Hindawi
Journal of Function Spaces
Volume 2021, Article ID 6637220, 3 pages
https://doi.org/10.1155/2021/6637220

*Research Article*

# A Global Universality of Two-Layer Neural Networks with ReLU Activations

**Naoya Hatano,[1,2,3] Masahiro Ikeda,[1,2] Isao Ishikawa,[2,4] and Yoshihiro Sawano[1,2,3]**

[1] *Department of Mathematics, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*
[2] *Center for Advanced Intelligence Project, Riken, Japan*
[3] *Department of Mathematics, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan*
[4] *Center for Data Science, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan*

Correspondence should be addressed to Naoya Hatano; n.hatano.chuo@gmail.com

Received 23 November 2020; Accepted 1 February 2021; Published 17 September 2021

Academic Editor: Alberto Fiorenza



In the present study, we investigate a universality of neural networks, which concerns a density of the set of two-layer neural networks in function spaces. There are many works that handle the convergence over compact sets. In the present paper, we consider a global convergence by introducing a norm suitably, so that our results will be uniform over any compact set.

## 1. Introduction

Neural network is a function that models a neuron system of a biological brain and is defined as alternate compositions of an affine map and a nonlinear map. The nonlinear map in a neural network is called the activation function. The neural networks have been playing a central role in the field of machine learning with a vast number of applications in the real world in the last decade. We refer to [1] and [2] for example.

We focus on a two-layer feed-forward neural network with ReLU (rectified linear unit) activation, which is a function $\mathrm{f} : \mathbb{R} \longrightarrow \mathbb{R}$ of the form of $f(x) = \sum_{\mathrm{i}=1}^{\mathrm{r}} \mathrm{c_i}\mathrm{ReLU}(a_i x + b_i)$ for some $a_1, b_1, c_1, \cdots, a_r, b_r, c_r \in \mathbb{R}$. Here, the function ReLU is called the rectified linear unit defined by

$$\mathrm{ReLU}(x) \coloneqq \max\,(x, 0)(x \in \mathbb{R}). \tag{1}$$

The ReLU is one of the most popular activation functions for feed-forward neural networks in practical machine learning tasks for real-world problems.

We consider the space of two-layer feedforward neural networks defined by the following linear space

$$\begin{aligned}\mathscr{X} &\coloneqq \mathrm{Span}(\{\mathrm{ReLU}(a\cdot +b) \colon a \neq 0, b \in \mathbb{R}\}) \\ &= \left\{\sum_{j=1}^{N} \lambda_j \mathrm{ReLU}(a_j \cdot + b_j) \colon N \in \mathbb{N}, \left\{(\lambda_j, a_j, b_j)\right\}_{j=1}^{N} \subset \mathbb{C}\times\mathbb{R}^n\times\mathbb{R}\right\}.\end{aligned} \tag{2}$$

Then, it is natural to ask ourselves whether $\mathscr{X}$ spans a dense subspace of a function space (topological linear space), which is called universality of $\mathscr{X}$. Historically, the density property of $\mathscr{X}$ in the space $\mathrm{C}(\mathbb{R})$ of continuous functions on $\mathbb{R}$ is investigated by several authors ([3–5]) since it is important to guarantee the existence of a feed-forward neural network $f \in \mathscr{X}$ that well approximates an unknown continuous function. Here, the topology of $\mathrm{C}(\mathbb{R})$ is generated by the seminorms $h \longmapsto \sup_{x\in K} |h(x)|$, where $K$ ranges over all compact sets in $\mathbb{R}$. Thus, the approximation property of two-layer feed-forward neural networks makes sense only on a local domain.

In this study, we prove an approximation property of $\mathscr{X}$ in a global sense. More precisely, we prove the space $\mathscr{X}$ is dense in the Banach subspace of $C(\mathbb{R})$ defined as

$$\mathscr{Y} \coloneqq \left\{ f \in C(\mathbb{R}) \lim_{x \longrightarrow \pm\infty} \frac{f(x)}{1+|x|} \text{ exists and it is finite} \right\}, \quad (3)$$

equipped with the norm

$$\|f\|_{\mathscr{Y}} \coloneqq \sup_{x \in \mathbb{R}} \frac{|f(x)|}{1+|x|}. \quad (4)$$

Note that any element in $\mathscr{Y}$, divided by $1+|\cdot|$, is a continuous function over $\bar{\mathbb{R}} \coloneqq \mathbb{R} \cup \{\pm\infty\}$. Our main result in this paper is as follows:

**Theorem 1.** *The linear subspace $\mathscr{X}$ is dense in $\mathscr{Y}$.*

Our main results claim that any function $f \in \mathscr{X}$ is close to a linear function both at $\infty$ and at $-\infty$. Near the origin, $\mathscr{X}$ approximates any continuous functions.

