

Kolmogorov-Arnold Networks (KANs): Альтернатива многослойным перцепtronам в Tabular DL

Данил Руденко, Булгаков Георгий

Московский физико-технический институт

22 апреля 2025 г.

Предпосылки и формулы

- **Теорема Колмогорова-Арнольда:** Любая многомерная функция может быть представлена как композиция одномерных функций и операций сложения.

$$f(x) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$$

- **Проблема MLPs:** Фиксированные активационные функции на узлах и линейные веса.
- **Решение:** Замена линейных весов на обучаемые одномерные функции, параметризованные сплайнами.

Идея KANs

• Архитектура:

- Каждый слой — матрица одномерных функций.
- Глубокие KANs позволяют лучше аппроксимировать сложные функции.

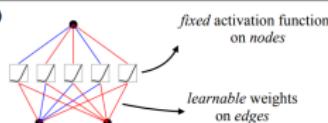
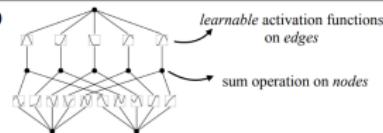
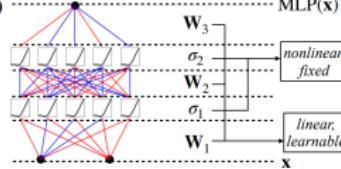
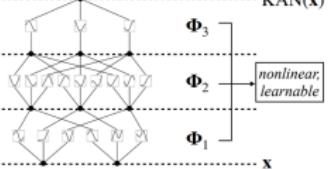
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(c)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) 	(b) 
Formula (Deep)	$\text{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\text{KAN}(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$
Model (Deep)	(c) 	(d) 

Рис.: Архитектура KANs

Почему Tabular DL

- Новые DL модели (такие как TabM, TabR, FT-T) начали показывать результаты сравнимые с GBDT
- Табличные данные зачастую имеет малую размерность
- В статьях [2] и [3] были проведены эксперименты KAN на табличных данных с многообещающими результатами
- Однако в статьях [2] и [3] не был изучен полный потенциал архитектуры KAN

Embeddings для численных фичей

- Идея: Преобразовать все численные атрибуты в векторы по некоторым (обучаемым) правилам \rightarrow на вход в основную модель
- Согласно статье [4], MLP оснащенная такими эмбедингами значительно обходит обычный MLP

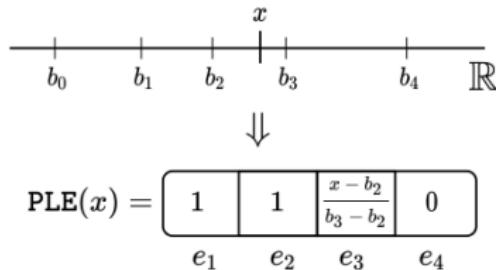
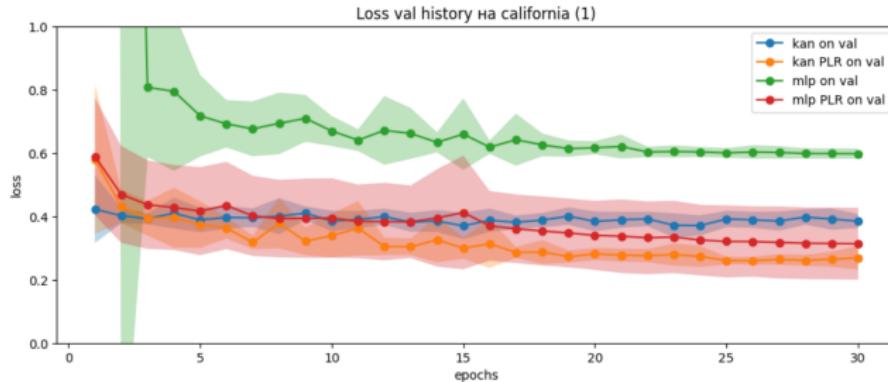


Figure 1: The piecewise linear encoding (PLE) in action for $T = 4$ (see Equation 1).

- Стоит попробовать применить такие эмбединги перед KAN

Результаты применения

- PLR (Periodic Linear Relu) и PLE-Q (Piecewise Linear Embeddings - Quantile) эмбединги показали лучшие результат среди различных вариантов
- Применение эмбедингов предотвращает переобучение KAN



- Эмбединги лучше влияют на MLP, поэтому KAN и MLP с эмбедингами сравнимы по итоговым результатам

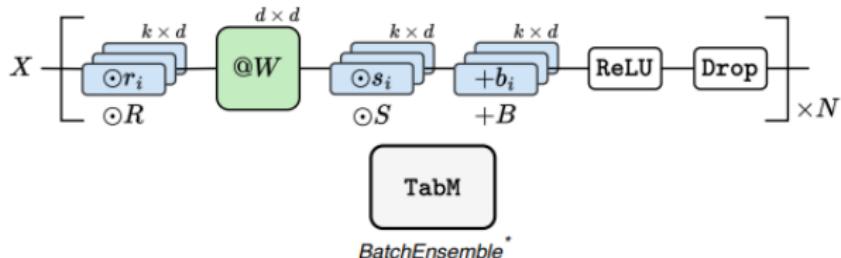
Различные виды полиномов/сплайнов

- В статье [1] в качестве нелинейности внутри KAN используют B-spline (преимущества: локальность, ограниченность)
- В статье [3] дополнительно предложили использовать полиномы Чебышева(Cheby-KAN) и Гауссовые нормальные функции (fast-KAN)

Model	Avg Rank	Train time Ratio	Test time Ratio
KAN	2,8	2,73	3,81
Small KAN	1,8	2,6	3,65
MLP	3	1	1
ChebyKAN	3,2	1,39	1,55
FastKAN	4	1,72	2,02

Ансамблирование

- В статье [5] был предложен метод эффективного моделирования k MLP различных моделей -> State of the art результаты



- Можно попробовать заменить MLP на KAN в этой архитектуре (ближайшая задача)

Наша работа до сих пор

- В подробностях разобрались в работе forward и back propagation
- Познакомились с работой модуля torch - реализация собственного оптимизатора, backprop в torch, а так же познакомились и реализовали передовые оптимизационные алгоритмы (Adam, AdamW, RMSProp, Adagrad, Sarah)
- Подробно изучили оригинальную статью по KAN ([1]), а также статьи с экспериментами ([2] и [3]), статьи по эмбедингам и ансамблированию ([4] и [5])
- Написали пайплайн для тюнинга и тестирования моделей с различными backbone и эмбедингами
- Проведение многочисленных экспериментов на популярных табличных датасетах

Заключение

- KAN стабильно обходят MLP в табличных задачах, что подтверждает результаты из статей [2] и [3]
- Однако по результатам экспериментов наличие Embeddings вносит больший вклад, чем разная архитектура
- Использование Chebyshev KAN не дает значительных улучшений и плохо работает с эмбедингами
- FastKAN и SmallKAN - лучше всего работают с примененными эвристиками.
- Есть перспектива использования ансамбля KAN