Сети Колмогорова-Арнольда vs Персептрон: Решение обратной задачи георазведки

Г.А. Куприянов[1,2], И.В. Исаев[2,3], С.А. Доленко[2]

[1]Физический факультет Московского государственного университета имени М.В. Ломоносова, Москва, Россия

[2] НИИ ядерной физики им. Д.В. Скобельцына Московского государственного университета им. М.В. Ломоносова, Москва, Россия

[3] Институт радиотехники и электроники им. Котельникова Российской академии наук, Москва, Россия

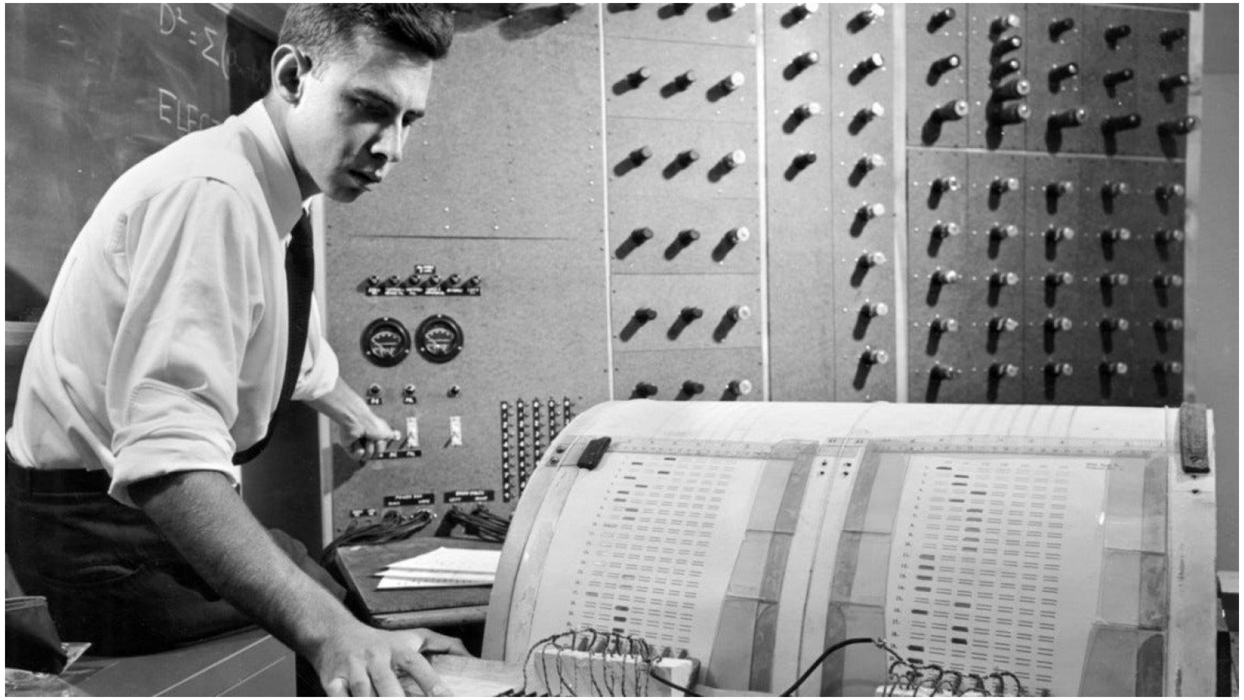
Исследование выполнено за счёт гранта Российского научного фонда №24-11-00266, https://rscf.ru/project/24-11-00222/

План доклада

1. Сети Колмогорова-Арнольда (КАН)
2. Предметная область: задача разведочной геофизики
3. Сравнение результатов применения персептрона и КАН

