Об улучшениях нейросетевой архитектуры для приближения кусочно-линейных функций

В. Г. Шишляков[1]

Работа представляет собой продолжение исследований в области оценок архитектур нейронных сетей, достаточных для приближения определенных классов функций. В работе рассматривается вопрос об улучшении оценки сверху архитектуры нейронной сети, которая хорошо приближает зависимости, описывающиеся кусочно-линейными функциями.

Ключевые слова: схема из функциональных элементов, нейронные сети, аппроксимация функций, кусочно-линейные функции.

1. Введение

С точки зрения схем функциональных элементов [1] нейронные сети были впервые рассмотрены в работе [2]. Также в работе [2] были рассмотрены оценки архитектур нейронных схем, восстанавливающих кусочно-постоянные функции над классическим базисом МакКаллока и Питтса, а также восстанавливающих кусочно-линейные функции над базисом $B = \{c, \gamma \cdot x, \sum_n(x_1, ..., x_n), \theta(x), F(x, y)\}$, где $F(x,y) = \begin{cases} x, & y \geqslant 0 \\ 0, & y < 0 \end{cases}$, $\theta(x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$.

В дальнейшем в работе [3] были получены результаты, оценивающие сверху количество нейронов в нейронных сетях над видоизмененным базисом $B_2 = \{c, \gamma \cdot x, \sum_n(x_1, ..., x_n), \prod_n(x_1, ..., x_n), \psi(x)\}$, где $\psi(x)$ - некоторая сигмоидная функция. В базисе B_2 рассматривалась задача аппроксимации кусочно-линейных функций, т.к. задача восстановления, вообще говоря, была уже невозможна.

В данной работе предлагается результат, улучшающий оценку сверху, представленную в работе [3] для базиса B_2, а также существенное улучшение оценки, которое получается при выборе в базисе B_2 функции $\psi(x)$, нечетной относительно уровня $\frac{1}{2}$.

[1] *Шишляков Владимир Геннадьевич* — аспирант каф. общих проблем управления мех.-мат. ф-та МГУ, e-mail: bolotmaks@yandex.ru.

Shishlyakov Vladimir Gennad'evich — graduate student, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of General Problems of Control.

2. Основные понятия и формулировка результата

Введем основные понятия, используемые в данной работе.

Базисом нейронной сети называется некоторое множество функциональных элементов, где каждый функциональный элемент представляет из себя пару $(S, f(x_1, ..., x_n))$, в которой $f(x_1, ..., x_n) : \mathbb{R}^n \to \mathbb{R}$, а S - сопоставленный ей графический объект с n входными стрелками и одной выходной (кратко – входы и выход объекта S). Входам объекта приписаны слева направо переменные $x_1, ..., x_n$ функции f, выходу приписан выход функции f.

Далее в тексте, для краткости, графические обозначения объектов опускаются.

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

В теоремах 1, 2, сформулированных ниже, рассматривается построение нейронных сетей над следующим базисом:

$$B_1 = \{c, \gamma \cdot x, \sum_n (x_1, ..., x_n), \prod_n (x_1, ..., x_n), \psi(x)\} \tag{1}$$

В базисе (1) используются следующие виды функций:

1) Сумматор – класс функций вида $\sum_n (x_1, ..., x_n) = x_1 + ... + x_n$

2) Продуктор – класс функций вида $\prod_n (x_1, ..., x_n) = x_1 \cdot ... \cdot x_n$

3) Константа – класс функций, каждая из которых выдает некоторую константу (все эти функции не имеют существенных переменных)

4) Усилитель – функция, умножающая пришедший на вход аргумент x на фиксированную константу γ

5) Функция активации – произвольная сигмоидная функция $\psi(x)$

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

Все нейроны первого типа будем называть нейронами-сумматорами, а нейроны второго типа – нейронами-продукторами.

Введем понятие слоя нейронной сети.

1) Множество нейронов, все входы которых не подсоединены к выходам каких-либо функциональных элементов, назовем нейронами первого слоя.

2) Пусть определено множество нейронов n-го слоя. Тогда $n+1$-ым слоем назовем все нейроны, для которых выполняется, что хотя бы один вход подсоединен к выходу нейрона n-го слоя, а все оставшиеся входы подсоединены либо к выходам нейронов из слоев $\{1, 2, ..., n\}$, либо не подсоединены ни к каким нейронам.

