On the expressiveness of rational ReLU neural networks with bounded depth

Gennadiy Averkov

BTU Cottbus, Germany

ICLR 2025, April 24-28, Singapore

Joint paper with Christopher Hojny and Maximilian Merkert

Meta question: high depth $\Rightarrow$ high expressive power?

depth $:=$ number of hidden layers

ReLU Networks: activation function $\mathrm{ReLU}(t) = \max\{0, t\}$

Exact Representations: motivated by integer quantization

Piecewise Linearity

Every ReLU network defines a continuous piecewise linear function.

Fundamental Theorem (Arora et al. 2018)

For ReLU networks, depth $\lceil \log_2(n+1) \rceil$ is enough to represent any continuous piecewise linear function in n variables.

Main Conjecture

For each $k \in \{1, \ldots, \lceil \log_2(n+1) \rceil\}$,
there exist functions in n variables,
representable with depth k but not with depth
$k-1$.

Equivalent Conjecture (Hertrich et al. 2021)

For $n = 2^t$, the function $\max\{0, x_1, \ldots, x_n\}$ admits a representation with depth $t + 1$ but not with depth t.

Partial Confirmation (Haase et al. 2023)

Main Conjecture and Equivalent Conjecture
are true for networks with integer weights.

New Result (ICLR 2025)

Integer weights $\rightarrow$ rational weights

Fractions as weights

N-ary fraction: $\frac{z}{N^e}$, with $z \in \mathbb{Z}$, $e \in \mathbb{Z}_{\geq 0}$.

Example of	
decimal fraction	3.14159265
binary fraction	11.001001_2

Necessary depth for weights being N-ary fractions

To represent $\max\{0, x_1, \ldots, x_n\}$ with N-ary fractional weights:
depth $\geq \lceil \log_p(n+1) \rceil$ is necessary,
when p is a prime that does not divide N.

$\Rightarrow$ depth $\geq \Omega(\frac{\ln n}{\ln \ln N})$
$\Rightarrow$ depth $\geq \lceil \log_3(n+1) \rceil$ when weights are decimal fractions.