Первая искусственная нейронная сеть

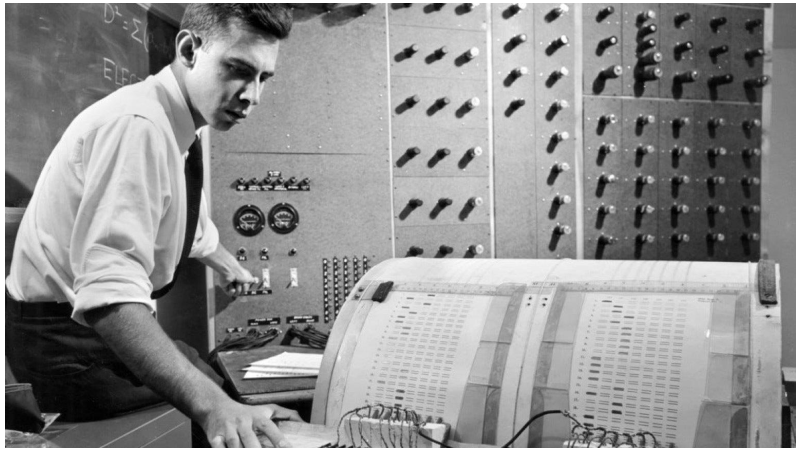
Фрэнк Розенблат
(1928-1971)



Первый персептрон «Марк-1»
(1960)

Первая искусственная нейронная сеть

Фрэнк
Розенблат



«Марк-1»

(1960)

$$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$$

Универсальная теорема аппроксимации персептроном (Цыбенко, 1989)

13-я проблема Гильберта

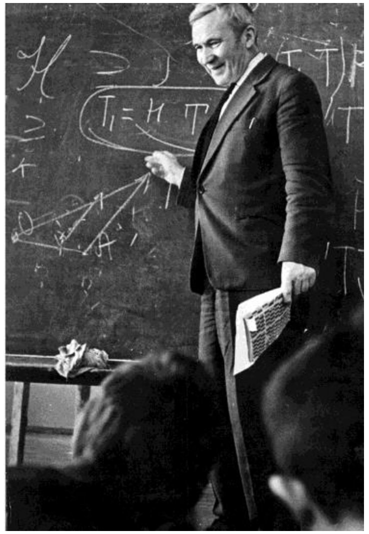
Можно ли представить функцию многих переменных в виде суперпозиции функций двух переменных?

13-я проблема Гильберта

Можно ли представить функцию <u>многих переменных</u> в виде <u>суперпозиции</u> функций <u>двух переменных</u>?

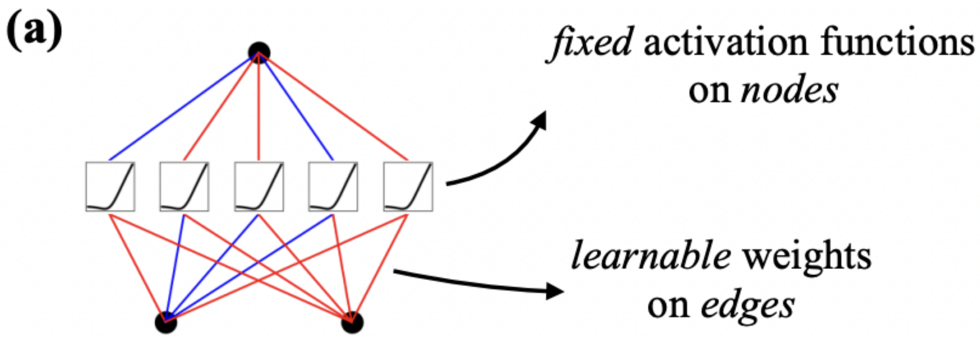
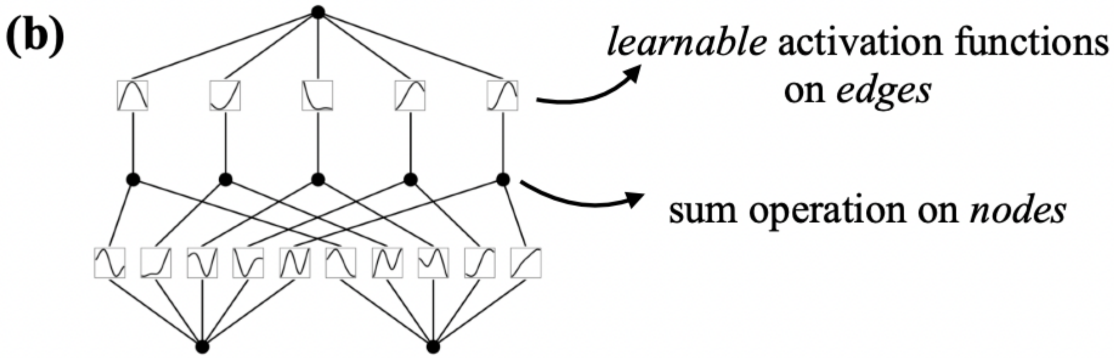
Можно, достаточно даже функций <u>одной</u> переменных:

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



Теорема Колмогорова-Арнольда

Персептрон и КАН

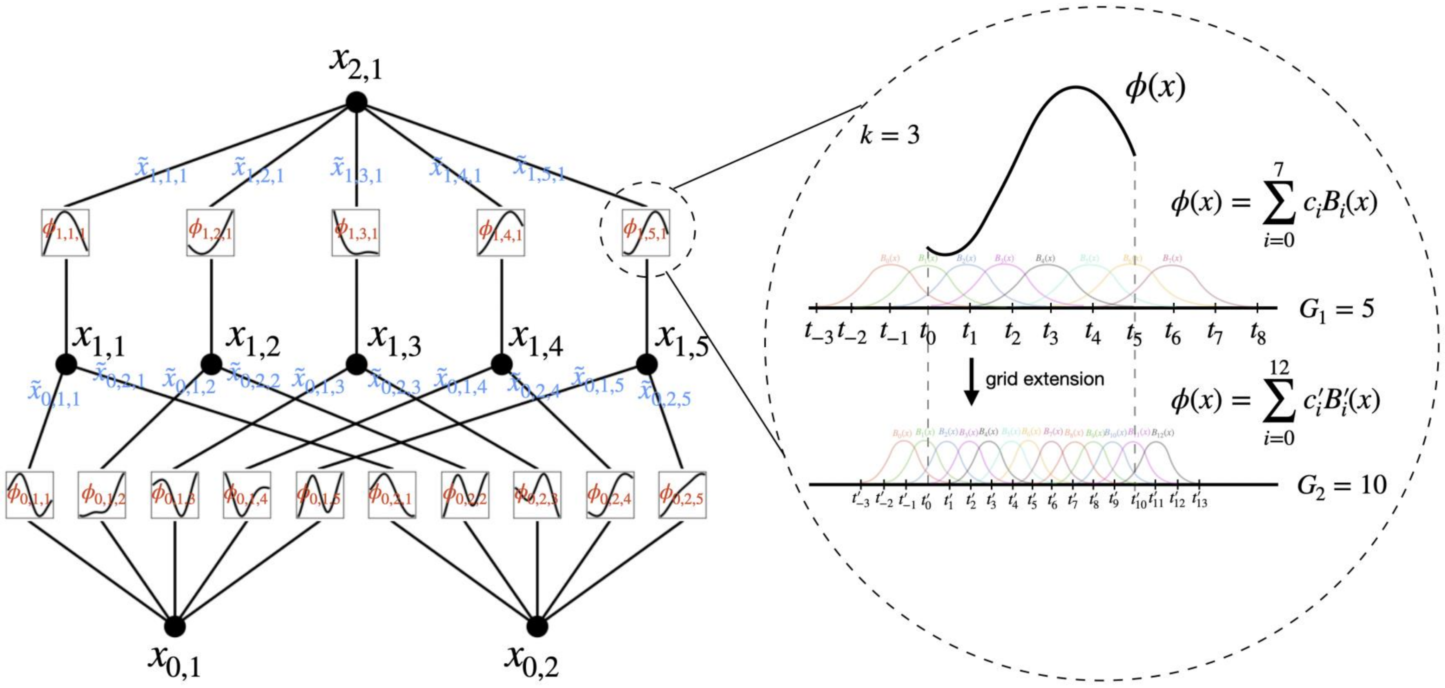
Model	**Multi-Layer Perceptron (MLP)**	**Kolmogorov-Arnold Network (KAN)**
Theorem	**Universal Approximation Theorem**	**Kolmogorov-Arnold Representation Theorem**
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$
Model (Shallow)		

