2 層ニューラルネットワークの大域的普遍近似定理

中央大学 理工学研究科 数学専攻
波多野修也 (Naoya HATANO)

概要

機械学習においてニューラルネットワークの普遍性定理がよく知られている．ニューラルネットワークとは, 脳神経を数理モデルとして定式化したものとして導入された．また, ニューラルネットワークによってコンパクト集合上で連続関数を近似できることを普遍近似定理と呼ばれ, よく知られている．そこで, 本研究では関数空間の言葉で置き換えることで, 定義域の取り方に依存しない大域的なものに拡張した．

1　導入

すべての連続関数は

$$\sum_{i=1}^{r} c_i \sigma(a_i x + b_i), \quad x \in \mathbb{R}, \tag{1}$$

の形の関数で近似できることがよく考えられている．ここで, $\sigma : \mathbb{R} \to \mathbb{R}$ を活性化関数と呼ばれている．活性化関数は以下のようなものがよく考えられている．

例 1.1. $\sigma(t) = \dfrac{1}{1+e^{-t}}$, $\mathrm{ReLU}(t)$, $\chi_{[0,\infty)}(t)$, etc.

$$\mathrm{ReLU}(t) := \max(0,t) = \begin{cases} t, & t \geq 0, \\ 0, & t < 0, \end{cases} \quad \chi_{[0,\infty)}(t) := \begin{cases} 1, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

さらに, $\sigma : \mathbb{R} \to \mathbb{R}$ が

$$\lim_{t \to \pm\infty} \sigma(t) = c_{\pm\infty} < \infty \text{ は収束する}, \quad c_{\infty} \neq c_{-\infty}$$

を満たす時, σ を sigmoidal という．

関数 (1) の族をニューラルネットワークと呼ばれている．

定義 1.2 (ニューラルネットワーク)**.** $\Omega \subset \mathbb{R}$ とする．活性化関数 $\sigma : \mathbb{R} \to \mathbb{R}$ に対し, Ω 上の関数の族を

$$H_\sigma(\Omega) := \left\{ \sum_{i=1}^{r} c_i \cdot \sigma(a_i \cdot + b_i) \,:\, r \in \mathbb{N},\, a_i \neq 0, b_i, c_i \in \mathbb{R} \right\}.$$

とする．

局所的には近似が可能であることが従来の普遍近似定理である．

定理 1.3 (Cybenko [1] and Funahashi [2], 1989)**.** $\sigma : \mathbb{R} \to \mathbb{R}$ を連続かつ sigmoidal または ReLU とする．任意の $f \in \mathrm{C}([0,1])$ と $\varepsilon > 0$ に対して，ある $g \in H_\sigma([0,1])$ が存在して

$$\sup_{x \in [0,1]} |f(x) - g(x)| < \varepsilon$$

を満たす．すなわち，$H_\sigma([0,1])$ は $(\mathrm{C}([0,1]), \|\cdot\|_{\mathrm{C}([0,1])})$ に稠密である．ただし，

$$\|f\|_{\mathrm{C}([0,1])} := \sup_{x \in [0,1]} |f(x)|$$

である．

本研究では, この結果を大域的な場合に拡張した.

2 主結果

それぞれの活性化関数について, 主結果を述べる.

関数空間

$$\mathrm{BUC}(\mathbb{R}) := \{ f \,:\, f \text{ は有界かつ一様連続な関数 } \}$$

にはノルム

$$\|f\|_{\mathrm{BUC}} := \sup_{x \in \mathbb{R}} |f(x)|$$

が備わっている．このとき, σ が sigmoidal の場合は以下のように述べられる．

定理 2.1 (N.-Ikeda-Ishikawa-Sawano)**.** $\sigma : \mathbb{R} \to \mathbb{R}$ を連続かつ sigmoidal とする． このとき, $(\mathrm{BUC}(\mathbb{R}), \|\cdot\|_{\mathrm{BUC}})$ で $H_\sigma(\mathbb{R})$ は閉部分空間

$$\mathfrak{X}(\mathbb{R}) := \left\{ f \in \mathrm{BUC}(\mathbb{R}) \,:\, \lim_{x \to \pm\infty} f(x) = f(\pm\infty) < \infty \text{ は収束する} \right\}$$

に稠密である．

さらに, 関数空間

$$\mathcal{Y}(\mathbb{R}) := \left\{ f \in \mathrm{C}(\mathbb{R}) \,:\, \lim_{x \to \pm\infty} \frac{f(x)}{1 + |x|} < \infty \text{ は収束} \right\}$$

