Minimum Width for Universal Approximation using ReLU Networks on Compact Domain

Namjun Kim[1] Chanho Min[2] Sejun Park[1]

[1]Korea University [2]Ajou University

ICLR 2024

Universal Approximation

- Neural networks are universal approximators if

For any continuous f^*, for any $\varepsilon > 0$, there exists a neural network f such that

distance (target f^*, NN f) $\leq \varepsilon$

Universal Approximation

- Neural networks are universal approximators if

For any continuous f^*, for any $\varepsilon > 0$, there exists a neural network f such that

distance (target f^*, NN f) $\leq \varepsilon$

- Popular distances:

L^p distance	Uniform distance
$\|f^* - f\|_p := (\int \|f^*(x) - f(x)\|_p^p dx)^{1/p}$	$\|f^* - f\|_\infty :=$ sup $\|f^*(x) - f(x)\|_\infty$

Universal Approximation

- Neural networks are universal approximators if

For any continuous f^*, for any $\varepsilon > 0$, there exists a neural network f such that

distance (target f^*, NN f) $\leq \varepsilon$

- Popular distances:

L^p distance

$$\|f^* - f\|_p := (\int \|f^*(x) - f(x)\|_p^p dx)^{1/p}$$

Uniform distance

$$\|f^* - f\|_\infty := \sup \|f^*(x) - f(x)\|_\infty$$

- Classical Results [Cybenko, 1989; Pinkus, 1999]:

Two-layer NNs are universal approximators for both L^p and uniform distances

I.e., the exact minimum depth for universal approximation is two

Universal Approximation

- Neural networks are universal approximators if

What is the minimum width for universal approximation?

- Classical Results [Cybenko, 1989; Pinkus, 1999]:

Two-layer NNs are universal approximators for both L^p and uniform distances

I.e., the exact minimum depth for universal approximation is two

Minimum Width for Universal Approximation

- The exact minimum width is only known for L^p distance and a few problem setups so far
  [Lu et al., 2017; Hanin & Sellke, 2017; Johnson, 2019; Kidger & Lyons, 2020; Park et al., 2021; Cai, 2023]

d_x, d_y: input, output dimensions of the target function

Reference	Distance	Domain	Activation	Exact Minimum Width
Park et al., 2021	L^p	$\mathbb{R}^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x + 1, d_y\}$
Cai, 2023	L^p	$[0, 1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$

Minimum Width for Universal Approximation

- The exact minimum width is only known for L^p distance and a few problem setups so far
  [Lu et al., 2017; Hanin & Sellke, 2017; Johnson, 2019; Kidger & Lyons, 2020; Park et al., 2021; Cai, 2023]

d_x, d_y: input, output dimensions of the target function

Reference	Distance	Domain	Activation	Exact Minimum Width
Park et al., 2021	L^p	$\mathbb{R}^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x + 1, d_y\}$
Cai, 2023	L^p	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$

- What is the minimum width for uniform distance?

Reference	Distance	Domain	Activation	Lower Bound
Park et al., 2021	Uniform	$[0,1]^{d_x}$	ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$
Cai, 2023	Uniform	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$

where $d_x = 1, d_y = 2$

Main Results

Contribution

1. A smaller width is sufficient to universally approximate target function on compact domain in L^p distance

d_x, d_y: input, output dimensions of the target function

Reference	Distance	Domain	Activation	Exact Minimum Width
Park et al., 2021	L^p	$\mathbb{R}^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x + 1, d_y\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$

Main Results

Contribution

1. A smaller width is sufficient to universally approximate target function on compact domain in L^p distance

d_x, d_y: input, output dimensions of the target function

Reference	Distance	Domain	Activation	Exact Minimum Width
Park et al., 2021	L^p	$\mathbb{R}^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x + 1, d_y\}$
Ours	L^p	$[0, 1]^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$

Can we generalize to other activation functions?

