

Обучение композиции операторов в сетях Колмогорова-Арнольда для обнаружения мультипликативных зависимостей



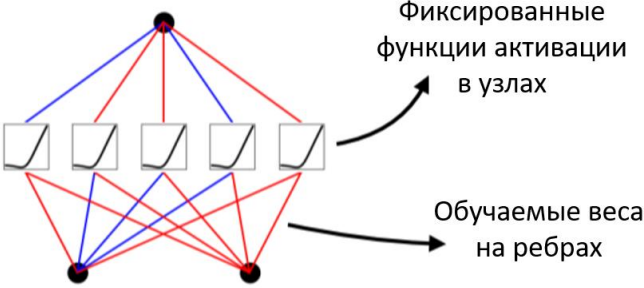
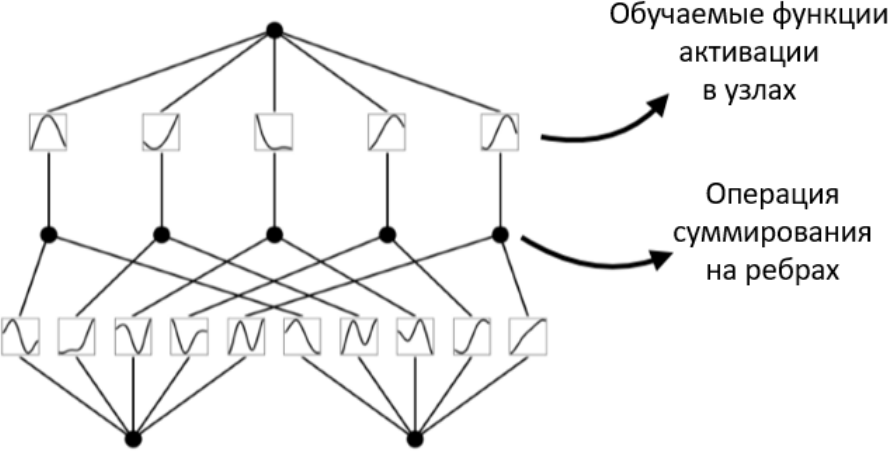
Шепелев Данил Игоревич
аспирант 2 курса
Научный руководитель:
Демяненко Яна Михайловна, к.т.н., доцент

Содержание



- Сети Колмогорова-Арнольда
- Задачи модифицированной модели
- KAN + обучаемая структура операций
- Визуальный анализ интерпретируемости
- Результаты для задач символьной регрессии и автоматического поиска признаков
- Итоги

Сети Колмогорова-Арнольда

Многослойный персептрон (MLP)	Сеть Колмогорова-Арнольда (KAN)
Теорема об универсальной аппроксимации $f(x) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(w_i x + b_i)$	Теорема Колмогорова-Арнольда $f(x) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
 Фиксированные функции активации в узлах Обучаемые веса на ребрах	 Обучаемые функции активации в узлах Операция суммирования на ребрах
$\text{MLP}(x) = (W_3 \circ \sigma_2 \circ W_2 \circ \sigma_1 \circ W_1)(x)$	$\text{KAN}(x) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(x)$

Обучаемые В-сплайны на рёбрах

Каждое ребро φ_{ij} — одномерный В-сплайн:

