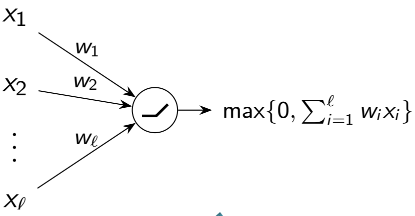
Lower Bounds on the Depth of Integral ReLU Neural Networks via Lattice Polytopes

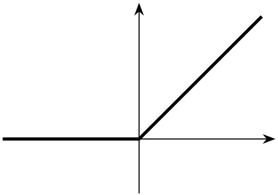
Christian Haase **Christoph Hertrich** Georg Loho

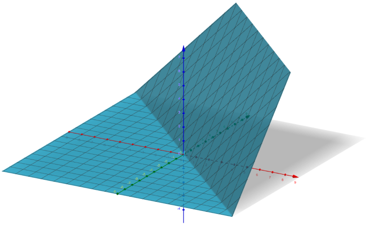
UNIVERSITY OF TWENTE.

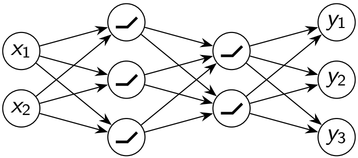
ICLR 2023

ReLU Neural Networks

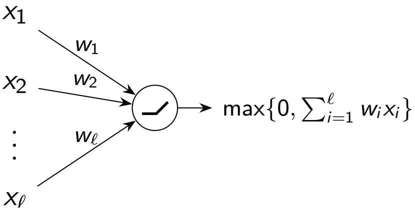


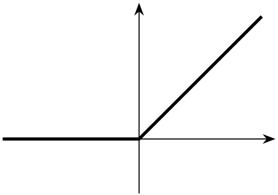


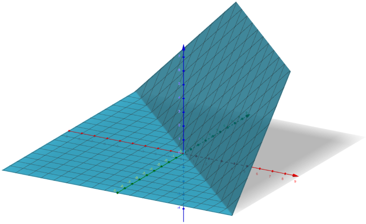


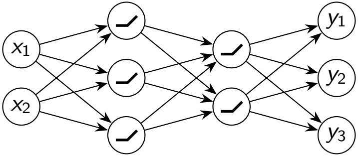


What is the class of functions computable by
ReLU Neural Networks
with a certain number of layers?



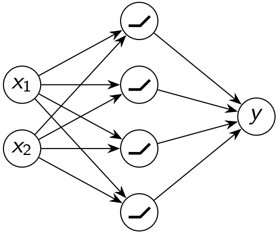






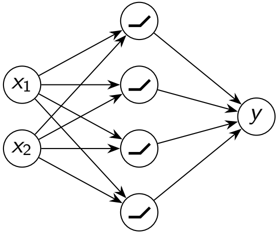
Universal approximation theorems:

One hidden layer enough to **approximate** any continuous function.



Universal approximation theorems:

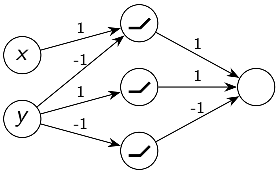
One hidden layer enough to **approximate** any continuous function.



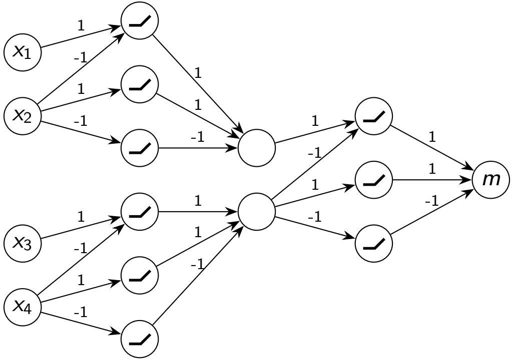
What about **exact** representability?

Example: Computing the Maximum of Two Numbers

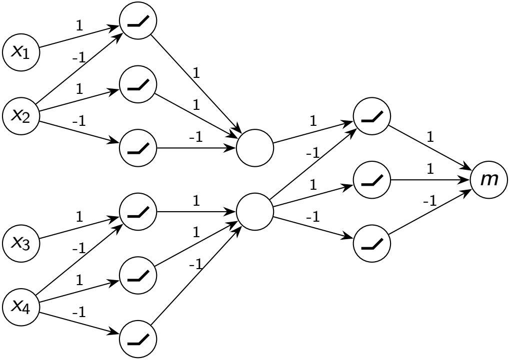
$$\max\{x, y\} = \max\{x - y, 0\} + y$$



Example: Computing the Maximum of Four Numbers



Example: Computing the Maximum of Four Numbers



- Inductively: Max of n numbers with $\lceil \log_2(n) \rceil$ hidden layers.