KAN: Kolmogorov-Arnold Networks, Cristian J. Vaca-Rubio, Luis Blanco, Roberto Pereira, Màrius Caus, https://arxiv.org/abs/2404.19756

Активационные функции в KAN

- Параметризуем функции активаций B-сплайнами

- Обучение = подбор коэффициентов при B-сплайнах



Преимущества B-сплайнов:

1. Параметризация вектором чисел
2. Локальность изменения
3. Возможность адаптивно изменять точность сплайна

Детали реализации

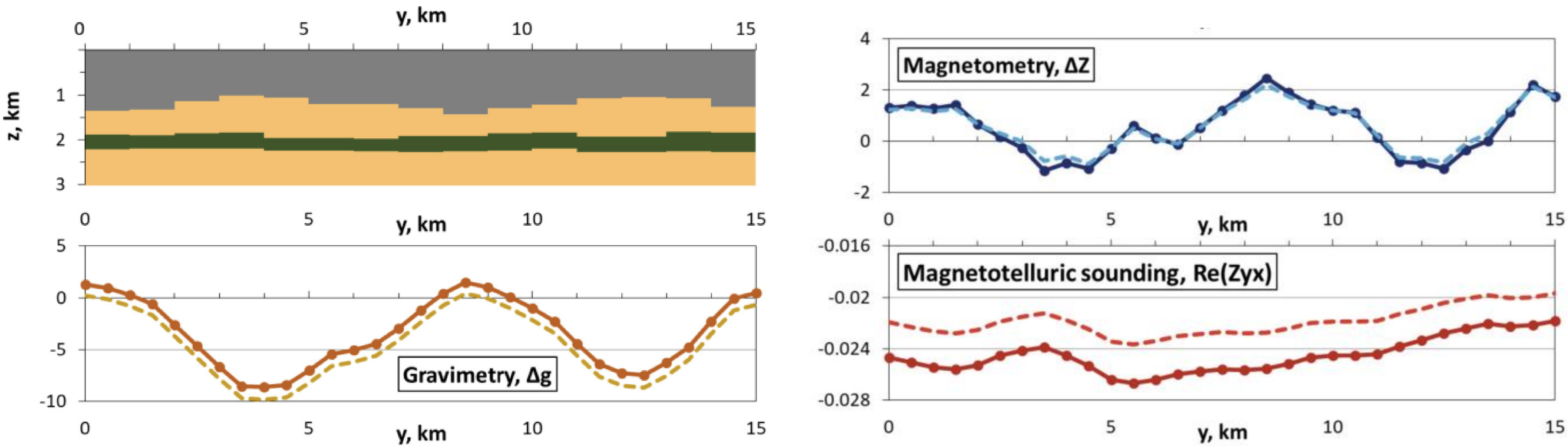
- На самом деле активация выглядит следующим образом (b(x) - аналог residuals connections)

$$\phi(x) = w_b b(x) + w_s \mathrm{spline}(x).$$
$$b(x) = \mathrm{silu}(x) = x/(1+e^{-x})$$

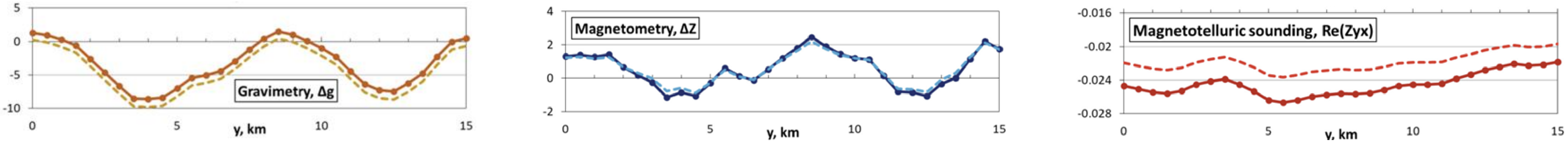
- Сетка, на которой строятся базисные сплайны, может адаптивно меняться в процессе обучения КАН
- Порядок B-сплайнов ~ 3