$$\varphi(x) = w_b b(x) + w_s \sum c_k B_k(x)$$

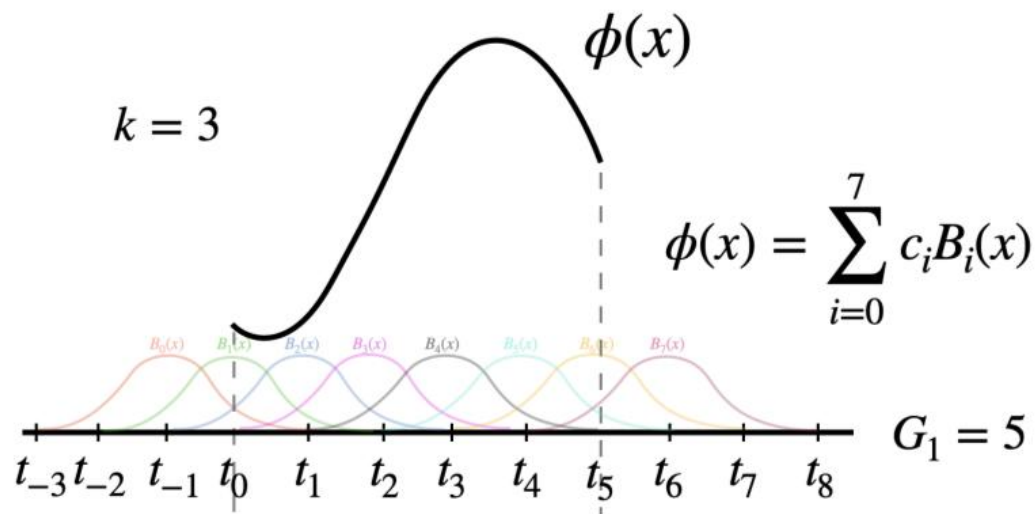
$b(x)$ — базовая активация (SiLU)

$B_k(x)$ — В-сплайн базисные функции

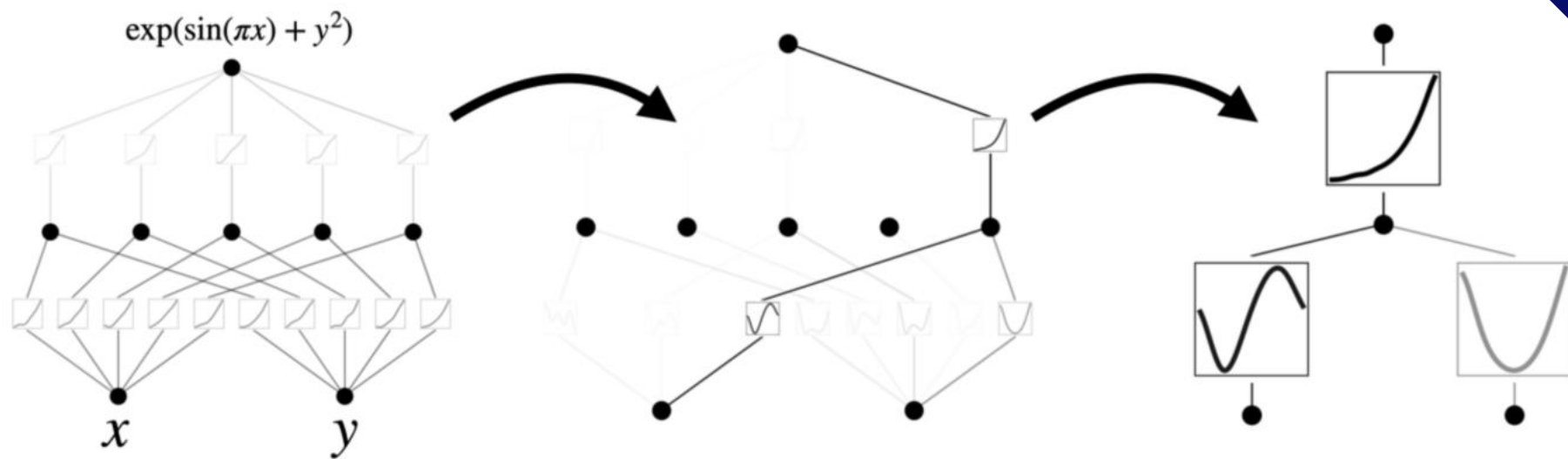
c_k — обучаемые коэффициенты

Выход слоя (аддитивная структура):

