Секция «Дискретная математика и математическая кибернетика»

О замене параметра в двухпараметрической оценке архитектуры нейросетевого аппроксиматора PL-функций

Научный руководитель – Половников Владимир Сергеевич

Шишляков Владимир Геннадьевич
Аспирант
Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра общих проблем управления, Москва, Россия
E-mail: bolotmaks@yandex.ru

Введение

Данная работа является продолжением работы [2] по построению нейросетевой архитектуры с оценкой количества нейронов в ней, приближающую всякую кусочно–линейную функцию на некотором достаточно объемном множестве с любой наперед заданной точностью. В работе [2] дается оценка архитектуры по двум параметрам — количество объемных классов эквивалентности и количество гиперплоскостей, порождающих эти классы. Однако, последний параметр трудно использовать на практике, поэтому в данной работе рассматривается замена данного параметра на размерность признакового пространства, которая всегда известна при постановке задачи.

Основные понятия и формулировка результата

Функцию $\psi : \mathbb{R} \to [0,1]$ назовем сигмоидной, если ψ не убывает на $\mathbb{R}$ и $\lim\limits_{x\to-\infty} \psi(x) = 0$, $\lim\limits_{x\to+\infty} \psi(x) = 1$. Функцию $\psi(x)$ назовем нечетной относительно y, если $\psi(x)-y = -(\psi(-x)-y)$.

Далее, ψ — сигмоидная функция, нечетная относительно $1/2$, $l_1, ..., l_k$ — гиперплоскости, разбивающие пространство $\mathbb{R}^n$ на классы эквивалентности $R^1, ..., R^s$.

Назовем класс R^i плоским, если $\exists l_j : R^i \subset l_j$. Все классы, не являющиеся плоскими, назовем объемными.

Возьмем $\forall \xi > 0$ и рассмотрим множества

$L_{i,\xi} = \{\overline{x} \in \mathbb{R}^n || l_i(\overline{x})| < \xi\}, i = 1, ..., k$. Обозначим $L_\xi = \bigcup\limits_{i=1}^{k} L_{i,\xi}$.

В теореме 1 рассматривается построение нейронных сетей над множеством $B_1 = \{c, \gamma \cdot x, \sum_n (x_1, ..., x_n), \prod_n (x_1, ..., x_n), \psi(x)\}$, которое назовем базисом построения.

Нейроном в базисе B_1 назовем всякую схему, вычисляющую одну из функций $\varphi(\sum\limits_{i=1}^{n} w_i \cdot x_i + c)$ (нейроны–сумматоры) или $\varphi(\prod\limits_{i=1}^{n} w_i \cdot x_i + c)$ (нейроны–продукторы), где $\varphi(x)$ полагается равной $\psi(x)$ или x.

Также дадим определение кусочно-линейной функции, следуя [1]. Будем говорить, что $f(\overline{x})$ является кусочно-линейной, если $f(\overline{x})|_{R^j} = \overline{b_j} \cdot \overline{x} + d_j$.

Основными выводами данной работы является теорема 1.

Теорема 1. *Пусть классы эквивалентности такие, что s' классов являются объемными, все ограниченные классы являются симплексами и все неограниченные классы образуются естественными продолжениями $n-1$-мерных граней симплексов, а $f(\overline{x})$ — кусочно–линейная функция, заданная над данными классами эквивалентности.*

Тогда $\forall\varepsilon > 0, \forall\xi > 0, \forall R > 0$ существует нейронная сеть $G(\overline{x})$ над базисом B_1 такая, что $\sup\limits_{\overline{x}\in O_R(\overline{0})\setminus L_\xi} |G(\overline{x}) - f(\overline{x})| < \varepsilon$. Причем данная нейронная сеть обладает следующей архитектурой:

1. На первом слое не более $s'\cdot(n+1)$ нейронов–сумматоров, имеющих функцию активации $\varphi(x) = \psi(x)$;

2. На втором слое $2s'' \leqslant 2s'$ нейронов, из которых s'' нейронов имеют функцию активации $\varphi(x) = \psi(x)$, а остальные s'' нейронов — $\varphi(x) = x$;

3. На третьем слое потребуется s' нейронов-продукторов с тождественной функцией активации;

4. На четвертом слое потребуется один нейрон-сумматор с тождественной функцией активации.

Автор выражает благодарность научному сотруднику Половникову В.С. и доценту Часовских А.А. за постановку задачи.

Источники и литература

1) Половников В. С. Об оптимизации структурной реализации нейронных сетей. Диссертация на соискание степени кандидата наук. МГУ. Москва. 2007.

2) Шишляков В. Г. Об улучшениях нейросетевой архитектуры для приближения кусочно-линейных функций. Интеллектуальные системы. Теория и приложения. 2021. Т. 25, №4. С.271–274.