Main Results

Contribution

1. A smaller width is sufficient to universally approximate target function on compact domain in L^p distance

2. The exact minimum width for ReLU networks also holds for ReLU-Like networks

d_x, d_y: input, output dimensions of the target function

Reference	Distance	Domain	Activation	Exact Minimum Width
Cai, 2023	L^p	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU-Like	$w_{\text{min}} = \max\{d_x, d_y, 2\}$

ReLU-Like: Softplus, Leaky-ReLU, ELU, CELU, SELU, GELU, SiLU, and Mish

Main Results

Contribution

1. A smaller width is sufficient to universally approximate target function on compact domain in L^p distance

2. The exact minimum width for ReLU networks also holds for ReLU-Like networks

3. Extends the pervious lower bounds to continuous monotone activation functions and general input/output dimensions

d_x, d_y: input, output dimensions of the target function

Reference	Distance	Domain	Activation	Lower Bound
Park et al., 2021	Uniform	$[0,1]^{d_x}$	ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$
Cai, 2023	Uniform	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$
Ours	Uniform	$[0,1]^{d_x}$	Conti. monotone	$w_{\text{min}} \geq d_y + \mathbf{1}_{d_x < d_y \leq 2d_x}$

$\mathbf{1}_A = 1$ if A is true, and $\mathbf{1}_A = 0$ otherwise

Main Results

Summary

1. Exact minimum width in L^p distance

Reference	Distance	Domain	Activation	Exact Minimum Width
Park et al., 2021	L^p	$\mathbb{R}^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x + 1, d_y\}$
Cai, 2023	L^p	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU-Like	$w_{\text{min}} = \max\{d_x, d_y, 2\}$

2. Lower bound on minimum width in uniform distance

Reference	Distance	Domain	Activation	Lower Bound
Park et al., 2021	Uniform	$[0,1]^{d_x}$	ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$
Cai, 2023	Uniform	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$
Ours	Uniform	$[0,1]^{d_x}$	Conti. monotone	$w_{\text{min}} \geq d_y + \mathbf{1}_{d_x < d_y \leq 2d_x}$

Main Results

Summary

1. Exact minimum width in L^p distance

Reference	Distance	Domain	Activation	Exact Minimum Width
Park et al., 2021	L^p	$\mathbb{R}^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x + 1, d_y\}$
Cai, 2023	L^p	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU	$w_{\text{min}} = \max\{d_x, d_y, 2\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU-Like	$w_{\text{min}} = \max\{d_x, d_y, 2\}$

2. Lower bound on minimum width in uniform distance

Reference	Distance	Domain	Activation	Lower Bound
Park et al., 2021	Uniform	$[0,1]^{d_x}$	ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$
Cai, 2023	Uniform	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\text{min}} = 3 > \max\{d_x + 1, d_y\}$
Ours	Uniform	$[0,1]^{d_x}$	Conti. monotone	$w_{\text{min}} \geq d_y + \mathbf{1}_{d_x < d_y \leq 2d_x}$

Proof Sketch: Upper Bounds

- Idea: A piecewise constant function can approximate continuous function within an arbitrary error in L^p distance
- Consider a partition of the domain and then map input vectors in most part of each partition to approximate target vector

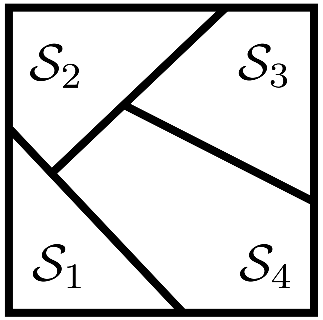


Illustration : $d_x = 2$ and the number of partitions is 4

Proof Sketch: Upper Bounds

- Idea: A piecewise constant function can approximate continuous function within an arbitrary error in L^p distance
- Consider a partition of the domain and then map input vectors in most part of each partition to approximate target vector