$$z_j = \sum_i \varphi_{ij}(x_i)$$



Интерпретируемость KAN



Задачи модифицированной модели

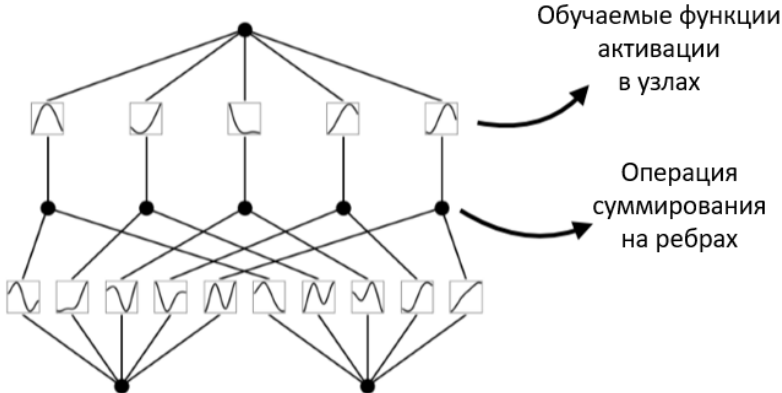
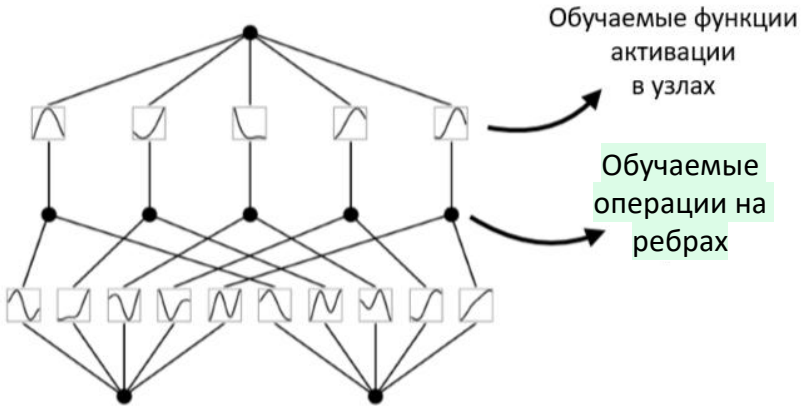
Символьная регрессия	Автоматическое извлечение признаков	Требуемые свойства модели
Найти аналитическую формулу $f(x)$, объясняющую данные. Ключевой вызов: формулы содержат как произведения, так и суммы, которые KAN не выделяет автоматически. Пример: $x_1 x_2 + \frac{x_3}{x_4} (x_3 * x_4^{-1})$	KAN-слой + полносвязные слои (end-to-end). Ожидание: KAN-часть автоматически находит нелинейные зависимости среди сырых признаков, полносвязная часть их комбинирует.	• Автоматическое обучения операций + и $\times$ • Дифференцируемость (end-to-end обучение)

Пример выражений для регрессии:

$$y = x_0 x_2 + x_4 \sin(2\pi x_3) + x_2 \quad | \quad y = x_0 / x_1 \quad | \quad y = x_0 / x_1 + x_2 / x_3 \quad | \quad y = (x_0 / x_1) / (x_2 / x_3)$$

LOKAN: KAN + обучаемая структура операции

Совместная оптимизация непрерывных параметров сплайнов и дискретной структуры операций

Сеть Колмогорова-Арнольда (KAN)	KAN + обучаемые операции
 Обучаемые функции активации в узлах Операция суммирования на ребрах	 Обучаемые функции активации в узлах Обучаемые операции на ребрах
$\text{KAN}(x) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(x)$	$\text{LOKAN}(x) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(x)$

MultKAN

(Вручную введенное наличие произведения при инициализации модели)

Структура операций

Идея модификации:

Гибкое распределение входных данных слоя (операндов) по G группам:

$$y = x_1x_5 + x_2x_3 + x_4 \quad | \quad y = x_1x_4x_5 + x_2 + x_3 \quad | \quad y = x_1 + x_2 + x_3 + x_4 + x_5 \quad | \quad \dots$$

$$G = \left\lfloor \frac{d_{in}}{2} \right\rfloor + 1, \quad \text{где}$$

$\left\lfloor \frac{d_{in}}{2} \right\rfloor$ – группы произведения, $+1$ – группа суммирования

$$y = x_1x_5 + x_2x_3 + x_4$$

$$G = 3$$

2 мультипликативные группы

+

1 аддитивная группа

$G \setminus X$	x_1	x_2	x_3	x_4	x_5
Гр. $\times_1$	1	0	0	0	1
Гр. $\times_2$	0	1	1	0	0
Гр. $+$	0	0	0	1	0

