МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение

высшего образования

«Национальный исследовательский

Нижегородский государственный университет им. Н.И. Лобачевского»

(ННГУ)

Институт информационных технологий, математики и механики

Кафедра: Алгебры, геометрии и дискретной математики

Направление подготовки: 01.03.01 – «Математика»

Профиль подготовки: «Общий профиль»

КУРСОВАЯ РАБОТА

Тема:

«Нейронные сети Колмогорова-Арнольда»

Выполнил:
студент группы 3822Б1МА1
Никитин И. И.
ф.и.о.

____
подпись

Научный руководитель:
д.ф.-м.н., доцент, Золотых Н. Ю.
учёная степень, учёное звание, ф.и.о.

____
подпись

Нижний Новгород

2024

Содержание

Введение

Целью данной работы является исследование нейронных сетей Колмогорова-Арнольда (KAN) и проверка утверждений, описанных в работе, в которой дано их первоначальное определение.

Для достижения цели необходимо было решить следующие **задачи**:

1. Исследовать теоретический фундамент KAN - теорему Колмогорова-Арнольда и её развитие;
2. Исследовать метод обобщения теоремы для определения структуры KAN по аналогии с MLP;
3. Проверить и разработать корректное утверждение о возможности предотвращения COD с помощью KAN.

Объем и структура работы. Курсовая работа состоит из введения, 2 глав и заключения. Полный объем работы составляет 11 страниц. Список литературы содержит 13 наименований.

Глава 1. Теорема Колмогорова-Арнольда

1.1. Определения

n-мерный единичный (гипер)куб – $\mathbb{I}^n := [0,1]^n$.

Сигмоидальная функция – функция $\sigma : \mathbb{R} \to \mathbb{I} : \quad \lim_{t\to-\infty} \sigma(t) = 0, \quad \lim_{t\to\infty} \sigma(t) = 1.$

Ступенчатая функция σ-типа – функция вида $f(x) = \sum_{i=1}^{k} a_i \sigma(b_i x + c_i)$.

Разбиение единицы – Набор неотрицательных функций $\{f_\beta : M \to \mathbb{I}\}$ на топологическом пространстве $\mathcal{M}$, такой, что $\forall x \in \mathcal{M} : \sum_\beta f_\beta(x) = 1$.

1.2. Теоретическая основа

Лемма 1.1. *(Колмогоров, 1956):* Любая $f \in C(\mathbb{I}^2)$ может быть равномерно приближена формулой вида:

$$P_1(x)Q(R_1(x) + y) + P_2(x)Q(R_2(x) + y), \tag{1.1}$$

причём $P_1, P_2, Q, R_1, R_2 \in C(\mathbb{R})$.

Схема доказательства.

Зафиксируем $\epsilon > 0$ и пусть

$$Q(x) = f(n\epsilon, y), \quad x \in [2n, 2n+1]$$

– функция, «склеенная» из сечений f;

$$R_1(x) = 4k, \quad \frac{x}{\epsilon} \in \left[-\frac{2}{3}, \frac{2}{3}\right] + 2k;$$

$$R_2(x) = 4k + 2, \quad \frac{x}{\epsilon} \in \left[-\frac{2}{3}, \frac{2}{3}\right] + 2k + 1.$$

– функции, переводящие x в отрезки в образе Q;

$$P_1(x), P_2(x)$$

– периодические функции разбиения единицы. Тогда (1.1) приближает f так, как и требовалось.

□

Впоследствии результаты о возможности приближения непрерывных функций были усилены утверждениями о наличии *точных представлений* непрерывных функций:

Лемма 1.2. *(Колмогоров, 1956 [1]):* Любая $f \in C(\mathbb{I}^n), n \geq 4$ выразима как суперпозиция функций *трёх* переменных $g \in C(\mathbb{R}^3)$.

Лемма 1.3. *(Арнольд, 1957 [2]):* Любая $f \in C(\mathbb{I}^3)$ выразима как суперпозиция функций *двух* переменных $g \in C(\mathbb{R}^2)$.

