Columbia University

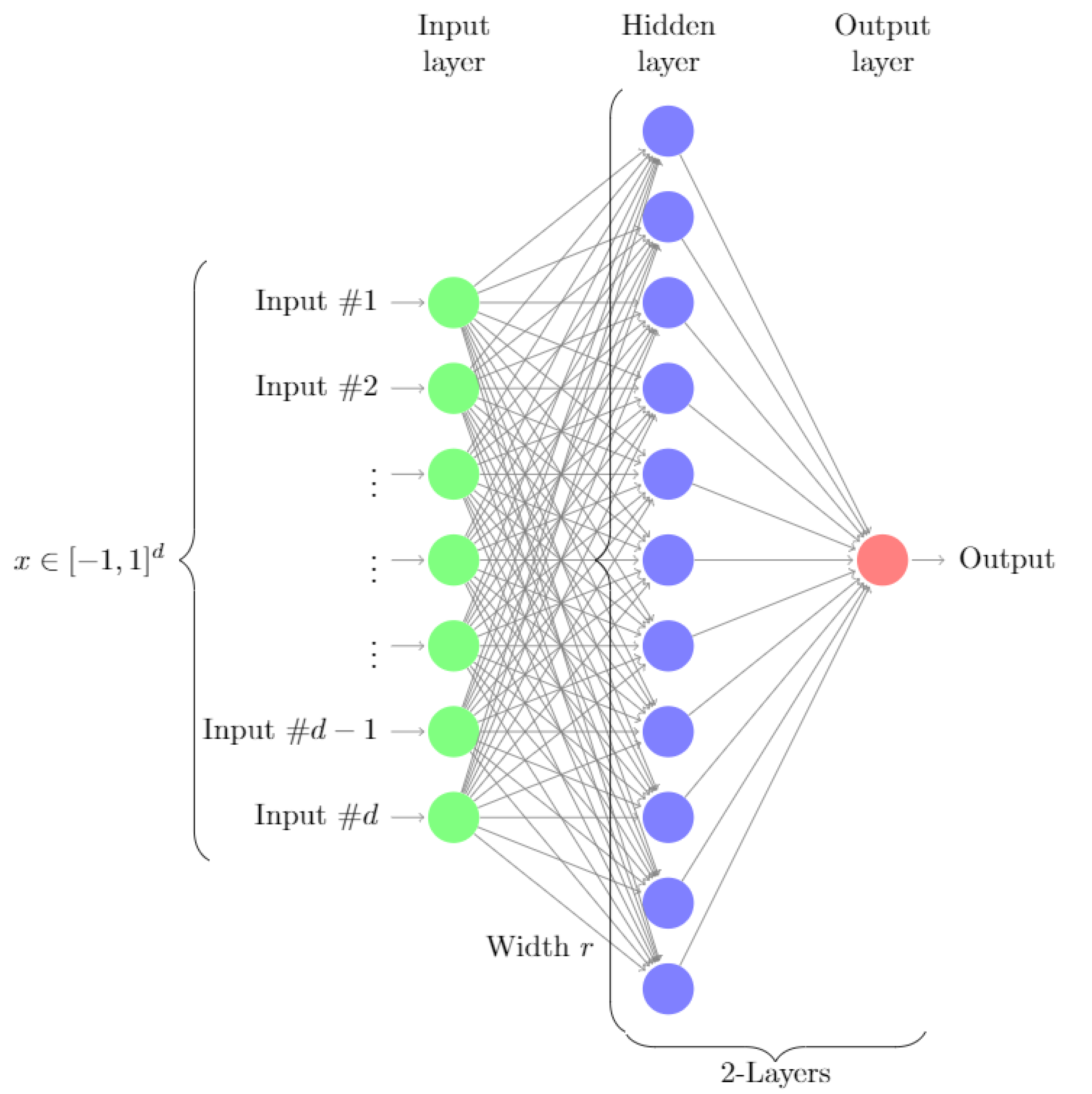
On the Approximation Power of Two-Layer Networks of Random ReLUs

Daniel Hsu, Clayton Sanford, Rocco A. Servedio, Emmanouil V. Vlatakis-Gkaragkounis
Columbia University

Conference on Learning Theory 2021

Our Problem

What are the approximation powers/limits of depth-2 NNs with ***randomly****-choosen bottom layer weights?*



Our Setting

Goal: *Find a 2-layer random bottom-layer (RBL) ReLU neural net that ε-approximates L-Lipschitz $f \in \mathcal{L}_2([-1,1]^d])$ w.h.p.*

- f belongs to $\mathcal{L}_2([-1,1]^d])$ if $\|f\|_2 \equiv \left(\int_{[-1,1]^d} f(x)^2 \, \mathrm{d}\mu(x)\right)^{1/2} < \infty$
- f is **L**−Lipschitz if for all $x, x' \in [-1,1]^d$,
$$|f(x) - f(x')| \leq L\|x - x\|_2.$$
- A 2-layer random bottom-layer (RBL) ReLU neural net can be written as
$$g(x) = \sum_{i=1}^{r} u^{(i)} \sigma(\langle \mathbf{w}^{(i)}, x\rangle - \mathbf{b}^{(i)}) \text{ for } (\mathbf{w}^{(i)}, \mathbf{b}^{(i)}) \sim \mathcal{D}$$
- g ε-approximates f if and only if $\|f - g\|_2 = \sqrt{\mathbb{E}[(f-g)^2(x)]} \leq \varepsilon$
- **MinWidth**$_{f,\varepsilon,\mathcal{D}}$ is the smallest r such that with prob. 0.9 over $(\mathbf{w}^{(i)}, \mathbf{b}^{(i)})_{i\in[r]}$, there exists an RBL g that ε-approximates f.