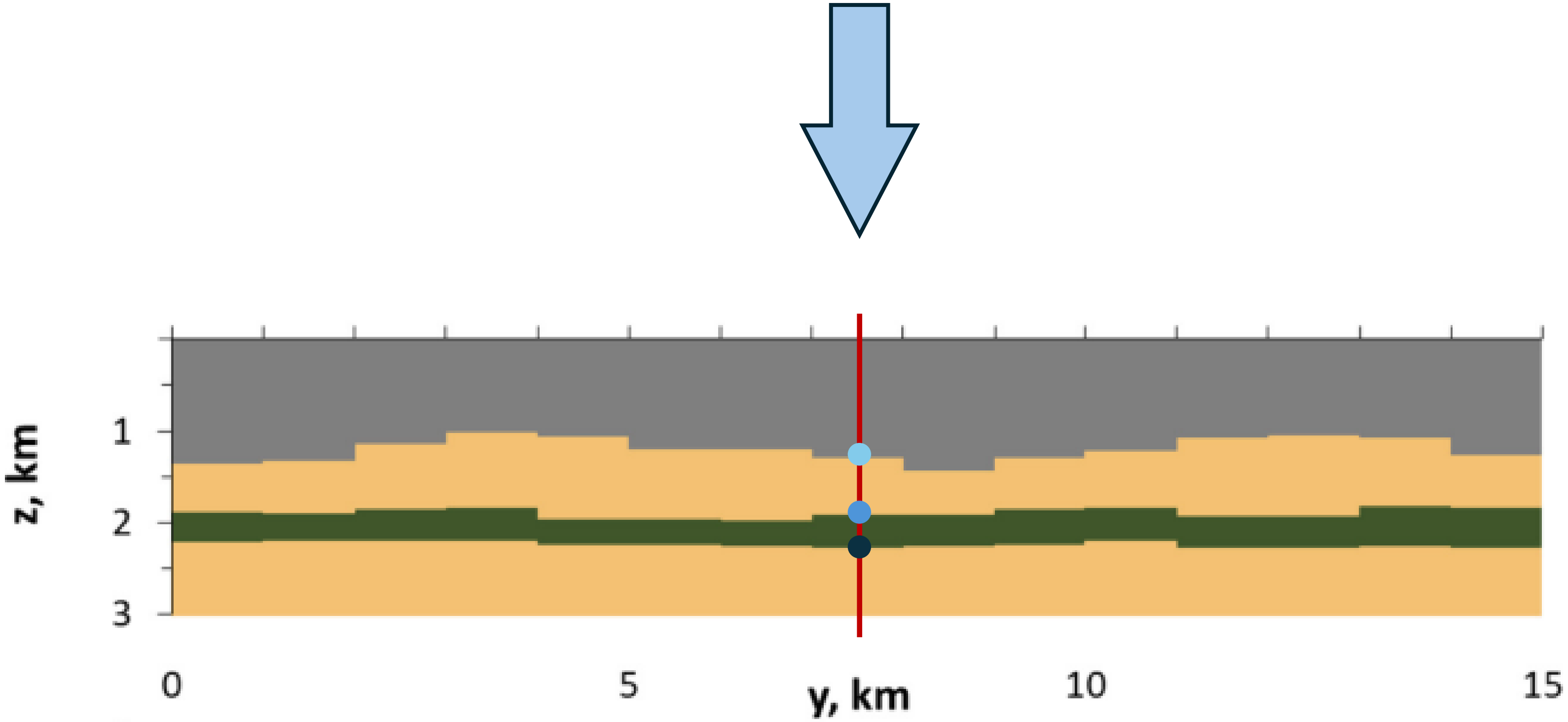
Обратная задача геофизики

Задача: По данным неинвазивных исследований (Гравиметрия, Магнитометрия, Магнитотеллурика) восстановить профиль залегающих пород