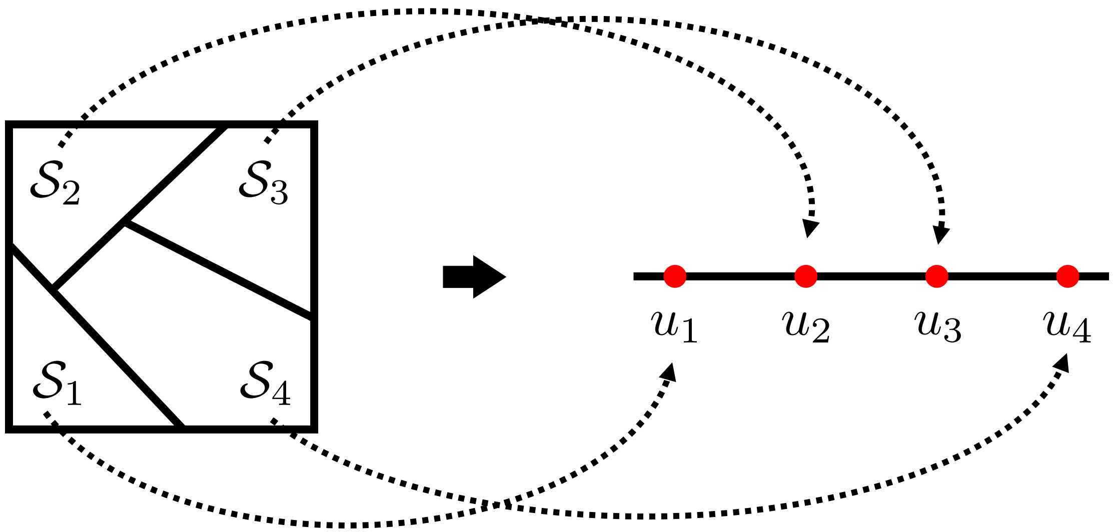
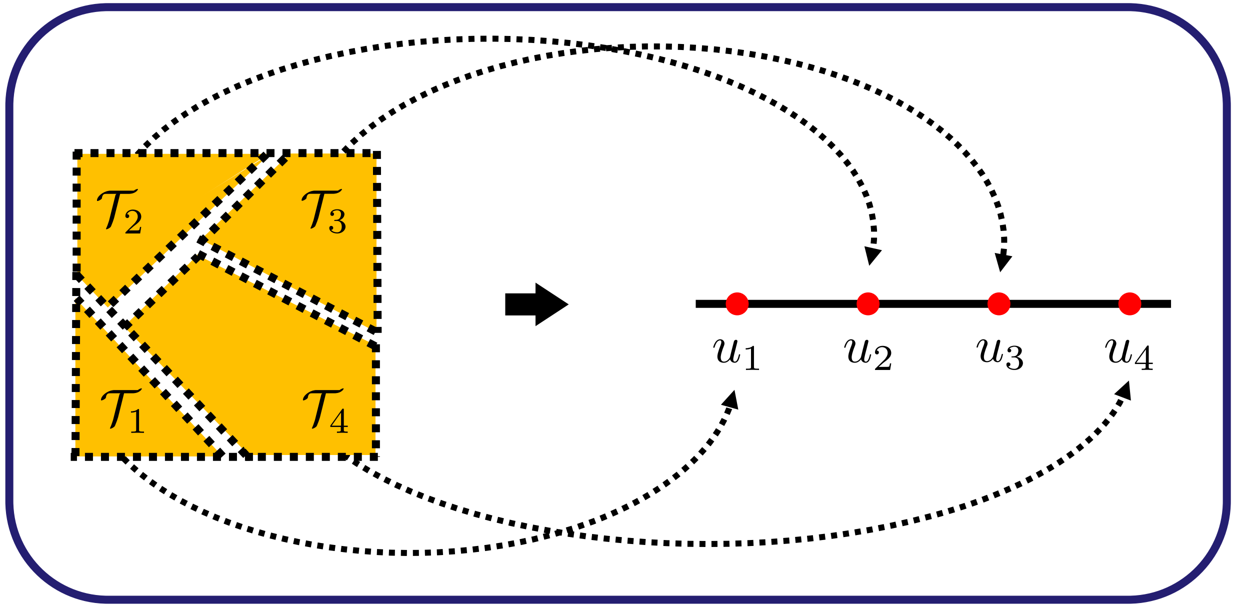