В частности, Лемма 1.3 имеет важное

Следствие. *(Гильберт, 1900, «13 проблема»):* Любое уравнение 7 степени имеет решение, выразимое как суперпозиция непрерывных функций двух переменных.

И, наконец, основной результат, являющийся теоретическим фундаментом для рассматриваемой архитектуры нейронных сетей.

Теорема 1.4. *(Колмогоров, 1957 [3]):* Пусть $f : \mathbb{I}^n \to \mathbb{R}$ – любая непрерывная функция n переменных. Тогда она выразима в виде:

$$f(x_1, ..., x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right), \tag{1.2}$$

где $\Phi_q \in C(\mathbb{R}), \phi_{q,p} \in C(\mathbb{I}^n)$ – функции *одной* переменной, причём $\phi_{p,q}$ можно выбрать *независимо* от f.

Замечание. В одной из формулировок теоремы функции ϕ могут быть липшицевыми [4], однако уже для гладких функций ϕ утверждение теоремы в общем случае неверно, даже если функция f гладкая. [5]

Теорема 1.4 в применении к нейронным сетям обсуждалась ранее, как в поддерживающих [6,7], так и в критикующих [8] работах. Один из основных аргументов критики основывается на том, что в теореме не указываются способы явно параметризовать функции ϕ. Однако в более поздних статьях указаны как, оказывающийся практически полезным, способ использовать «аппроксимационный» вариант теоремы (в случае, когда приближения функций ϕ известны) [6], так и конструктивный вариант доказательства самой теоремы [9].

Самым распространённым конструктивным методом построения функции аппроксимации на основе нейронной сети является метод обратного распространения ошибки («backpropagation»). Для его применения «передаточные функции» (в нашем случае это Φ и ϕ) должны быть дифференцируемы, что существенно мешает точно представить функцию в виде, который предлагает Теорема 1.4. Однако, когда достаточно получить только приближение функции с допущением некоторой погрешности ϵ, мы можем воспользоваться следующим результатом.

Теорема 1.5. *(Куркова, 1991 [10]):* Для любого $n > 2$, $\sigma : \mathbb{R} \to \mathbb{I}$ – сигмоидальной функции, $f : \mathbb{I}^n \to \mathbb{R}$ – непрерывной функции и $\epsilon > 0$ найдутся $k \in \mathbb{N}$ и ступенчатые функции σ-типа Φ_q и $\phi_{q,p}$ такие, что для любого $\boldsymbol{x} \in \mathbb{I}^n$:

$$\left| f(x_1, ..., x_n) - \sum_{q=1}^{k} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right) \right| < \epsilon. \tag{1.3}$$

Теорема 1.5 оказывается практически полезной для обучения сетей с помощью обратного распространения ошибки. Далее определим структуру нейронной сети по аналогии со структурой MLP, но используя вместо универсальной теоремы об аппроксимации теорему Колмогорова-Арнольда.

Глава 2. Нейронные сети Колмогорова-Арнольда

2.1. Структура нейронной сети

Мы рассмотрим конструкцию сети, предложенную в работе [11].

Определение. *Слой* – матрица $\mathbf{\Phi} = \{\phi_{q,p}\}$, где $p = 1, 2, ..., n_{\text{out}}, \quad q = 1, 2, n_{\text{in}}$, а функции ϕ имеют параметры, изменяемые в процессе обучения сети.

Определение. *Сигнатура сети* – для сети из L слоёв – упорядоченный набор $(n_0, ..., n_L) \subset \mathbb{N}$, где n_i - число узлов в i-м слое сети.

