

# Kolmogorov-Arnold Networks (KANs): Альтернатива многослойным перцептронам

Данил Руденко, Булгаков Георгий

Московский физико-технический институт

18 марта 2025 г.

- **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.

## Архитектура:

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

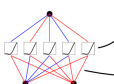
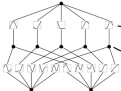
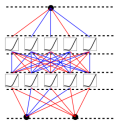
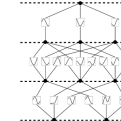
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	Universal Approximation Theorem	Kolmogorov-Arnold Representation Theorem
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(e)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left( \sum_{p=1}^n \phi_{q,p}(x_p) \right)$
Model (Shallow)	(a)  fixed activation functions on nodes learnable weights on edges	(b)  learnable activation functions on edges sum operation on nodes
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)  MLP(x) $\mathbf{W}_3$ $\sigma_2$ $\mathbf{W}_2$ $\sigma_1$ $\mathbf{W}_1$ x nonlinear, fixed linear, learnable	(d)  KAN(x) $\Phi_3$ $\Phi_2$ $\Phi_1$ x nonlinear, learnable

Рис.: Архитектура KANs

# Возможности и преимущества

- **Компактность:** KANs демонстрируют более быстрые законы масштабирования по сравнению с MLPs.
- **Интерпретируемость:** Легко визуализировать и понимать, как работает модель.

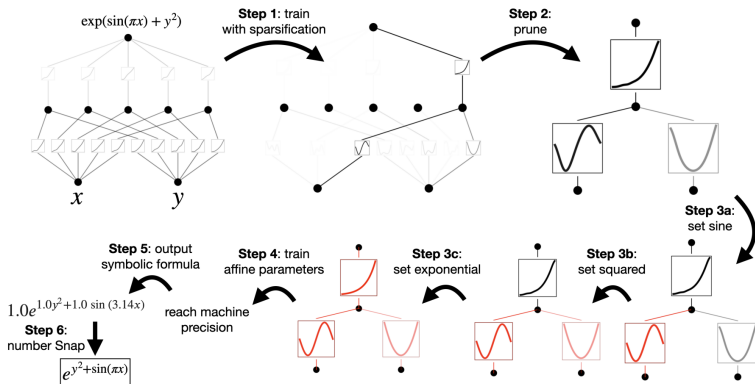


Figure 2.4: An example of how to do symbolic regression with KAN.

- **Сложность обучения:** Не является GPU эффективной, поэтому плохо прикладывается к задачам большой размерности
- **Вычислительная сложность:** Требуют больше ресурсов для обучения и выполнения.
- **Ограниченные эксперименты:** Пока что KANs тестировались только на задачах малого масштаба.

# Что уже было сделано в оригинальной статье

- **Теоретическое обоснование:** Обобщение теоремы Колмогорова-Арнольда на произвольные ширину и глубину.
- **Эксперименты:** Демонстрация превосходства KANs над MLPs на задачах аппроксимации функций и решения дифференциальных уравнений.

# Наша работа до сих пор

- В подробностях разобрались в работе forward и back propogation
- 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 большую часть оригинальной статьи по KAN

- Реализовать KAN своими силами, повторить эксперименты из статьи
- Заменить feed forward слои известных моделей на KAN и сравнить результаты
- Попробовать применить KAN в маломерных задачах (Например, tabular dl)



- KANs представляют собой перспективную альтернативу MLPs, особенно в задачах, где важны точность и интерпретируемость.
- Они открывают новые возможности для улучшения современных моделей глубокого обучения.
- Будущие исследования могут значительно расширить область применения KANs и улучшить их производительность.