Before we conclude this section, we will offer some words on some existing results. See [6] for the $L^2$-approximation over the real line. Other attempts have been made to grasp the neural network by the use of the Radon transform [7] or by considering some other topologies [5, 8].

## 2. Proof of the Main Theorem

*Definition 2.* We define a linear operator $A : f \in \mathscr{Y} \longmapsto f/1+|\cdot| \in BC(\bar{\mathbb{R}})$.

**Lemma 3.** *The operator $A : \mathscr{Y} \longrightarrow BC(\bar{\mathbb{R}})$ is an isomorphism from $\mathscr{Y}$ to $BC(\bar{\mathbb{R}})$.*

A tacit understanding here is that we extend $f/1+|\cdot|$, which is initially defined over $\mathbb{R}$, continuously to $\bar{\mathbb{R}}$.

Thus, any continuous functional on $\mathscr{Y}$ is realized by a Borel measure over $\bar{\mathbb{R}}$.

Our theorem can recapture the case where the underlying domain is bounded. Indeed, if the domain $\Omega$ is contained in $[-R, R]$ for some $R > 0$, then, we have

$$\|f\|_{L^\infty(\Omega)} \le (1+R)\|f\|_{\mathscr{Y}} (f \in \mathscr{Y}), \quad (5)$$

which will give results by Cybenko [3] and Funahashi [4].

Now we start the proof of Theorem 1. As Cybenko did in [3], take any measure $\mu$ over $\bar{\mathbb{R}}$ such that $\mu$ annihilates $\mathscr{X}$. We will show that $\mu = 0$. Once this is proved, from the Riesz representation theorem, we conclude that the only linear functional that vanishes on $\mathscr{X}$ is zero. Using the Hahn-Banach theorem, we see that $\mathscr{X}$ is dense in $\mathscr{Y}$.

Remark that

$$\begin{aligned} \max(1-|x-1|, 0) = \mathrm{ReLU}(x) + \mathrm{ReLU}(x-2) \\ - 2\mathrm{ReLU}(x-1)\ (x \in \mathbb{R}). \end{aligned} \quad (6)$$

Thus, any element in $C_c(\mathbb{R})$ can be approximated by a function $\mathscr{X}$ in the $L^\infty$-norm. Since $\mu$ annihilates $C_c(\mathbb{R})$, it follows that $\mu$ is not supported on $\mathbb{R}$. Or equivalently, $\mu$ is supported on $\pm\infty$. It remains to show that $\mu(\{\pm\infty\}) = 0$. Consider

$$f(x) = \mathrm{ReLU}(x) - \mathrm{ReLU}(x-1)\ (x \in \mathbb{R}). \quad (7)$$

Remark that

$$0 = \int_{\bar{\mathbb{R}}} f(x) d\mu(x) = \mu(\{\infty\}). \quad (8)$$

Likewise, if we test the condition on $g = f(-\cdot)$, we obtain $\mu(\{-\infty\}) = 0$.

Thus, we conclude that $\mathscr{X}$ is dense in $\mathscr{Y}$.

*Remark 4.* The set $\{f, g\} \cup C_c(\mathbb{R})$ spans a dense subspace in $\mathscr{Y}$, where $f$ and $g$ are functions given in the above proof.

## Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Authors' Contributions

The four authors contributed equally to this paper. All of them read the whole manuscript and approved the content of the paper.

## Acknowledgments

This work was supported by a JST CREST Grant (number JPMJCR1913, Japan). This work was also supported by the RIKEN Junior Research Associate Program. The second author was supported by a Grant-in-Aid for Young Scientists Research (no. 19K14581), Japan Society for the Promotion of Science. The fourth author was supported by a Grant-in-Aid for Scientific Research (C) (19K03546), Japan Society for the Promotion of Science.

## References

[1] B. Hanin, "Universal function approximation by deep neural nets with bounded width and ReLU activations," *Mathematics*, vol. 7, no. 10, p. 992, 2019.

[2] A. Pinkus, "Approximation theory of the MLP model in neural networks," *Acta Numerica*, vol. 8, pp. 143–195, 1999.

[3] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.

[4] K. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, no. 3, pp. 183–192, 1989.

[5] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[6] S. Sonoda and N. Murata, “Neural network with unbounded activation functions is universal approximator,” *Applied and Computational Harmonic Analysis*, vol. 43, no. 2, pp. 233–268, 2017.

[7] M. Carroll and B. W. Dikinson, “Construction of neural nets using the Radon transform,” in *Proceedings of the IEEE 1989 International Joint Conference on Neural Networks*, vol. 1, pp. 607–611, New York, 1989.

[8] Y. T. Sun, A. Gilbert, and A. Tewari, “On the approximation properties of random ReLU features,” https://arxiv.org/pdf/1810.04374.pdf.