Обозначим i-й узел в l-м слое сети как (l, i), а его численное значение - $x_{l,i}$. Между слоями l и $l+1$ находится $n_l * n_{l+1}$ функций ϕ. Функцию, связывающую узел (l, i) с узлом $(l+1, j)$ обозначим

$$\phi_{l,j,i}, \quad l = 0, ..., L-1, \quad i = 1, ..., n_l, \quad j = 1, ..., n_{l+1}. \tag{2.1}$$

Аргумент функции $\phi_{l,j,i}$ мы назовём «входным сигналом» $x_{l,i}$, а её значение – «выходным сигналом». Затем определим выходной сигнал узла $(l+1, j)$: $x_{l+1,j} = \sum_{i=1}^{n_l} \phi_{l,j,i}(x_{l,i}), \quad j = 1, ..., n_{l+1}$. В матричной записи выходной сигнал слоя сети будет иметь вид:

$$\boldsymbol{x}_{l+1} = \mathbf{\Phi}_l \boldsymbol{x}_l = \begin{pmatrix} \phi_{l,1,1}(\cdot) & \dots & \phi_{l,1,n_l}(\cdot) \\ \vdots & & \vdots \\ \phi_{l,n_{l+1},2}(\cdot) & \dots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix} \boldsymbol{x}_l. \tag{2.2}$$

Полностью сеть, как и в случае с MLP, описывается как композиция слоёв. Для вектора входных данных $\boldsymbol{x}_0 \in \mathbb{R}^{n_0}$:

$$\begin{gathered} \operatorname{kan}(\boldsymbol{x_0}) = (\mathbf{\Phi}_{L-1} \circ \mathbf{\Phi}_{L-2} \circ ... \circ \mathbf{\Phi}_0)\boldsymbol{x_0}, \\ \operatorname{mlp}(\boldsymbol{x_0}) = (\mathrm{W}_{L-1} \circ \sigma \circ \mathrm{W}_{L-2} \circ \sigma \circ ... \circ \mathrm{W}_0)\boldsymbol{x_0}. \end{gathered} \tag{2.3}$$

Теорема 1.4 показывает, что внутренние функции образуют слой, где $n_{\text{in}} = n$ и $n_{\text{out}} = 2n+1$, а внешние функции образуют слой, где $n_{\text{in}} = 2n+1$ и $n_{\text{out}} = 1$. Таким образом, справедливо следующее

Утверждение 2.1. Выражения Колмогорова-Арнольда в уравнении (1.2) являются частным случаем сети из 2 слоёв.

Доказательство. Используя общую формулу:

$$f(x_0, ..., x_{n_0}) = (y_0, ..., y_{n_L}),$$

$$y_{i_L} = \sum_{i_{L-1}=1}^{n_{L-1}} \phi_{L-1,i_L,i_{L-1}} \left(\sum_{i_{L-2}=1}^{n_{L-2}} \phi_{L-2,i_{L-1},i_{L-2}} \left(\cdots \left(\sum_{i_0=1}^{n_0} \phi_{0,i_1,i_0}(x_{i_0}) \right) \cdots \right) \right)$$

для сигнатуры $(n_0, n_1, n_2) \equiv (n, 2n+1, 1),\ i_1 \equiv q,\ i_0 \equiv p,\ \phi_{1,1,q} \equiv \Phi_q,\ \phi_{0,q,p} \equiv \phi_{q,p}$:

$$y_{i_2} = \sum_{i_1=1}^{n_1} \phi_{1,1,i_1} \left(\sum_{i_0=1}^{n_0} \phi_{0,i_1,i_0}\big(x_{i_0}\big) \right) \Rightarrow f(x_1, ..., x_n) = y_1 = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{0,q,p}\big(x_p\big) \right).$$

□

2.2. Оценка ограничений архитектуры

Целью этого раздела является проверка предложения, данного в работе [11] о том, что предложенная структура нейронной сети позволяет избежать «проклятия размерности». В теории глубокого обучения этот термин может иметь следующее

Определение. *Проклятие размерности (COD)* – экспоненциальная зависимость количества параметров $P(n)$ приближающей $f(x_1, ..., x_n)$ функции $\tilde{f}$ от размерности многообразия, на котором задана функция f.

Теорема 2.2. Пусть $f \in C(\mathbb{I}^n)$ имеет представление вида

$$f(\boldsymbol{x}) \equiv \mathrm{kan}(\boldsymbol{x}) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \cdots \circ \Phi_0)\boldsymbol{x}, \tag{2.4}$$

причём $\boldsymbol{\Phi} \in C^k(\mathbb{I})$. Тогда KAN, соответствующая этому представлению, имеет $P(n) = n^{2+2/k} \cdot \epsilon^{-1/k}$ параметров.