Математика операций

$$y = x_1x_5 + x_2x_3 + x_4$$

G \ X	x_1	x_2	x_3	x_4	x_5
Гр. $\times_1$	1	0	0	0	1
Гр. $\times_2$	0	1	1	0	0
Гр. $+$	0	0	0	1	0

Мягкое произведение (для сохранения градиентов и знаков):

$$\begin{aligned} prod_{o,g} &= \prod term_{i,o,g} term_{i,o,g} \\ &= 1 - p_{i,o,g} + p_{i,o,g} \cdot postspline y_{i,o} \end{aligned}$$

Итоговый выход нейрона O :

$$z_o = sum_o + \sum prod_{o,g}$$

Аддитивный вклад:

$$sum_o = \sum p_{i,o,G} \cdot postspline y_{i,o}$$

Непрерывная релаксация структуры операций

- Проблема: выбор структуры операций – задача невыпуклой дискретной оптимизации
- Решение: непрерывная релаксация через *softmax* с температурой τ – сведение к выпуклой оптимизации

G \ X	x_1	x_2	x_3	x_4	x_5
Гр. $\times_1$	1	0	0	0	1
Гр. $\times_2$	0	1	1	0	0
Гр. +	0	0	0	1	0

$$\text{softmax}\left(\frac{\text{logits}}{\tau}\right)$$

$$p_{i,o,g} = \frac{\exp\left(\frac{l_{i,o,g}}{\tau}\right)}{\sum g' \exp\left(\frac{l_{i,o,g'}}{\tau}\right)}$$

$l_{i,o,g}$ – обучаемые логиты (параметры)
 $\tau \rightarrow 0$: распределение стремится к дискретному
 $\tau \rightarrow \infty$: равномерное распределение (начало обучения)

Функция потерь и регуляризация

$$L = L_{task} + \lambda_{ops}(t) \cdot H_{ops} + \lambda_{L1} \cdot L1_{acts} + \lambda_{ent} \cdot H_{acts}$$

L_{task}	H_{ops}
MSE или BCE Основная задача: соответствие прогноза целевым значениям. Регрессия → MSE Классификация → бинарная кросс-энтропия.	Энтропия распределения операций по группам. Минимизация → дискретные назначения входов к группам.
$L1_{acts}$	H_{acts}
L1-норма масштабов активаций сплайнов. Поощряет разреженность: неважные рёбра получают нулевой вес.	Энтропия матрицы активаций по строкам и столбцам. Поощряет фокусировку: один доминирующий вход на нейрон.

Стратегия обучения: три расписания

$$f(t) = f_{start} \cdot \left(\frac{f_{end}}{f_{start}} \right)^{\frac{t}{T}} - \text{общая форма расписания}$$

$lr(t)$ - скорость обучения	$\tau(t)$ - температура softmax	$\lambda_{H_{ops}}(t)$ - коэффициент регуляризации H_{ops}
$lr_{end} = 1e-3$ $lr_{start} = 0.1$	$\tau_{end} = 0.1$ $\tau_{start} = 2.0$	$\lambda_{end} = 10.0$ $\lambda_{start} = 1e-2$

