DATA GEOMETRY AND TOPOLOGY-DEPENDENT BOUNDS ON DEEP RELU NETWORK WIDTHS

Sangmin Lee [1] , Abbas Mammadov [1] and Jong Chul Ye [2]

1) *Department of Mathematical Sciences, KAIST, Daejeon, KOREA*

2) *Kim Jaechul Graduate School of AI, KAIST, Daejeon, KOREA*

Corresponding Author : Jong Chul Ye, jong.ye@kaist.ac.kr

ABSTRACT

The geometrical perspective of deep ReLU networks has been extensively explored to comprehend the training capabilities of neural networks. In this paper, we delve into the intricate relationship between the geometric and topological attributes of datasets and the architecture of ReLU networks. Specifically, we establish upper and lower bounds on network widths in terms of the number of faces in convex polytopes. Furthermore, we demonstrate that network widths are not exclusively determined by topological features alone. However, by imposing certain geometric assumptions, we also provide width bounds based on the Betti numbers of the data manifold.

INTRODUCTION

Since Cybenko firstly proved the universal approximation property (UAP) of two-layer neural networks [1], research on UAP of neural networks have been widely studied. Despite numerous prior studies, the relationship between the characteristics of the training dataset and network architectures has not been thoroughly investigated. The goal of this paper is to study the following fundamental approximation problem: *For a given topological space $\mathcal{X}$ representing the dataset and $\varepsilon > 0$, what is a bound on the widths of a neural network $\mathcal{N}$ such that $\mathcal{N}(\boldsymbol{x}) = 1$ for $\boldsymbol{x} \in \mathcal{X}$, and it vanishes outside on the ε neighborhood of $\mathcal{X}$?* We address this question based on the geometric and topological characteristics of the data manifold $\mathcal{X}$.

Notations.

The ReLU activation function is denoted by $\sigma(x) := \text{ReLU}(x) = \max\{x, 0\}$, and the sigmoid function is denoted by $\texttt{SIG}(x) := \frac{1}{1+e^{-x}}$. We denote the architecture of neural networks by arrow between hidden nodes with the activation functions, for example, $d \overset{\sigma}{\to} d_1 \overset{\sigma}{\to} d_2 \overset{\sigma}{\to} d_3 \overset{\texttt{SIG}}{\to} 1$. For a given topological space $\mathcal{X}$, if a neural network with architecture $\mathcal{A}$ has UAP on $\mathcal{X}$, then we say that $\mathcal{A}$ is a *feasible architecture on $\mathcal{X}$*.

RESULTS

When the given dataset $\mathcal{X}$ forms a convex polytope in $\mathbb{R}^d$, we provide upper and lower bounds for the feasible architecture on $\mathcal{X}$ in the following theorem.

Theorem 1. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a convex polytope enclosed by l hyperplanes. Then, $d \xrightarrow{\sigma} l \xrightarrow{\sigma} 1$ is a feasible architecture on $\mathcal{X}$ with minimal depth. Conversely, if $d \xrightarrow{\sigma} d_1 \xrightarrow{\sigma} d_2 \xrightarrow{\sigma} \cdots \xrightarrow{\sigma} d_k \xrightarrow{\sigma} 1$ is a feasible architecture on $\mathcal{X}$, then*

$$d_1 \cdot \prod_{j=2}^{k}(2d_j+1) \geq \begin{cases} \left\lceil \frac{l}{2} \right\rceil + (d-2), & \textit{if} \quad l \geq 2d+1, \\ 2d-1, & \textit{if} \quad l = 2d-1, 2d, \\ d+1, & \textit{if} \quad l < 2d-1. \end{cases} \tag{1}$$

This lower bound is optimal when $k = 1$ and $d = 2$.

This theorem elucidates how the geometric complexity of the dataset affects on the network architectures. Before deriving a similar result for topological complexity, we first mention that topological characteristics cannot be the only factor determining the feasible architecture on $\mathcal{X}$.

Proposition 2. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a topological space and $\mathcal{A}$ be a feasible architecture on $\mathcal{X}$. Then, there exists a topological space $\mathcal{X}'$ which is homeomorphic to $\mathcal{X}$, but $\mathcal{A}$ is not a feasible architecture on $\mathcal{X}'$.*

Nevertheless, we can derive topology-dependent bounds on widths by imposing certain geometric shapes on $\mathcal{X}$. The following theorem provides upper and lower bounds of network widths based on the Betti numbers of $\mathcal{X}$, which represent the topological complexity of $\mathcal{X}$.

Theorem 3. *Let $\mathcal{X}$ be a topological space obtained by removing some disjoint prism-shaped convex polytopes from a convex polytope. Let l be the maximum number of faces of these polytopes. Let β_k be the k-th Betti number of $\mathcal{X}$. Then,*

$$d \xrightarrow{\sigma} \left(l + 2(\beta_0 - 1) + \sum_{k=1}^{d} \left(l - 2(d-k-1)\right) \beta_k \right) \xrightarrow{\sigma} \left(\sum_{k=0}^{d} \beta_k \right) \xrightarrow{\sigma} 2 \xrightarrow{\sigma} 1 \tag{2}$$

is a feasible architecture on $\mathcal{X}$. Conversely, for any such $\mathcal{X}$, suppose $d \xrightarrow{\sigma} d_1 \xrightarrow{\sigma} d_2 \xrightarrow{\sigma} \cdots \xrightarrow{\sigma} d_k \xrightarrow{f} 1$ is a feasible architecture on $\mathcal{X}$ where the last activation function f is either σ or `SIG`. *Then, the network widths should satisfy*

$$\sum_{i=1}^{k} \prod_{j=i}^{k} d_j \geq 2 \sum_{k=0}^{d} \beta_k - 2. \tag{3}$$

CONCLUSION

In this work, we investigated the intricate relationship between geometric and topological characteristics of datasets and the architectures of neural networks. Specifically, when the given manifold $\mathcal{X}$ forms a closed convex polytope, we established upper and lower bounds for the widths of neural networks. Regarding the topological features of manifolds, we proved that two homeomorphic topological spaces may require different network architectures. However, by imposing certain geometrical restrictions on the datasets, we derived upper and lower bounds for widths in terms of their Betti numbers. Overall, our results theoretically illustrated how the complexity of a dataset influences neural network architectures.

REFERENCES

1. Cybenko, G., “Approximation by superpositions of a sigmoidal function, *Mathematics of control*, signals and systems, 2(4):303–314, 1989.