Illustration : $d_x = 2$ and the number of partitions is 4

Proof Sketch: Upper Bounds

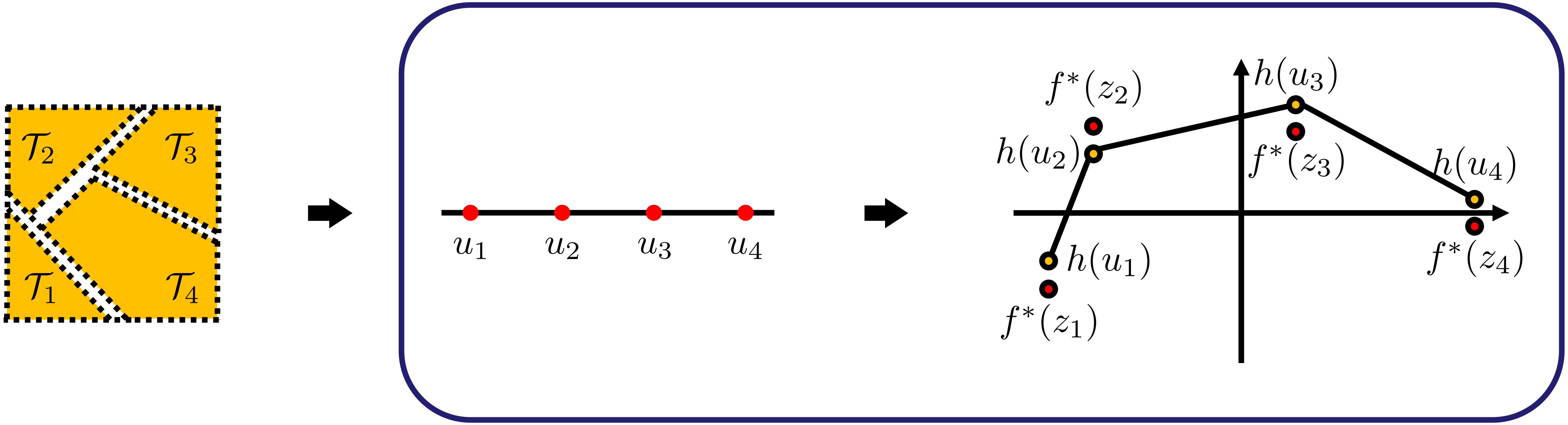
- Idea: A piecewise constant function can approximate continuous function within an arbitrary error in L^p distance
- Consider a partition of the domain and then map input vectors in most part of each partition to approximate target vector



width max$\{d_x, 2\}$ ReLU network

Proof Sketch: Upper Bounds

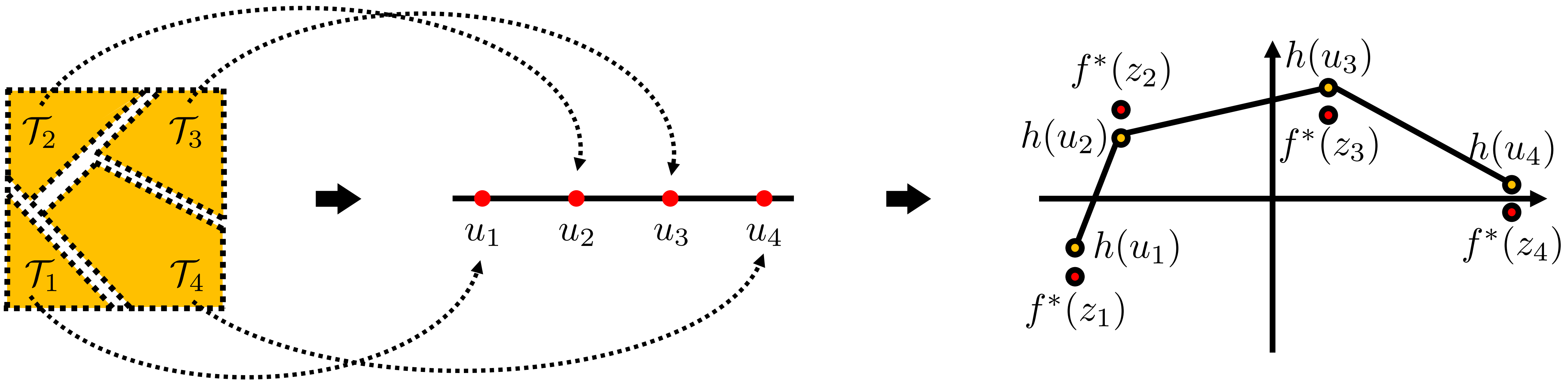
- Idea: A piecewise constant function can approximate continuous function within an arbitrary error in L^p distance
- Consider a partition of the domain and then map input vectors in most part of each partition to approximate target vector



