$\|f\|_{\mathcal{B}_p} = \inf_\rho \left(\mathbb{E}_\rho[|w|^p(\|\mathbf{a}\|_1 + |b|)^p]\right)^{1/p}, \qquad 1 \le p \le \infty, \qquad (4)$$

Note that, by Hölder's inequality, it is straightforward that

$$\mathcal{B}_\infty \subset \cdots \subset \mathcal{B}_2 \subset \mathcal{B}_1.$$

To see this, given $p \leq q$, we have that, for all $f \in C(\Omega)$,

$$\left(\int_\Theta |a|^p d\rho\right)^{1/p} \leq \left(\int_\Theta |a|^q d\rho\right)^{1/q} \left(\int_\Theta d\rho\right)^{\frac{q-p}{q}} \Rightarrow \|f\|_{\mathcal{B}_p} \leq \|f\|_{\mathcal{B}_q}.$$

The opposite inclusions are also true (see [1]):

Proposition 1

For any $f \in \mathcal{B}_1$, we have $f \in \mathcal{B}_\infty$ and

$$\|f\|_{\mathcal{B}_1} = \|f\|_{\mathcal{B}_\infty}.$$

As a consequence, we have that for any $1 \leq p < \infty$, $\mathcal{B}_p = \mathcal{B}_\infty$ and $\|f\|_{\mathcal{B}_p} = \|f\|_{\mathcal{B}_\infty}$, this is, there is just one Barron space and one Barron norm that we denote by $\mathcal{B}$ and $\|\cdot\|_{\mathcal{B}}$, respectively.

Having defined the Barron space, it is natural to think what kind of functions belong to it. To answer this question, it is necessary to refer to Barron's seminal paper of 1993 [1], in which he showed that functions verifying an spectral regularity property would improve the approximation bound by SNNs in L^2 helping to partially overcome the curse of dimensionality.

Before stating an extended version of its main result (proven in [2]), we recall one basic definition in harmonic analysis:

The *Fourier transform* of a function $f : \mathbb{R}^d \to \mathbb{R}$ is

$$\hat{f}(\xi) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-i\xi\cdot\mathbf{x}} \mathbf{dx}.$$

and verifies the property $\widehat{Df}(\xi) = i\xi\hat{f}(\xi)$, which means it transforms derivatives in products.

Theorem 3.1

For a function $F^* : \Omega \to \mathbb{R}$, let $\hat{F}^*$ be the Fourier transform of any extension of F^* to $\mathbb{R}^d$. If

$$\gamma(F^*) := \inf_{\hat{F}} \int_{\mathbb{R}^d} \|\xi\|_1^2 |\hat{F}^*(\xi)| d\xi = \|\widehat{D^2 F^*}\|_1 < +\infty,$$

then for any $M > 0$ there exists a SNN $F_M(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M w_i \sigma(\mathbf{a_i}^T \mathbf{x} + b_i)$ satisfying

$$\|F_M - F^*\|_{L^2(\Omega)}^2 \leq \frac{3\gamma(F^*)^2}{M},$$

and $\frac{1}{M} \sum_{i=1}^M |w_i|(\|\mathbf{a_i}\|_1 + |b_i|) \leq \frac{2\gamma(F^*)}{M}$.

As we can see, the theorem states that the L^1-integrability of the Fourier transform of the second derivative of a function F^* is a sufficient condition to ensure the disappearance of the exponential dependence of the width M with the dimension in the error bound in norm L^2.

Note that the numerator $\gamma(F^*)$ is an integral over $\mathbb{R}^d$ and could therefore bring back to the bound an exponential dependence on dimension. This means that the finiteness of $\gamma(F^*)$ if generally not enough to avoid the curse of

$$\|\theta\|_{\mathcal{P}} := \frac{1}{M}\sum_{i=1}^{M} |w_i|(\|\mathbf{a_i}\|_1 + |b_i|),$$

where θ denotes its correspondent set of parameters $\{(w_i, \mathbf{a}_i, b_i)\}_{i=1}^{M}$. As the analog of $\|\cdot\|_{\mathcal{B}}$, it is natural to consider two-layer neural networks with bounded path norm when studying their approximation properties.