More Generally ...

Theorem (Arora, Basu, Mianjy, Mukherjee (2018))

Every continuous, piecewise linear function $f : \mathbb{R}^n \to \mathbb{R}$ *can be represented by a ReLU NN with* $\lceil \log_2(n+1) \rceil$ *hidden layers.*

More Generally ...

Theorem (Arora, Basu, Mianjy, Mukherjee (2018))

Every continuous, piecewise linear function $f : \mathbb{R}^n \to \mathbb{R}$ can be represented by a ReLU NN with $\lceil \log_2(n+1) \rceil$ hidden layers.

- Is logarithmic depth best possible?

Conjecture (Hertrich, Basu, Di Summa, Skutella (2021))

Yes, there are functions which need $\lceil \log_2(n+1) \rceil$ *hidden layers!*

Conjecture (Hertrich, Basu, Di Summa, Skutella (2021))

Yes, there are functions which need $\lceil \log_2(n+1) \rceil$ *hidden layers!*

This is equivalent to:

Conjecture (Hertrich, Basu, Di Summa, Skutella (2021))

$\max\{0, x_1, \ldots, x_{2^k}\}$ *cannot be represented with* k *hidden layers.*

Results

Conjecture (Hertrich, Basu, Di Summa, Skutella (2021))

$\max\{0, x_1, \ldots, x_{2^k}\}$ *cannot be represented with k hidden layers.*

Results

Conjecture (Hertrich, Basu, Di Summa, Skutella (2021))

$\max\{0, x_1, \ldots, x_{2^k}\}$ *cannot be represented with k hidden layers.*

- true for $k = 1$ (Mukherjee, Basu (2017)):
  "$\max\{0, x_1, x_2\}$ is not representable with 1 hidden layer"

Results

Conjecture (Hertrich, Basu, Di Summa, Skutella (2021))

$\max\{0, x_1, \ldots, x_{2^k}\}$ *cannot be represented with k hidden layers.*

- true for $k = 1$ (Mukherjee, Basu (2017)):
  "$\max\{0, x_1, x_2\}$ is not representable with 1 hidden layer"
- open for $k \geq 2$:
  "Is $\max\{0, x_1, x_2, x_3, x_4\}$ representable with 2 hidden layers?"

Results

Conjecture (Hertrich, Basu, Di Summa, Skutella (2021))

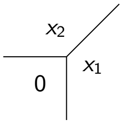
$\max\{0, x_1, \ldots, x_{2^k}\}$ *cannot be represented with k hidden layers.*

- true for $k = 1$ (Mukherjee, Basu (2017)):
  "$\max\{0, x_1, x_2\}$ is not representable with 1 hidden layer"
- open for $k \geq 2$:
  "Is $\max\{0, x_1, x_2, x_3, x_4\}$ representable with 2 hidden layers?"
- **We show:**
  Conjecture holds for all k if network has only integer weights!

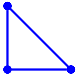
Proof Techniques

- Use tropical geometry to represent NNs as lattice polytopes. (Compare Zhang, Naitzat, Lim (2018))

Example for $\max\{0, x_1, x_2\}$:



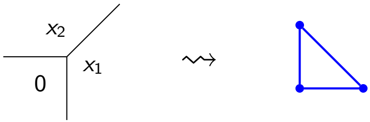
$\rightsquigarrow$



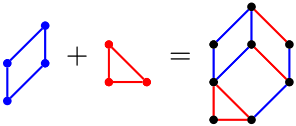
Proof Techniques

- Use tropical geometry to represent NNs as lattice polytopes. (Compare Zhang, Naitzat, Lim (2018))

Example for $\max\{0, x_1, x_2\}$:



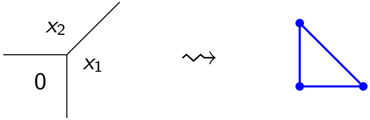
- Subdivide polytopes “layer by layer” into “easier” polytopes.



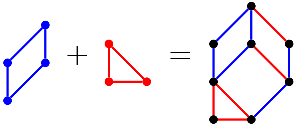
Proof Techniques

- Use tropical geometry to represent NNs as lattice polytopes. (Compare Zhang, Naitzat, Lim (2018))

Example for $\max\{0, x_1, x_2\}$:



- Subdivide polytopes "layer by layer" into "easier" polytopes.



- Separate via parity of the normalized volume.

Outlook

To prove general conjecture ...

- Polytopes and subdivisions seem promising.
- Replace volume argument by different separation.

Outlook

To prove general conjecture ...

- Polytopes and subdivisions seem promising.
- Replace volume argument by different separation.

Thanks for watching!