

Kolmogorov Arnold Networks as an alternative to multilayered perceptrons in tabular DL

*Д.В. Руденко¹, Г.А. Булгаков¹
Г.Л. Молодцов^{1,2}, Д.О. Медяков^{1,2}*

¹Московский физико-технический институт (национальный исследовательский университет)

²Лаборатория методов оптимизации(BRAIn Lab)

Введение

Исторически методы глубокого обучения (DL) демонстрировали ограниченную эффективность при работе с табличными данными, состоящими из категориальных, числовых и бинарных признаков. Традиционно классические модели машинного обучения, такие как XGBoost и CatBoost, значительно превосходили нейронные сети в этой области. Однако в последние годы направление Tabular DL активно развивалось, и современные архитектуры на основе полносвязных перцептронов (MLP) смогли превзойти лучшие решения среди ансамблей деревьев.

В данной работе исследуется возможность применения этих достижений к новой архитектуре нейронных сетей — Kolmogorov-Arnold Networks (KANs) [1], основанной на теореме Колмогорова-Арнольда. Согласно этой теореме, любая многомерная функция может быть представлена в виде композиции одномерных функций. В отличие от MLP, где используются фиксированные активационные функции и линейные преобразования, KANs заменяют линейные веса на обучаемые одномерные функции, параметризованные сплайнами. Такой подход потенциально повышает адаптивность модели и точность аппроксимации сложных зависимостей в табличных данных.


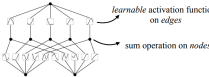
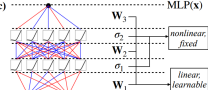
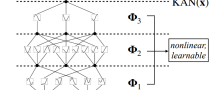
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{M(x)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a) 	(b) 
Formula (Deep)	$MLP(\mathbf{x}) = (W_3 \circ \sigma_2 \circ W_2 \circ \sigma_1 \circ W_1)(\mathbf{x})$	$KAN(\mathbf{x}) = (\Phi_3 \circ \Phi_2 \circ \Phi_1)(\mathbf{x})$
Model (Deep)	(c) 	(d) 

рис 1: Архитектура MLP vs KAN

Методология

В работах [2] и [3] авторы провели масштабные эксперименты с KAN на табличных данных, продемонстрировавшие многообещающие результаты. Как показали исследования, KAN стабильно превосходит традиционные MLP по качеству предсказаний на различных датасетах. В работе [3] дополнительно были предложены модификации базовой архитектуры KAN - FastKAN и ChebyshevKAN, использующие альтернативные подходы к параметризации нелинейных функций. Однако в этих исследованиях не был полностью раскрыт потенциал как MLP, так и KAN-архитектур.

Для повышения эффективности моделей в настоящей работе предлагается два ключевых усовершенствования:

1. Применение эмбедингов числовых признаков. В частности, используются методы Piecewise Linear Encoding (PLE) и Periodic Linear ReLU (PLR) [4], которые преобразуют числовые атрибуты в векторные представления. Этот подход позволяет:
 - Существенно снизить риск переобучения
 - Улучшить обобщающую способность модели
2. Использование продвинутых методов ансамблирования [5]. Предложенный подход позволяет эмулировать работу ансамбля из k моделей, при этом:
 - Усреднение предсказаний обеспечивает более высокую точность
 - Уменьшается дисперсия ошибок

Экспериментальная часть и результаты

Для обеспечения объективного сравнения моделей был проведен тщательный подбор гиперпараметров как для KAN-based, так и для MLP-based архитектур. Оптимизация выполнялась с использованием модуля Optuna для автоматизированного поиска оптимальных параметров, а также фреймворка Weights & Biases (wandb) для трекинга экспериментов и визуализации результатов

Полученные результаты демонстрируют следующие ключевые закономерности:

1. KAN-архитектуры стабильно показывают более высокое качество по сравнению с MLP в задачах Tabular DL, что подтверждает выводы предыдущих исследований [2, 3]
2. Наибольший вклад в улучшение метрик вносит не сама архитектура KAN, а применение современных методов эмбедингов:
 - Модели с PLE и PLR демонстрируют сравнимую эффективность независимо от базовой архитектуры
 - Качество MLP с продвинутыми эмбедингами приближается к показателям KAN
3. Сравнительный анализ вариантов KAN выявил:
 - Наилучшую производительность у Fast-KAN и Small-KAN
 - Относительно низкую эффективность Cheby-KAN
 - Оптимальное соотношение точности и вычислительных затрат у Small-KAN

Заключение

Работа демонстрирует, что KANs являются перспективной альтернативой MLP в задачах Tabular DL, особенно в сочетании с современными методами эмбедингов. Дальнейшие исследования могут быть направлены на оптимизацию архитектуры KANs, разработку новых методов параметризации функций и интеграцию KANs в ансамблевые подходы.

Литература

- [1] Ziming Liu and Yixuan Wang and Sachin Vaidya and Fabian Ruehle and James Halverson and Marin Soljačić and Thomas Y. Hou and Max Tegmark KAN: Kolmogorov-Arnold Networks 2025
- [2] Poeta, Eleonora and Giobergia, Flavio and Pastor, Eliana and Cerquitelli, Tania and Baralis, Elena A Benchmarking Study of Kolmogorov-Arnold Networks on Tabular Data 2024
- [3] Ali Eslamian and Alireza Afzal Aghaei and Qiang Cheng TabKAN: Advancing Tabular Data Analysis using Kolmogorov-Arnold Network 2025
- [4] Yury Gorishniy and Ivan Rubachev and Artem Babenko On Embeddings for Numerical Features in Tabular Deep Learning 2023
- [5] Yury Gorishniy and Akim Kotelnikov and Artem Babenko TabM: Advancing Tabular Deep Learning with Parameter-Efficient Ensembling 2025