Начало: $\tau \uparrow, \lambda \downarrow \rightarrow$
мягкое исследование

$t = 0$

Середина: структура
формируется

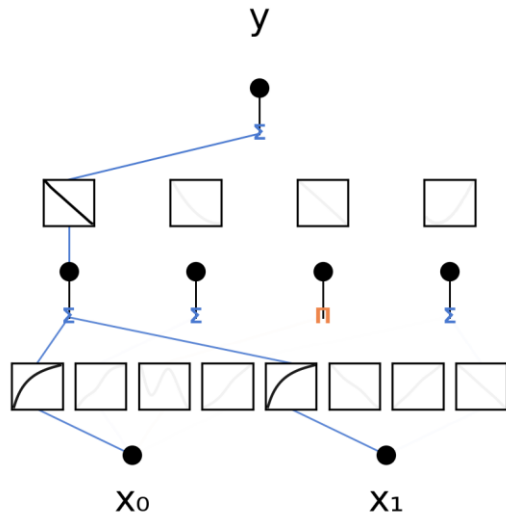


Конец: $\tau \downarrow, \lambda \uparrow \rightarrow$
чёткие группы

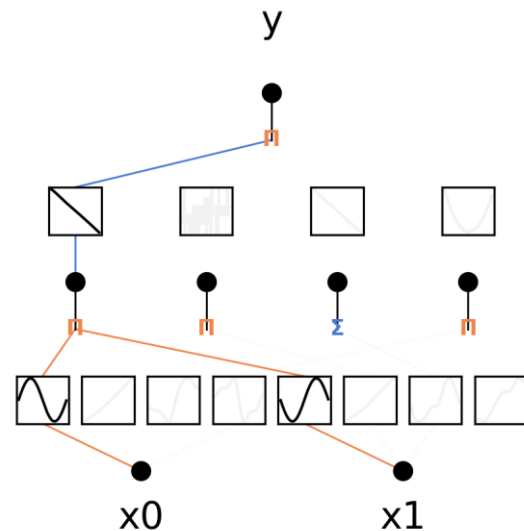
$t = T$

Визуализация: символьная регрессия

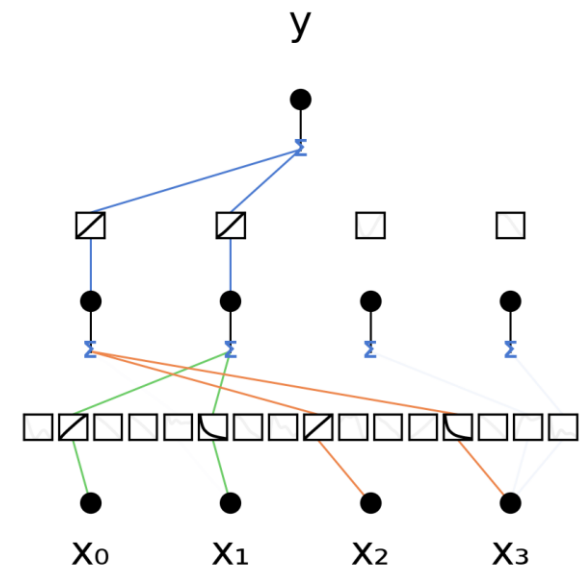
Inverse pair: $1/x_0 + 1/x_1$



Sine product: $\sin(2\pi x_0) \cdot \sin(2\pi x_1)$



Stage 2: $x_0/x_1 + x_2/x_3$



Бенчмарк: формулы Фейнмана

N_RUNS = 3 | EPOCHS = 500 | Порог успеха: $R^2 > 0.99$

Уравнение	Выражение	d_{in}	LOKAN	MultKAN	MLP	PySR
I.6.20a	$\sqrt{2} \cdot \exp(-\theta^2/2)/(2\sqrt{\pi})$	1	3/3 ✓	3/3 ✓	0/3	3/3 ✓
I.14.4	$0.5 \cdot k_s \cdot x^2$	2	3/3 ✓	3/3 ✓	0/3	3/3 ✓
I.8.4	$\sqrt{((x_2-x_1)^2+(y_2-y_1)^2)}$	4	1/3	0/3	0/3	1/3
I.11.19	$x_1y_1 + x_2y_2 + x_3y_3$	6	3/3 ✓	2/3	3/3 ✓	3/3 ✓
I.12.2	$q_1q_2/(4\pi\epsilon r^2)$	4	1/3	0/3	0/3	3/3 ✓
I.12.4	$q_1/(4\pi\epsilon r^2)$	3	2/3	2/3	0/3	3/3 ✓