にはノルム

$$\|f\|_{\mathcal{Y}} := \sup_{x \in \mathbb{R}} \frac{|f(x)|}{1 + |x|}$$

が備わっている．このとき, σ が ReLU の場合は以下のように述べられる．

定理 2.2 (N.-Ikeda-Ishikawa-Sawano, [3])**.** $H_{\mathrm{ReLU}}(\mathbb{R})$ は $(\mathcal{Y}(\mathbb{R}), \|\cdot\|_{\mathcal{Y}})$ に稠密である．

最後に, 関数空間

$$\mathcal{C}(\mathbb{R}) := \{ f \,:\, f \text{ は右連続かつ左極限をもつ } \}$$

には, ノルム

$$\|f\|_{\mathcal{C}} := \sup_{x \in \mathbb{R}} |f(x)|$$

が備わっている．このとき, σ が Heaviside 関数 $\chi_{[0,\infty)}$ の場合は以下のように述べられる．

定理 2.3 (N.-Ikeda-Ishikawa-Sawano). $(\mathcal{C}(\mathbb{R}), \|\cdot\|_{\mathcal{C}})$ で $H_{\chi_{[0,\infty)}}(\mathbb{R})$ は閉部分空間

$$\mathcal{D}(\mathbb{R}) := \left\{ f \in \mathcal{C}(\mathbb{R}) : \lim_{x \to \pm\infty} f(x) = f(\pm\infty) < \infty \text{ は収束する} \right\}$$

に稠密である.

3 補足

本講演では, 定理 2.2 の証明のみを紹介する. 証明のアイデアは以下の補題による関数空間 $\mathcal{Y}(\mathbb{R})$ の言い換え後, 双対性を用いる.

補題 3.1. 写像 $\mathcal{Y}(\mathbb{R}) \ni f \mapsto \dfrac{f}{1+|\cdot|} \in \mathrm{BC}(\overline{\mathbb{R}})$ により, 関数空間 $\mathcal{Y}(\mathbb{R})$ と $\mathrm{BC}(\overline{\mathbb{R}})$ は同相となる. ただし, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ とする.

線形汎関数がニューラルネットワークの上で消滅していると仮定して, 全空間でも消滅していることを示すことで Hahn-Banach の定理から稠密性が証明できるという点は Cybenko や Funahashi によるアイデアによるものだが, それに加え, 上補題を用いた無限遠方での制御を行うことで本研究成果が得られた.

注意 3.2. それぞれの普遍近似定理定理の比較:

(1) ノルム評価

$$\|f\|_{\mathrm{C}([0,1])} \le \|f\|_{\mathrm{BUC}}, \quad \|f\|_{\mathrm{C}([0,1])} \le 2\|f\|_{\mathcal{Y}}$$

が成り立っているため, 定理 2.1 と定理 2.2 は単に定義域を制限することで *Cybenko, Funahashi* の結果, 定理 1.3 が再現できる.

(2) 定理 2.1 は活性化関数 σ の取り方によらず, 閉部分空間 $\mathfrak{X}(\mathbb{R})$ に稠密である.

(3) $\sigma_0(t) := \mathrm{ReLU}(t) - \mathrm{ReLU}(t-1)$ は sigmoidal な連続関数であり,

$$\sigma_0 \in H_{\mathrm{ReLU}}(\mathbb{R})$$

が成り立っているため, 活性化関数を ReLU とした方が sigmoidal のときよりも多くの関数を近似できる. さらに定理 2.2 の証明によると, $\mathrm{C_c}(\mathbb{R}) \cup \{\sigma_0, \sigma_0(-\cdot)\}$ は $\mathcal{Y}(\mathbb{R})$ に稠密であることもわかる. ただし, $\mathrm{C_c}(\mathbb{R})$ はコンパクト台を持つ連続関数全体とする.

(4) $\sigma_\varepsilon(t) := \mathrm{ReLU}(t+\varepsilon) - \mathrm{ReLU}(t)$, $\varepsilon > 0$ は sigmoidal な連続関数であり,

$$\lim_{\varepsilon \downarrow 0} \sigma_\varepsilon = \chi_{[0,\infty)} \in H_{\chi_{[0,\infty)}}(\mathbb{R})$$

が成り立っているため, 活性化関数を $\chi_{[0,\infty)}$ とした方が sigmoidal のときよりも多くの関数を近似できる.

参考文献

[1] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Systems (1989) 2:303–314.

[2] K. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Networks **2**, 183–192.

[3] N. Hatano, M. Ikeda, I. Ishikawa and Y. Sawano, A global universality of two-layer neural networks with ReLU activations, J. Funct. Spaces 2021.