Доказательство. Разбиение $[a,b] \subset \mathbb{R}$ осуществляется с помощью набора узлов $\boldsymbol{\tau} := (\tau_i)_1^{n+k}$. Определим мелкость разбиения $|\boldsymbol{\tau}|$ и размер сетки G:

$$|\boldsymbol{\tau}| := \max_i \Delta\tau_i, \quad G := |\boldsymbol{\tau}|^{-1}. \tag{2.5}$$

Пусть $\Phi_q^G, \phi_{q,p}^G$ – приближения функций $\Phi_q, \phi_{q,p}$ B-сплайнами. Тогда погрешность приближения функции f сетью будет ограничена следующим образом:

$$\begin{gathered} \|f - \mathrm{kan}\| \leq \sum_{q=1}^{2n+1} \big(A_q(n,G) + B_q(G)\big), \\ A_q(n,G) = \left| \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p} \right) - \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}^G \right) \right|, \quad B_q(G) = \big|\Phi_q - \Phi_q^G\big|. \end{gathered} \tag{2.6}$$

Предположим, что функции Φ_q – L_q-липшицевы, тогда:

$$A_q(n,G) = L_q \sum_{p=1}^{n} \big|\phi_{q,p} - \phi_{q,p}^G\big|. \tag{2.7}$$

Обозначим погрешности приближения и определим L:

$$\epsilon_\Phi = \max_{q,G} \big|\Phi_q - \Phi_q^G\big|, \quad \epsilon_\phi = \max_{q,p,G} \big|\phi_{q,p} - \phi_{q,p}^G\big|, \quad L = \max_q L_q, \tag{2.8}$$

тогда

$$\|f - \mathrm{kan}\| \leq (2n+1) \cdot \big(\epsilon_\Phi + L \cdot n \cdot \epsilon_\phi\big). \tag{2.9}$$

Таким образом,

$$\|f - \mathrm{kan}\| \leq \epsilon \quad \Longrightarrow \quad \epsilon_\Phi = O\Big(\frac{\epsilon}{n}\Big), \quad \epsilon_\phi = O\Big(\frac{\epsilon}{n^2}\Big). \tag{2.10}$$

Из теории приближений [12] известно, что для функции $g \in C^k(\mathbb{I})$ B-сплайнами порядка k выполняется следующее неравенство типа Джексона — Стечкина:

$$\big\|g - g^G\big\| \leq \hat{C}_k \cdot G^{-k} \cdot \big\|D^k g\big\| = C_k \cdot G^{-k}, \tag{2.11}$$

где G-число функций в базисе сплайна. Отсюда получаем оценку для числа узлов G:

$$\epsilon_g \sim C_k G^{-k} \quad \Longrightarrow \quad G \sim \sqrt[k]{\frac{C_k}{\epsilon_g}}. \tag{2.12}$$

Из (2.10) получаем, что

$$G_\Phi \sim \sqrt[k]{\frac{C_k}{\epsilon_\Phi}}, \quad G_\phi \sim \sqrt[k]{\frac{C_k}{\epsilon_\phi}} \quad \Longrightarrow \quad G_\Phi \sim \sqrt[k]{\frac{C_k n}{\epsilon}}, \quad G_\phi \sim \sqrt[k]{\frac{C_k n^2}{\epsilon}}. \tag{2.13}$$

Оценим количество параметров в представлении:

$$\begin{aligned} P(n) \; &= \; (2n+1)\big(G_\Phi + nG_\phi\big) \; \sim \; (2n+1)\left(\sqrt[k]{\frac{C_k n}{\epsilon}} + n\sqrt[k]{\frac{C_k n^2}{\epsilon}}\right) \; \sim \\ &\sim \; \big(n^{1+1/k} + n^{2+2/k}\big) \cdot \epsilon^{-1/k} \; \sim \boxed{n^{2+2/k} \cdot \epsilon^{-1/k}.} \end{aligned} \tag{2.14}$$