The subsequent theorem finally sheds some light on a class of functions that belong to $\mathcal{B}$:

Theorem 3.2

Let $F^* \in C(\Omega)$ and assume that F^* satisfies $\gamma(F^*) < \infty$. Then F^* admits an integral representation (3). Moreover,

$$\|F^*\|_{\mathcal{B}} \leq 2\gamma(F^*) + 2\|\nabla F^*(0)\|_1 + 2|F^*(0)|.$$

A necessary condition for a function $F*$ to verify that $\gamma(F^*) < \infty$ is the boundedness of all its first order partial derivatives, which constitutes a big restriction for those functions. On the other hand, it is enough for a function to satify $\gamma(F^*) if F^*$ has all partial derivatives of order less or equal than s in the space $L^2(\mathbb{R}^d)$, being $s = \lceil 1 + d/2 \rceil$.

By finiteness of $\gamma(F^*)$, it is immediate the following result:
Corollary 1
All **gaussian** functions, **positive definite** functions, **linear** functions and **radial** functions belong to $\mathcal{B}$.

Having defined the Barron space and characterized some functions that belong to it, we state the two results, extracted from [5] and proven there, that identify this space as the appropriate one to be approximated by SNNs.

Theorem 3.3

(Theorem of direct approximation)
For any $F^* \in \mathcal{B}$ and $M > 0$, there exists a SNN $F_M(\mathbf{x}) = \frac{1}{M}\sum_{i=1}^{M} w_i \sigma({\mathbf{a_i}}^T \mathbf{x} + b_i)$ satisfying

$$\|F_M(\cdot;\theta) - F^*(\cdot)\|_{L^2(\Omega)}^2 \leq \frac{3\|F^*\|_{\mathcal{B}}^2}{M}.$$

Furthermore, we have $\|\theta\|_{\mathcal{P}} \leq 2\|F^*\|_{\mathcal{B}}$.

To state the reciprocal final result, which is very important as it states the characterization of $\mathcal{B}$ as the right approximation function space with respect to the path norm, we define $\mathcal{N}_Q := \{F_M(\cdot;\theta) : \ \|\theta\|_{\mathcal{P}} \leq Q, m \in \mathbb{N}^+\}$.

Theorem 3.4

(Theorem of inverse approximation)
Let $F^* \in C(\Omega)$. Assume there exists a constant Q and a sequence of functions $(F_M) \subset \mathcal{N}_Q$ such that

$$\lim_{M\to\infty} F_M(\mathbf{x}) = F^*(\mathbf{x}),$$

for all $\mathbf{x} \in \Omega$. Then $F^* \in \mathcal{B}$ and $\|F^*\|_{\mathcal{B}} \leq Q$.

We give a brief idea of the proof:

is tight. By Prokhorov's Theorem, there exists a subsequence (ρ_{M_k}) and a probability measure ρ^* such that ρ_{M_k} converges weakly to ρ^*. This measure ρ^* on the parameter space Θ is the one associated to F^* that makes it belong to $\mathcal{B}$ with a Barron norm bounded by Q.

4 Conclusions and open questions

The Barron space $\mathcal{B}$ catches all the functions that can be approximated by SNNs with bounded path norm, and the bound of the corresponding error partially overcomes the curse of dimensionality.
Due to the statement of the two theorems, $\mathcal{B}$ can be seen as the closure of the space of DNNs $\mathcal{H}_\Theta$ with respect to the path norm, which means it is the largest function space that is well approximated by SNNs, and the Barron norm is the natural norm associated to it. Target functions outside $\mathcal{B}$ may be increasingly difficult to approximate by SNNs as dimension increases.

We end this post by listing three open paths for research on the subject:
1. More specific descriptions of the functions that belong to $\mathcal{B}$.
2. Further study on the possibilities and conditions for PDE solutions to lie on the Barron space (see [3]). The goal is the application of SNNs to approximate the solutions of high-dimensional PDEs.
3. Extension of this theory to more complex architectures like Deep Neural Networks. In the specific case of Residual Neural Networks, there is current research on this direction (see [5]).
\end{enumerate}

