Neural network expressivity and simplex subdivisions

Egor Bakaev*

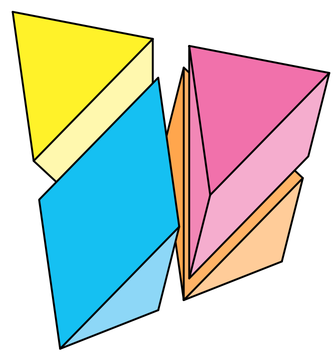

Figure 1: Subdividing the three-simplex into four identical rhombic pyramids.

This work studies the expressivity of ReLU neural networks with a focus on their depth. Functions represented by ReLU networks are *continuous and piecewise linear* (CPWL), and every such function on $\mathbb{R}^n$ can be computed by a ReLU network [1]. This naturally leads to the complexity-theoretic question: what is the minimal depth needed to represent CPWL functions? An important CPWL function for understanding neural network depth is the function

$$\mathsf{MAX}_n(x) = \mathsf{MAX}_n(x_1, \ldots, x_n) = \mathsf{max}\{x_1, \ldots, x_n\}.$$

Wang and Sun [3] showed that every function in CPWL_n can be written as a linear combination of MAX_{n+1} functions applied to some affine functions. As observed by Arora, Basu, Mianjy, and Mukherjee [1], this fact implies that

$$\mathsf{CPWL}_n \subseteq \mathsf{ReLU}_{n,\lceil \log_2(n+1) \rceil}\,.$$

In words, every CPWL function on $\mathbb{R}^n$ can be represented with $\lceil \log_2(n+1) \rceil$ hidden layers by implementing it in a binary tree manner. Hertrich, Basu, Di Summa, and Skutella [2] conjectured that the $\lceil \log_2(n) \rceil$ upper bound on the

*joint work with Florestan Brunck, Christoph Hertrich, Jack Stade, Amir Yehudayoff required number of hidden layers for MAX_n is sharp. This conjecture led to a sequence of works that proved it in a variety of special cases.

Our main result is that the conjecture is false. More precisely, we prove that the maximum of n numbers can be computed with $\lceil \log_3(n-2) \rceil + 1$ hidden layers.

Theorem 1. *For $n \geq 1$, we have $\mathsf{MAX}_{3^n+2} \in \mathsf{ReLU}_{n+1}$.*

By the discussion above, this implies that every CPWL function defined on $\mathbb{R}^n$ can be represented with $\lceil \log_3(n-1) \rceil + 1$ hidden layers.

Theorem 2. *For $n \geq 3$, we have $\mathsf{CPWL}_n = \mathsf{ReLU}_{n, \lceil \log_3(n-1) \rceil + 1}$.*

It is noteworthy that Theorem 1 already improves the depth complexity of MAX_5.

Proposition 3. *The minimum number of hidden layers needed for computing MAX_5 is exactly two.*

The proof of our result is short and elementary. However, it hides a beautiful geometric intuition that guided our search for the construction and that may be useful to further improve the current upper or lower bounds.

The geometric intuition is based on relations between neural networks and geometry. We use such geometrical notions as valuation and full additivity of the identity map.

We explain the geometric intuition behind our construction of the two-hidden-layer ReLU neural network for MAX_5. Geometrically, we subdivide Δ^4 to four pieces, each in $\mathcal{P}_2$. This subdivision leads to the two-hidden-layer ReLU network for MAX_5. This is a nice 3D puzzle, and an illustration of the solution is given in Figure 1.

References

[1] Raman Arora, Amitabh Basu, Poorya Mianjy, and Anirbit Mukherjee, *Understanding deep neural networks with rectified linear units*, International Conference on Learning Representations, 2018.

[2] Christoph Hertrich, Amitabh Basu, Marco Di Summa, and Martin Skutella, *Towards lower bounds on the depth of ReLU neural networks*, Advances in Neural Information Processing Systems **34** (2021), 3336–3348.

[3] Shuning Wang and Xusheng Sun, *Generalization of hinging hyperplanes*, IEEE Transactions on Information Theory **51** (2005), no. 12, 4425–4431.