width max$\{d_y, 2\}$ ReLU network

Proof Sketch: Upper Bounds

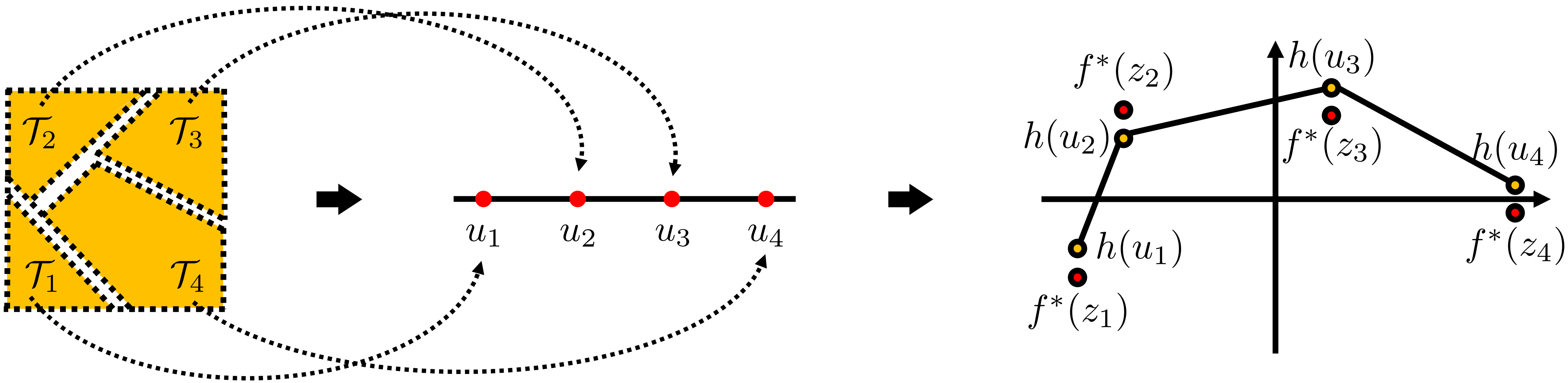
- Idea: A piecewise constant function can approximate continuous function within an arbitrary error in L^p distance
- Consider a partition of the domain and then map input vectors in most part of each partition to approximate target vector



ReLU networks of width max$\{d_x, d_y, 2\}$ are universal approximators in L^p distance

Proof Sketch: Upper Bounds

- Idea: A piecewise constant function can approximate continuous function within an arbitrary error in L^p distance
- Consider a partition of the domain and then map input vectors in most part of each partition to approximate target vector



Also, width max$\{d_x, d_y, 2\}$ networks using ReLU-Like activation functoins are universal approximatiors in L^p distance

Main Results

Summary

1. Exact minimum width in L^p distance

Reference	Distance	Domain	Activation	Exact Minimum Width
Park et al., 2021	L^p	$\mathbb{R}^{d_x}$	ReLU	$w_{\min} = \max\{d_x + 1, d_y\}$
Cai, 2023	L^p	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\min} = \max\{d_x, d_y, 2\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU	$w_{\min} = \max\{d_x, d_y, 2\}$
Ours	L^p	$[0,1]^{d_x}$	ReLU-Like	$w_{\min} = \max\{d_x, d_y, 2\}$

