Приложение: теоремы об универсальных моделях

Потанин М. С., Вайсер К. О., Жолобов В. А., Стрижов В. В.

1 Теоремы об универсальных моделях

Теоремы 1, 2, 3 изложены соответственно в [1–3] и приведены ниже в обозначениях, принятых в данной работе.

Теорема 1. *Каждая непрерывная функция $f(\mathbf{x})$, заданная на единичном кубе d-мерного пространства, представима в виде*

$$f(\mathbf{x}) = \sum_{i=1}^{2d+1} \sigma_i \left(\sum_{j=1}^{d} g_{ij}(x_j) \right), \text{ где } \mathbf{x} = [x_1, \dots, x_d]^\mathsf{T}.$$

Функции $\sigma_i(\cdot)$, $g_{ij}(\cdot)$ непрерывны, причем $g_{ij}(\cdot)$ не зависят от выбора f.

Теорема 2. *Пусть φ — любая непрерывная сигмоидная функция, например, $\varphi(\xi) = 1/(1+e^{-\xi})$. Тогда, если дана любая непрерывная функция действительных переменных f на $[0,1]^n$ (или любое другое компактное подмножество $\mathbb{R}^n$) и $\varepsilon > 0$, то существуют векторы $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N, \boldsymbol{\alpha}$ и $\boldsymbol{\theta}$ и параметризованная функция $g(\cdot, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\theta}) : [0,1]^n \to \mathbb{R}$ такая, что*

$$|g(\mathbf{x}, \mathbf{W}, \alpha, \theta) - f(\mathbf{x})| < |\varepsilon|, \quad \mathbf{x} \in [0,1]^n,$$

где

$$g(\mathbf{x}, \mathbf{W}, \alpha, \theta) = \sum_{i=1}^{N} \alpha_i \varphi(\mathbf{w}_i^\mathsf{T} \mathbf{x} + \theta_i))$$

и

$$\mathbf{w}_i \in \mathbb{R}^n, \quad \alpha_i, \theta_i \in \mathbb{R}, \quad \mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_N), \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N),$$

и

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N).$$

Для следующей теоремы введем обозначения. В дальнейшем используется только функция активации ReLU для любого $n \geq 1$:

$$\sigma = \mathrm{ReLU}(x_1, \dots, x_n) = (\max\{0, x_1\}, \dots, \max\{0, x_n\}). \qquad (1)$$

Для $n \geq 1$ и непрерывной функции $f : [0,1]^n \to \mathbb{R}$ вводится норма

$$||f||_{C_0} := \sup_{\mathbf{x} \in [0,1]^n} |f(\mathbf{x})|.$$

Введем модуль непрерывности

$$\omega_f(\varepsilon) := \sup\{|f(\mathbf{x}) - f(\mathbf{y})| : |\mathbf{x} - \mathbf{y}| \le \varepsilon\}.$$

Обозначим множеством D_1 множество всех возможных нейронных сетей с входным слоем размером n и размером скрытых слоев s, которые достаточно точно аппроксимируют любую положительную непрерывную функцию на $[0,1]^n$. Добавим обозначение минимальной размерности скрытых слоев нейронной сети с входным размером n из D_1 как

$$\omega_{\min}(n) := \min_{s \in D_1}\{s\}.$$

Также обозначим множеством D_2 множество всех возможных нейронных сетей с входным слоем размером n и размером скрытых слоев s, которые достаточно точно аппроксимируют любую положительную выпуклую функцию на $[0,1]^n$. Добавим обзначение минимальной размерности скрытых слоев нейронной сети с входным размером n из D_2 как

$$\omega^{\text{conv}}_{\min}(n) := \min_{s \in D_2}\{s\}.$$

Теорема 3. *Пусть $n \ge 1$ и $f : [0,1]^n \to \mathbb{R}_+$ — положительная функция с нормой $||f||_{C_0} = 1$. Тогда:*

1. *Если f — непрерывная, то существует последовательность нейронных сетей прямого распространения $\mathcal{N}_k$ с функциями активации* (1), *размер входного слоя — n, скрытого — $n+2$, выходного слоя — 1 такие, что*

$$\lim_{k\to\infty} ||f - f_{\mathcal{N}_k}||_{C^0} = 0. \tag{2}$$

   В частности, $\omega_{\min}(n) \le n+2$ Кроме того, если зафиксировать $\varepsilon > 0$ и взять модуль непрерывности функции f $\omega_f(\varepsilon)$, то найдется нейронная сеть прямого распространения $\mathcal{N}_\varepsilon$ с функциями активации ReLu (1), *размерами входного слоя n, скрытого слоя — $n+3$, выходного слоя — 1, и*

$$\operatorname{depth}(\mathcal{N}_\varepsilon) = \frac{2 \cdot n!}{\omega_f(\varepsilon)^n}.$$

   такая, что

$$||f - f_{\mathcal{N}_\varepsilon}||_{C_0} \le \varepsilon.$$

2. *Если f — выпуклая, то существует последовательность нейронных сетей прямого распространения $\mathcal{N}_m$ с функциями активации $ReLU$* (1), *размерами входного слоя — n, скрытого слоя — $n+1$, и выходного слоя — 1, такие, что*

$$\lim_{m\to\infty} ||f - f_{\mathcal{N}_m}||_{C_0} = 0. \tag{3}$$

   В частности, верна такая оценка $\omega^{\text{conv}}_{\min}(n) \le n+1$. Более того, найдется $C > 0$ такая, что если f одновременно и выпуклая, и липшицева с константой

Липшица L, тогда нейронные сети $\mathcal{N}_m$ в (3) *могут быть выбраны такими, что они удовлетворяют*

$$\text{depth}(\mathcal{N}_m) = m, \ ||f - f_{\mathcal{N}_m}||_{C_0} \leq CLn^{\frac{3}{2}}m^{-\frac{2}{n}}.$$

3. *Если f — гладкая, то найдутся константа K, зависящая только от n и константа B, зависящая только от максимума первых K производных от f такие, что для каждого $m \geq 3$ нейронные сети $\mathcal{N}_m$ с размерностью $n+2$ в* (2) *могут быть выбраны такими что*

$$\text{depth}(\mathcal{N}_m) = m, \ ||f - f_{\mathcal{N}_m}||_{C_0} \leq B(m-2)^{-\frac{1}{n}}.$$

Список литературы

[1] *Kolmogorov A.* On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables// Proceedings of the USSR Academy of Sciences, 108 (1956), pp. 179–182; English translation: Amer. Math. Soc. Transl., 17 (1961), pp. 369–373.

[2] *Cybenko, G. V.* Approximation by Superpositions of a Sigmoidal function // Mathematics of Control Signals and Systems. — 1989. — T. 2, № 4. — C. 303—314.

[3] *Hanin B.* Universal function approximation by deep neural nets with bounded width and relu activations //arXiv preprint arXiv:1708.02691. – 2017.