□

Из теории приближений в многомерных пространствах [13] известно, что нижняя граница погрешности выразима в виде

$$\epsilon \geq C \cdot P(n)^{-k/n}. \tag{2.15}$$

Таким образом, если при $n \to \infty$ фиксировать ϵ, мы получим полиномиальный рост количества параметров: $P(n) = n^{2+2/k}$. Так действительно можно предотвратить COD. Однако, чтобы применение приближения было практически полезно, необходимо, чтобы $\epsilon \to 0$ при $n \to \infty$, в силу того, что для пространств большой размерности данные становятся более разреженными. Поэтому, положив, к примеру, $\epsilon = n^{-\alpha}$ мы будем иметь:

$$P(n) = O\big(n^{2+1/k+\alpha/k}\big). \tag{2.16}$$

Если взять $\alpha = k \Longrightarrow P(n) = O(n^3)$, и COD не наблюдается. Но это ограничение на α возможно только для малого класса функций, предположительно, гладких (например такого, в котором Теорема 1.4 даёт представление с количеством слагаемых $S \ll n$). Для функций общего вида обыкновенно необходимо, чтобы $\alpha \propto n$, и COD сохраняется.

Заключение

Основные результаты работы заключаются в следующем:

1. На основе анализа предыдущих исследований о способах определения нейронных сетей с помощью теоремы Колмогорова-Арнольда была разобрана работа [11]. Проверены основные, наиболее амбициозные утверждения статьи. Утверждения, подвергнутые сомнению, были исправлены, либо пояснены.
2. Определены перспективы дальнейшего исследования: формальное доказательство эквивалентности MLP и KAN со структурной (алгебраической) точки зрения. Также предлагается определить «место» рассмотренной архитектуры в алгебраической теории CDL.

Словарь терминов

CDL: Categorial Deep Learning (Алгебраическая теория категориального глубокого обучения) 9

COD: Curse of Dimensionality (Проклятие размерности) 2, 7, 8

KAN: Kolmogorov-Arnold Network (Нейронная сеть Колмогорова-Арнольда) 2, 7, 9

MLP: Multi-layer perceptron (Многослойный перцептрон) 2, 5, 6, 9

Список литературы

1. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных. Доклады Академии Наук СССР, 1956. т. 108, № 2. сс. 179–182
2. Арнольд В. И. О функциях трёх переменных. Доклады Академии Наук СССР, 1957. т. 114, № 4. сс. 679–681
3. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения. Доклады Академии Наук СССР, 1957. т. 114, № 5. сс. 953–956
4. Sprecher D. On the Structure of Continuous Functions of Several Variables. American Mathematical Society, 1965. т. 115. сс. 340–355
5. Фридман Б. Л. Улучшение гладкости функций в теореме А. Н. Колмогорова о суперпозициях. Доклады Академии Наук СССР, 1967. т. 177. сс. 1019–1022
6. Kůrková V. Kolmogorov's Theorem Is Relevant. Neural Computation, 1991. т. 3, № 4. сс. 617–622
7. Ismayilova A., Ismailov V. On the Kolmogorov neural networks [электронный ресурс]. 2023. URL: https://arxiv.org/abs/2311.00049
8. Girosi F., Poggio T. Representation Properties of Networks: Kolmogorov's Theorem Is Irrelevant. Neural Computation, 1989. т. 1. сс. 465–469
9. Braun J., Griebel M. On a Constructive Proof of Kolmogorov’s Superposition Theorem. Constructive Approximation, 2009. т. 30. сс. 653–675
10. Kůrková V. Kolmogorov's theorem and multilayer neural networks // Neural Networks. 1992. т. 5. сс. 501–506
11. Liu Z. и др. KAN: Kolmogorov-Arnold Networks [электронный ресурс]. 2024. URL: https://arxiv.org/abs/2404.19756
12. Boor C. de. A Practical Guide to Splines. 1978. сс. 145–155
13. Pinkus A. Approximation theory of the MLP model in neural networks. Acta Numerica, 1999. т. 8. сс. 143–195