2. Lower bound on minimum width in uniform distance

Reference	Distance	Domain	Activation	Lower Bound
Park et al., 2021	Uniform	$[0,1]^{d_x}$	ReLU	$w_{\min} = 3 > \max\{d_x + 1, d_y\}$
Cai, 2023	Uniform	$[0,1]^{d_x}$	Leaky-ReLU	$w_{\min} = 3 > \max\{d_x + 1, d_y\}$
Ours	Uniform	$[0,1]^{d_x}$	Conti. monotone	$w_{\min} \geq d_y + \mathbf{1}_{d_x < d_y \leq 2d_x}$

Proof Sketch: Lower Bound

Our choice of target function f^*

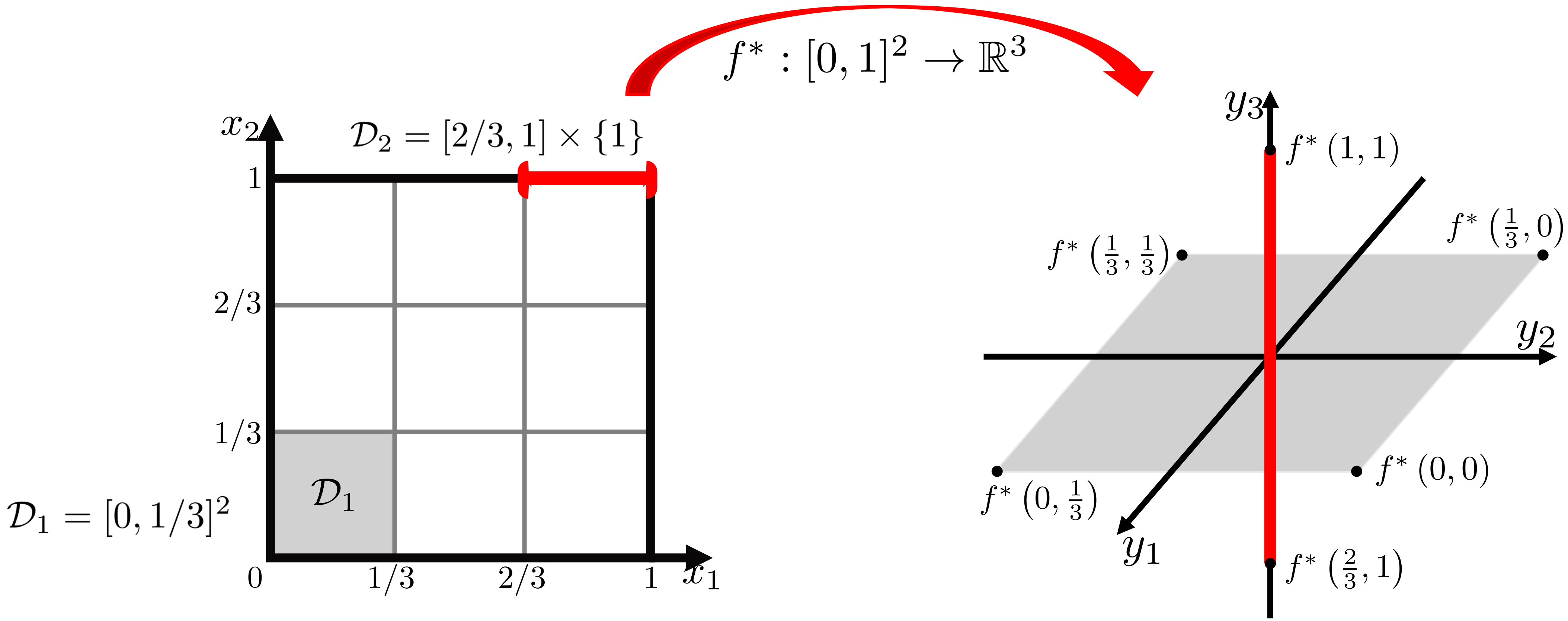


Illustration : $d_x = 2$ and $d_y = 3$

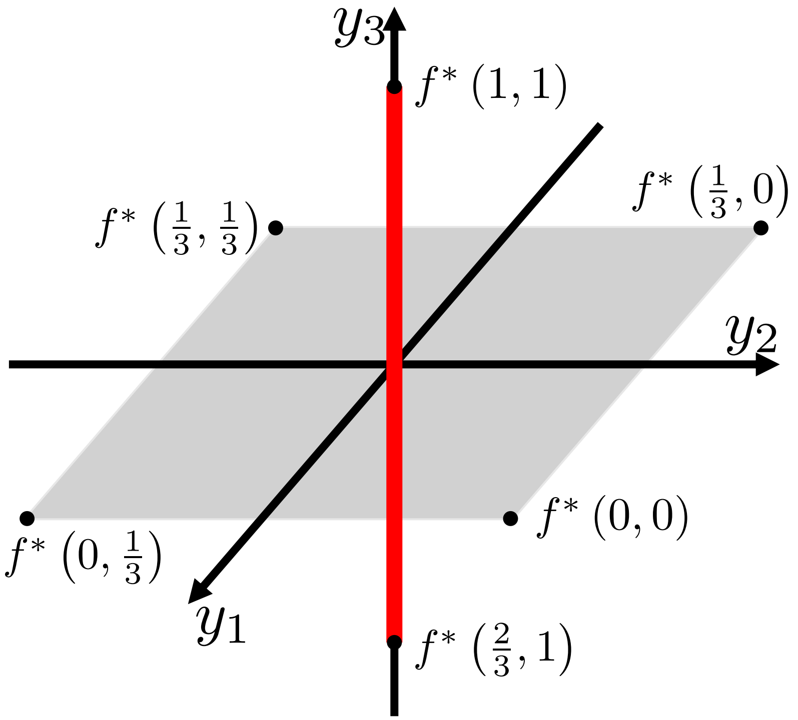
Proof Sketch: Lower Bound

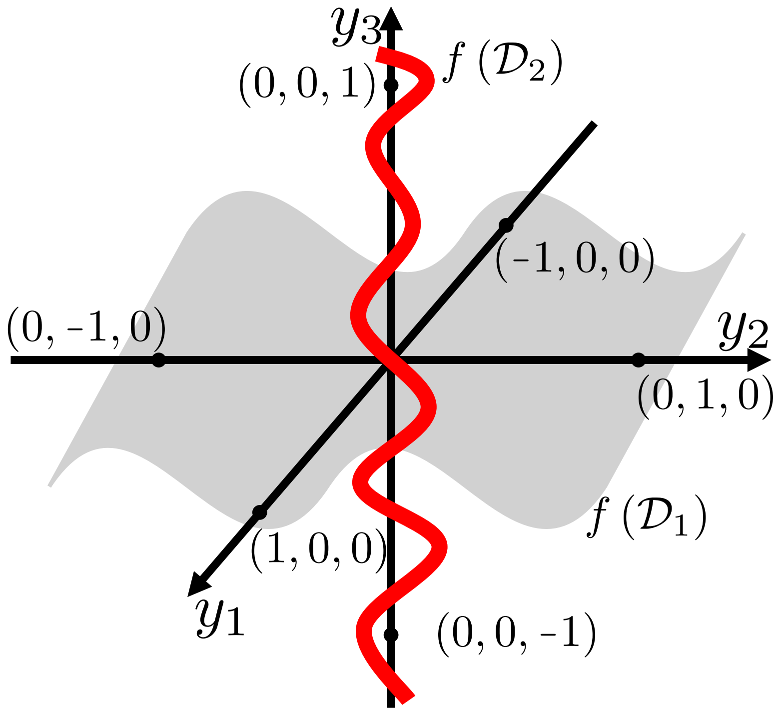
- Fact: Networks using continuous monotone activation function can be approximated by networks using continuous injection activation function in uniform distance

Proof Sketch: Lower Bound

- Fact: Networks using continuous monotone activation function can be approximated by networks using continuous injection activation function in uniform distance

- However, width d_y continuous injection networks, which can uniformly approximate f^* in small uniform error, must have an intersection.





Summary

- We first prove that the minimum width of networks on compact domain using RELU or RELU-LIKE activation function is exactly max$\{d_x, d_y, 2\}$

- We improve the previous lower bound on the minimum width for universal approximation in uniform distance: general activation functions & input/output dimensions

For more details and additional results,
read our paper and come to our poster session!

Wed 8 May 10:45am @ Halle B