References

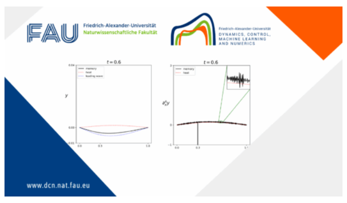
[1] A. Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: IEEE Transactions on Information Theory 39.3 (1993), pp. 930– 945. doi: 10.1109/18.256500.
[2] L. Breiman. "Hinging hyperplanes for regression, classification, and function approximation". In: IEEE Transactions on Information Theory 39.3 (1993), pp. 999–1013. doi: 10.1109/18.256506.
[3] Z. Chen, J. Lu, and Y. Lu. "On the representation of solutions to elliptic pdes in barron spaces". In: Advances in neural information processing systems 34 (2021), pp. 6454–6465.
[4] G. Cybenko. "Approximation by superpositions of a sigmoidal function". In: Mathematics of control, signals and systems 2.4 (1989), pp. 303–314.
[5] C. Ma, L. Wu, et al. "The Barron space and the flow-induced function spaces for neural network models". In: Constructive Approximation 55.1 (2022), pp. 369– 406.
[6] A. Pinkus. "Approximation theory of the MLP model in neural networks". In: Acta Numerica 8 (1999), pp. 143–195.

|| Go to the Math & Research main page

You might like!

FAU Friedrich-Alexander-Universität Research Center for Mathematics of Data | MoD
FAU MoD Lecture Series
Optimization-based control for large-scale and complex systems: When and why does it work?
Lars Grüne
UNIVERSITÄT BAYREUTH
TUE. JUNE 3, 2025 • 15:00H
WWW.MOD.FAU.EU

FAU MoD Lecture: Optimization-based control for large-scale and



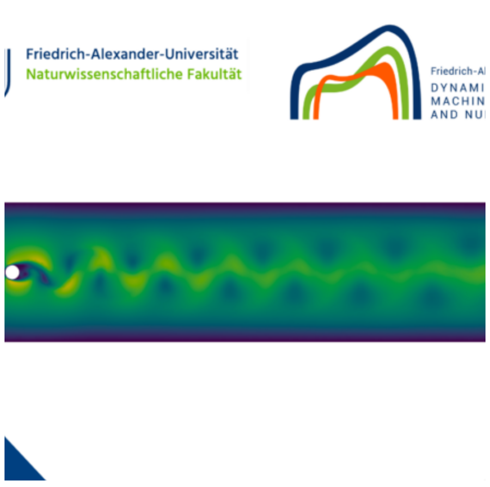
Hybrid parabolic-hyperbolic effect for heat equations

AI for maths and maths for AI

Interior Control in a Diffusive Lotka-volterra Model

Towards a Mathematical Foundation of Deep Learning: From Phenomena to Theory

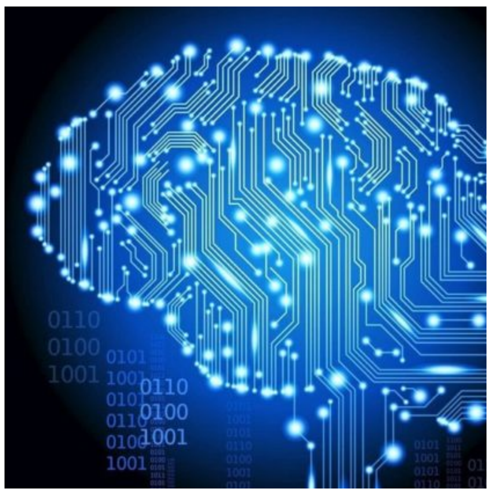
YOU MAY ALSO LIKE



Published April 1, 2023

Convolutional autoencoders for very low-dimensional parametrizations of nonlinear fluid flow

Convolutional autoencoders for very low-dimensional parametrizations of nonlinear fluid flow 1. Introduction The control of large-dimensional nonlinear dynamical



Published October 1, 2020

Neural networks and Machine Learning

Neural networks and Machine Learning By Marius Yamakou Neural Networks with time delayed connections Neurons communicate with each other through electrical signals. […]

‹ DTN MODCONFLEX. MODELLING AND CONTRO…

FEDERATED LEARNING: PROTECT YOUR DATA… ›

FAU Friedrich-Alexander-Universität Naturwissenschaftliche Fakultät

Last news

FAU MoD Lecture: Optimization-based control for large-scale and complex systems: When and why does it work?

Hybrid parabolic-hyperbolic effect for heat equations with memory

FAU MoD Lecture: AI for maths and maths for AI

Boundary and Interior Control in a Diffusive Lotka-volterra Model

FAU MoD Course Towards a Mathematical Foundation of Deep Learning: From Phenomena to Theory

Alexander von HUMBOLDT STIFTUNG