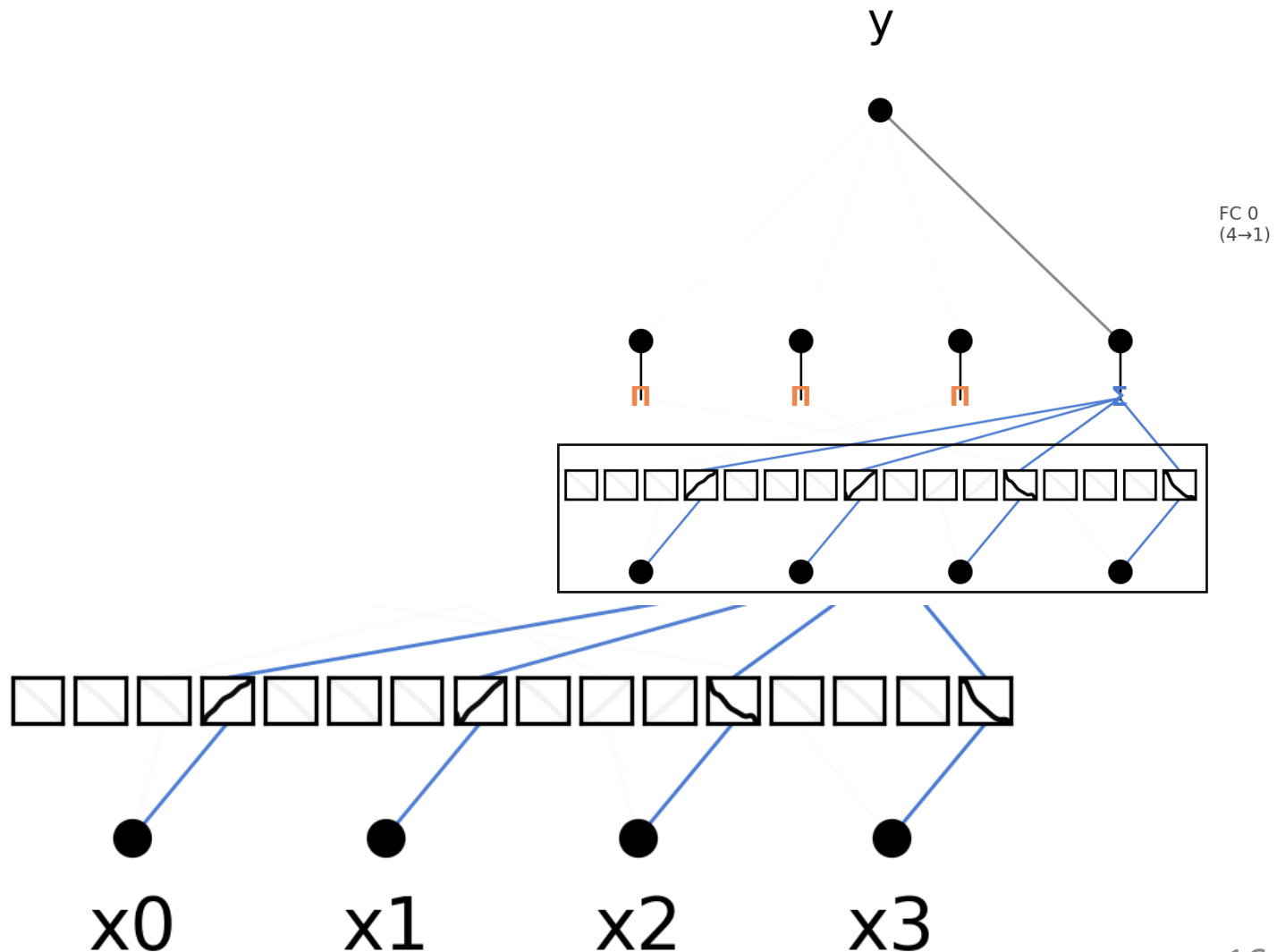
Бенчмарк: символьная регрессия (отношение и взаимодействие)

N_RUNS = 3 | EPOCHS = 500 | Метрика: R^2 (успех > 0.99)