Таким образом, можно комбинировать не отдельные элементы базиса схемы, а целые схемы, реализующие нейроны. При комбинации нейронов друг с другом, будут получаться различные функции, которые и будут исследоваться в данной работе. Особый интерес для данного исследования представляют так называемые кусочно-линейные функции. Дадим их определение, следуя определениям из [2].

Пусть пространство $\mathbb{R}^n$ разбивается на классы эквивалентности $R^1, ..., R^s$ гиперплоскостями $l_1, ..., l_k$. Будем говорить, что $f(\overline{x})$ является кусочно-линейной, если $f(\overline{x})|_{R^j} = \overline{b_j} \cdot \overline{x} + d_j$.

Также введем ряд определений и обозначений, которые позволяют облегчить формулирование дальнейших теорем.

Пусть даны гиперплоскости $l_1, ..., l_k$, разбивающие пространство $\mathbb{R}^n$ на классы $R^1, ..., R^s$. Назовем класс R^p плоским, если $\exists l_i : R^p \subset l_i$. Все классы, не являющиеся плоскими, назовем объемными.

Также введем несколько полезных обозначений. Пусть $l_1, ..., l_k$ - гиперплоскости. 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 $L_\xi = \bigcup\limits_{i=1}^{k} L_{i,\xi}$.

Будем говорить, что функция $f(x)$ является нечетной относительно уровня y, если $f(x) - y = -(f(-x) - y)$.

Основными выводами данной работы являются две теоремы, сформулированные ниже.

Теорема 1. *Пусть $l_1, ..., l_k$ - гиперплоскости, которые разбивают пространство $\mathbb{R}^n$ на s классов эквивалентности $R^1, ..., R^s$, из которых s' классов являются объемными, 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 (1) такая, что выполняется $\sup\limits_{\overline{x} \in O_R(\overline{0}) \setminus L_\xi} |G(\overline{x}) - f(\overline{x})| < \varepsilon$. Причем данная нейронная сеть обладает следующей архитектурой:

1. На первом слое потребуется не более $2k$ нейронов-сумматоров, имеющих функцию активации $\varphi(x) = \psi(x)$;

2. На втором слое потребуется $2s'' \leqslant 2s'$ 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.

Теорема 2. *Пусть выполняются все условия теоремы 1 с тем лишь условием, что функция $\psi(x)$ является нечетной относительно уровня $\frac{1}{2}$. Тогда количество нейронов на первом слое в нейронной сети $G(\overline{x})$ из теоремы 1 можно сократить с $2k$ до k, оставив все оставшиеся слои без изменений.*

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

Список литературы

[1] Яблонский С.В., *Введение в дискретную математику*, eds. физ.-мат.лит., «Наука», Москва, 1986.

[2] Половников В.С., *Об оптимизации структурной реализации нейронных сетей*, дисс. … канд. физ.-матем. наук, МГУ, Москва, 2007.

[3] Шишляков В.Г., *Построение архитектуры нейронной сети, достаточной для приближения всякой кусочно-линейной функции с любой наперед заданной точностью*, Проблемы теоретической кибернетики. Материалы XIX международной конференции (Казань, 27 сентября – 01 октября 2021 года), eds. Ю.И. Журавлев, МАКС Пресс, Москва, 2021.

Improvements to neural network architecture for approximating particle-linear functions

Shishlyakov V.G.

This work continues series of works in neural networks architectures estimatons approximating some classes of functions. The following paper considers the problem of improving upper-bound estimation of neural network closely approximating particle-linear dependences.

Keywords: schema of functional elements, neural networks, functions approximation, particle-linear functions.

References

[1] YAblonskij S.V., *Introduction to discrete mathematics*, eds. fiz.-mat.lit., «Science», Moscow, 1986 (In Russian).

[2] Polovnikov V.S., *On optimization of the structural implementation of neural networks*, Ph.D. Thesis, MSU, Moscow, 2007 (In Russian).

[3] Shishlyakov V.G., *Building a neural network architecture sufficient to approximate any particle-linear function with any predetermined accuracy*, Problems of theoretical cybernetics. Materials of the XIX international conference (Kazan, 27 sep. – 01 oct. 2021), eds. YU.I. ZHuravlev, MAKS Press, Moscow, 2021 (In Russian).