Our Contributions

We provide **necessary and sufficient conditions** to ε-approximate an L-Lipschitz function $f \in \mathcal{L}_2([-1,1]^d)$:

- **MinWidth**$_{f,\varepsilon,\mathcal{D}}$ is poly(d) if $L/\varepsilon = \Theta(1)$
- **MinWidth**$_{f,\varepsilon,\mathcal{D}}$ is poly(L/ε) if $d = \Theta(1)$
- **MinWidth**$_{f,\varepsilon,\mathcal{D}}$ is $\exp(\Theta(d))$ if $L/\varepsilon = \Theta(\sqrt{d})$

Previous Work on Width of Depth-2 NNs

Upper bounds

- Universal approx. theorem [Cybenko '89, Funahashi '89, HSW '89].
- $\mathcal{L}_2$ approx. for bounded Fourier coefficients [Barron '93].
- $\mathcal{L}_2$ approx. with width $d^{\mathcal{O}(L^2/\varepsilon^2)}$ [APVZ '14].
- $\mathcal{L}_\infty$ approx. with width $(L/\varepsilon)^{\mathcal{O}(d)}$ [Bach '17].

Lower bounds

- $\exp(d)$ for Sobolev smooth with small ε [Maiorov '99].
- $\exp(d)$ for poly(d)-Lipschitz using *RBL nets* [YS '19, KMS '20].

Our Results

- $Q_{k,d} = \left|\{K \in \mathbb{Z}^d : \|K\|_2 \leq k\}\right| = \binom{k^2+d}{d}^{\Theta(1)}$

Theorem (Upper Bound)

For any L, d, ε, there exists symmetric $\mathcal{D}$ such that for all L-Lipschitz $f \in \mathcal{L}_2([-1,1]^d)$ with $|\mathbb{E}[f]| \leq L$:
$$\textbf{MinWidth}_{f,\varepsilon,\mathcal{D}} \leq Q_{2L/\varepsilon,d}^{\mathcal{O}(1)}$$

Theorem (Lower Bound)

For any L, d, ε, and any symmetric $\mathcal{D}$, there exists an L-Lipschitz $f(x) = \sin(L\langle u, x\rangle)$ such that:
$$\textbf{MinWidth}_{f,\varepsilon,\mathcal{D}} = \Omega\left(Q_{L/18\varepsilon,d}\right)$$

Applications

Depth Separation

- Question from [SES '19]: Are there 1-Lipschitz functions that separate poly-size depth-2 NNs from poly-size depth-3 NNs?
- Our answer: Not for $\mathcal{L}_2$-approximation.

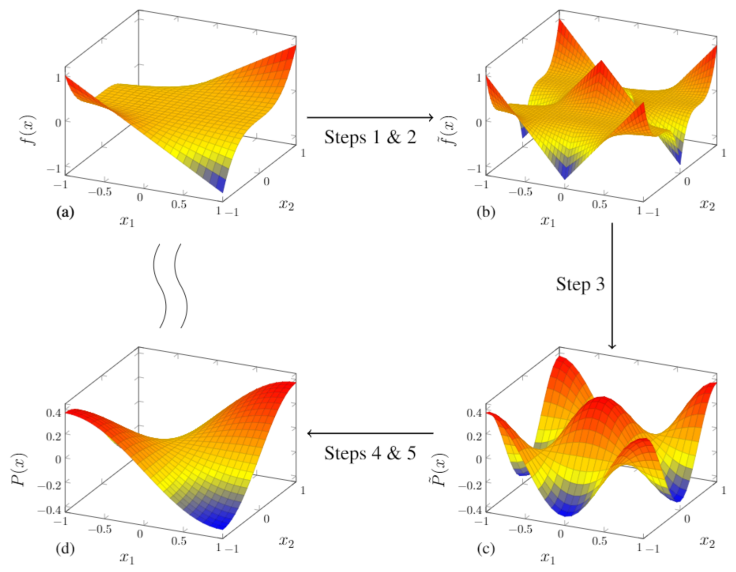
Learnability

- [MYSS '21]: Hardness of approximation with depth-3 ⇒ hardness of learning with any poly-size NN.
- Our bounds can strengthen statement to "hardness of approximation with depth-2."

Upper Bound Sketch

Every L-Lipschitz f can be ε-approximated by a trig. polynomial of degree $\mathcal{O}(L/\varepsilon)$.

$\exists$ symmetric $\mathcal{D}_k$ such that every k-degree trig. polynomial P has **MinWidth**$_{P,\varepsilon,\mathcal{D}} = Q_{k,d}^{\mathcal{O}(1)}$.



Lower Bound Sketch

For orthonormal $\varphi_1, \cdots, \varphi_N \in \mathcal{L}_2([-1,1]^d)$ and $N \gg r$, then at least one φ_i will be ***inapproximable*** *by the span of r functions.*

The family of $\mathcal{F}_k = \{x \mapsto \sqrt{2}\sin(\pi\langle K, x\rangle) : \|K\|_2 \leq k\}$ contains $\Theta(Q_{k,d})$ orthonormal $\Theta(k)$-Lipschitz functions.

Full version

- For more details, explicit lower-bounds and replacing Lipschitzness with Sobolev smoothness, **check**:

https://arxiv.org/abs/2102.02336