Задача	LOKAN R^2	MultKAN R^2	MLP R^2	PySR R^2	LOKAN успех
x_0/x_1 (2D)	0.9999 ± 0.0001	0.9963 ± 0.0021	0.8378 ± 0.0891	1.0000 ± 0.0000	3/3
$x_0/x_1 + x_2/x_3$ (4D)	0.9973 ± 0.0035	0.9673 ± 0.0246	0.7288 ± 0.0417	1.0000 ± 0.0000	3/3
$x_0 \cdot x_1 + x_2 \cdot x_3$ (4D)	0.9921 ± 0.0110	0.9971 ± 0.0017	0.9217 ± 0.0044	1.0000 ± 0.0000	2/3
$(x_0/x_1 + x_2/x_3) / (x_4/x_5 + x_6/x_7)$ (8D)	0.8823 ± 0.0561	0.9275 ± 0.0272	0.7798 ± 0.0089	0.7857 ± 0.1193	0/3

Визуализация: поиск признаков

Classification: $x_0/x_3 + x_1/x_2 > 2.2$



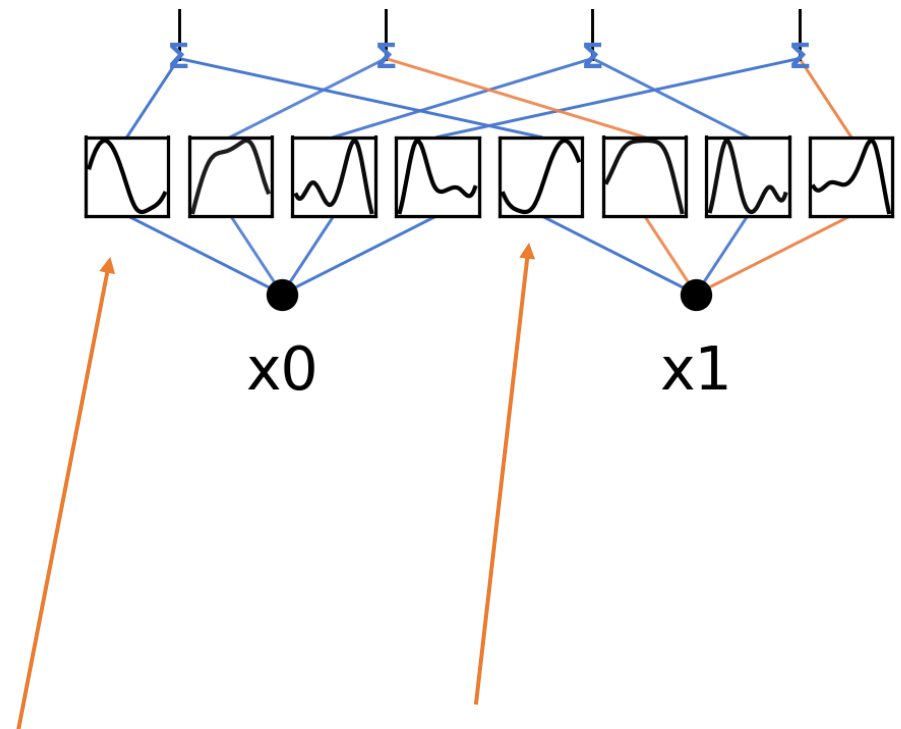
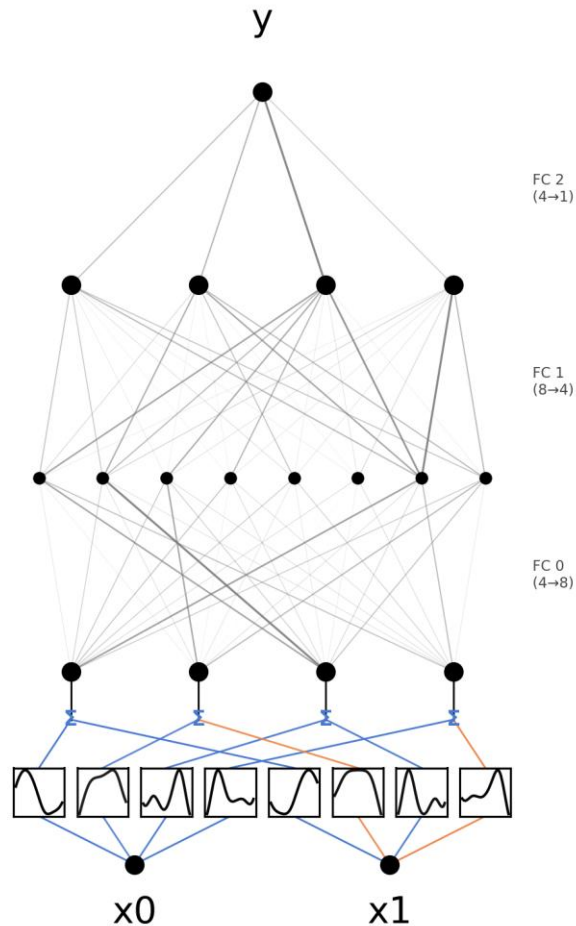
Бенчмарк: автоматическое извлечение признаков и классификация