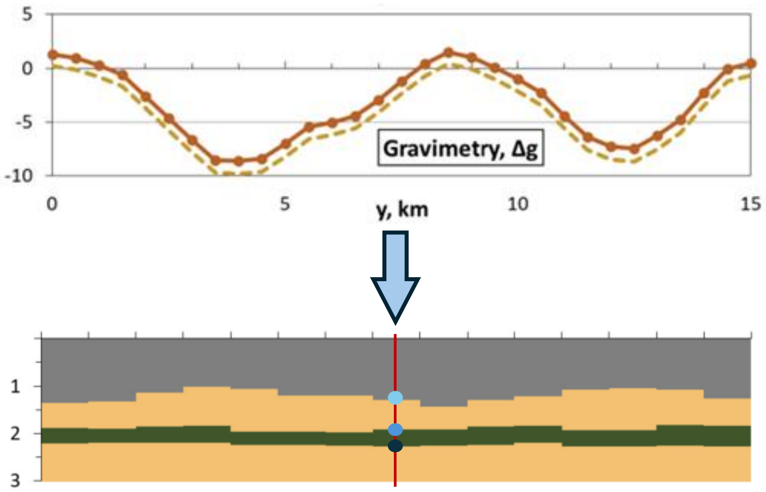


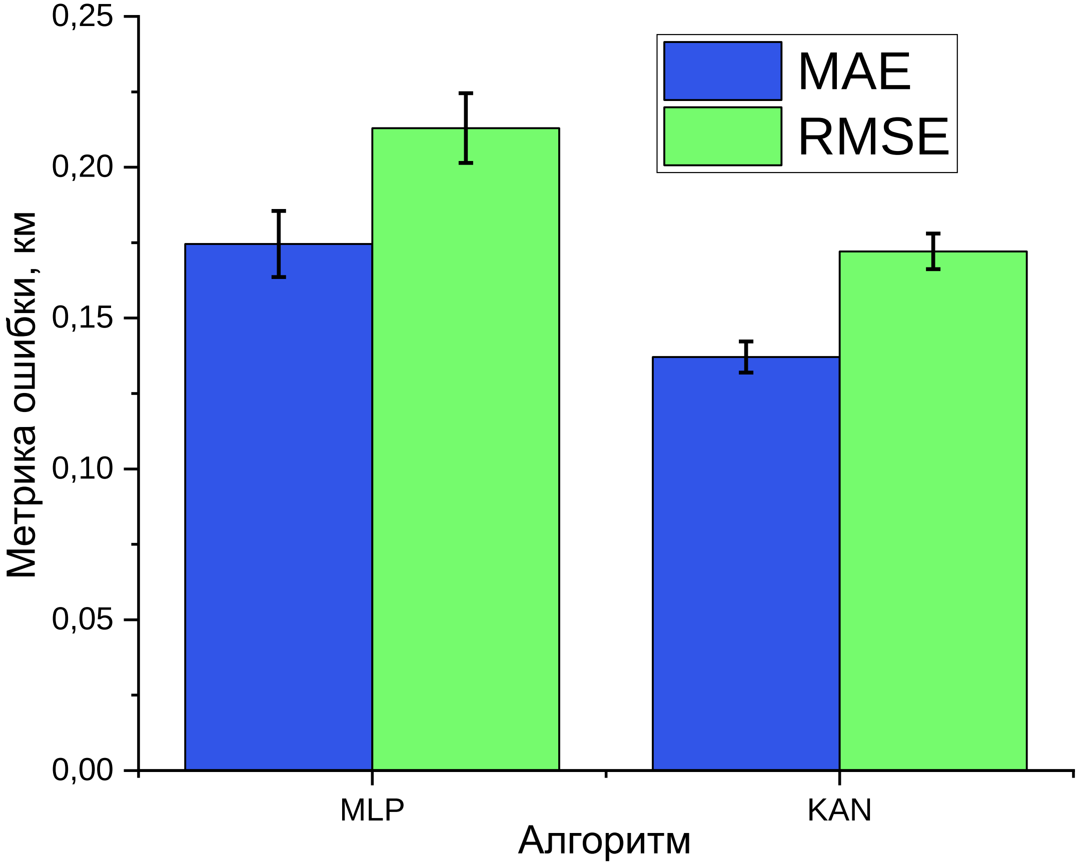
Обратная задача геофизики