Архитектура: LOKAN([in→4→4]) + FC([4→sigmoid]) | N_RUNS = 3 | EPOCHS = 500

Задача	LOKAN-FC Acc	MLP-small Acc	MLP-large Acc	Ep→90% (LOKAN)
T1: $x_0/x_1 + x_2/x_3 > 2.2$ (4D)	0.982 ± 0.005	0.980 ± 0.004	0.985 ± 0.007	5.3 эпох
T2: $(x_0+x_1) \cdot (x_2+x_3) > 0.49$ (4D)	0.968 ± 0.013	0.982 ± 0.013	0.998 ± 0.002	12.0 эпох
T3: $x_0 \cdot x_1 \cdot x_2 + x_3 \cdot x_4 \cdot x_5 > 0.18$ (6D)	0.967 ± 0.006	0.920 ± 0.016	0.972 ± 0.010	5.7 эпох

Визуализация: поиск признаков

Hybrid: $\sin(2\pi x_0) \cdot \sin(2\pi x_1)$ -- KAN+FC network



Бенчмарк: обнаружение нелинейностей (LOKAN + FC)

Архитектура: LOKAN([in→4→4]) + FC([4→8→4→1]) | N_RUNS = 3 | EPOCHS = 500

Задача	LOKAN+FC R ²	MLP R ²	Ep→R ² ≥0.9 5 (LOKAN)	Ep→R ² ≥0.9 5 (MLP)	Детекция нелинейности
T1: $1/x_0 + 1/x_1$	1.0000 ± 0.0000	0.9996 ± 0.0001	6.7	23.0	2/2 ✓ (1/x ✓)
T2: $\sin(2\pi x_0) \cdot \sin(2\pi x_1)$	0.9994 ± 0.0003	0.9962 ± 0.0025	7.0	77.0	2/2 ✓ (sin ✓)
T3: $\exp(x_0) + 1/(0.5+x_1)$	0.9999 ± 0.0001	0.9994 ± 0.0003	6.7	8.0	1/2 (exp ✓, 1/x X → linear)
T4: $\sin(2\pi x_0) + 1/(0.5+x_1) + x_2^2$	0.9971 ± 0.0039	0.9983 ± 0.0009	40.7	100.7	1/3 (sin ✓, 1/x X, x ² X)

Итоги



- Разработана модификация KAN повышающая интерпретируемость
- Предложена непрерывная релаксация дискретной структуры операций через softmax с температурой
- Обеспечено сквозное обучение моделей: сплайны + структура операций + полносвязные слои совместно
- LOKAN превосходит MultKAN на задачах с поиском отношений и взаимодействий
- Ускорение сходимости LOKAN по сравнению с MLP в 2–10 раз (регрессия и классификация + полносвязные слои)



Спасибо за внимание!

Шепелев
Данил Игоревич

аспирант
младший научный сотрудник
НИТЦ нейротехнологий ЮФУ