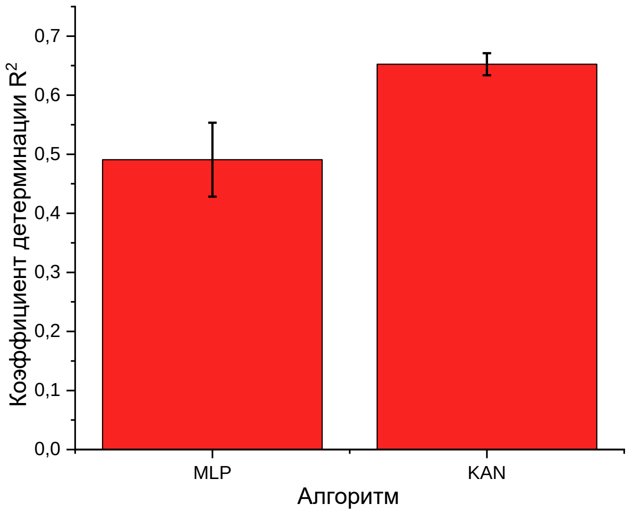




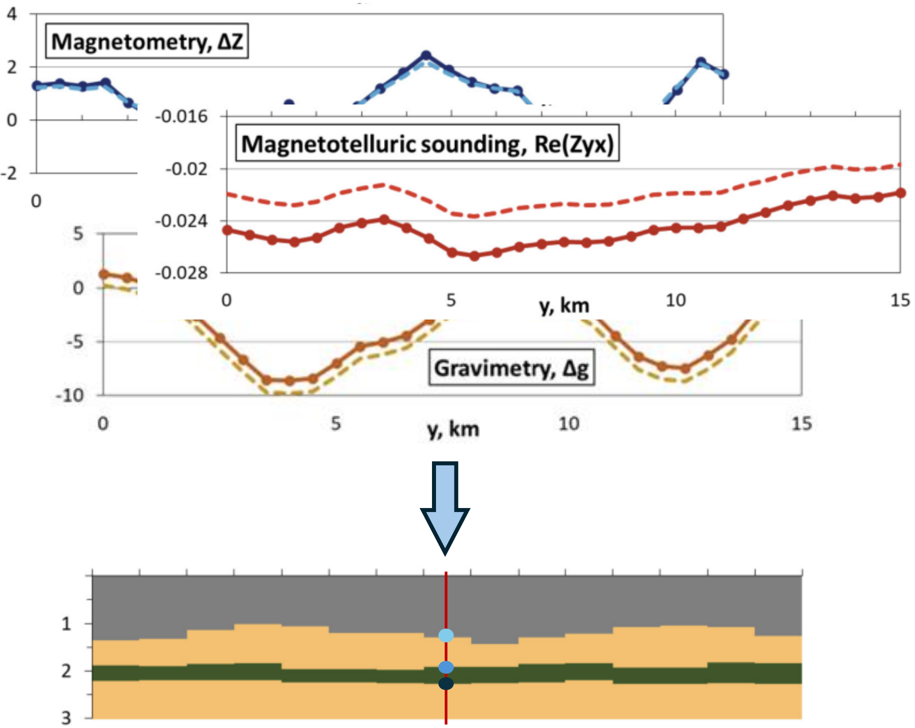
Сравнение персептрона и КАН

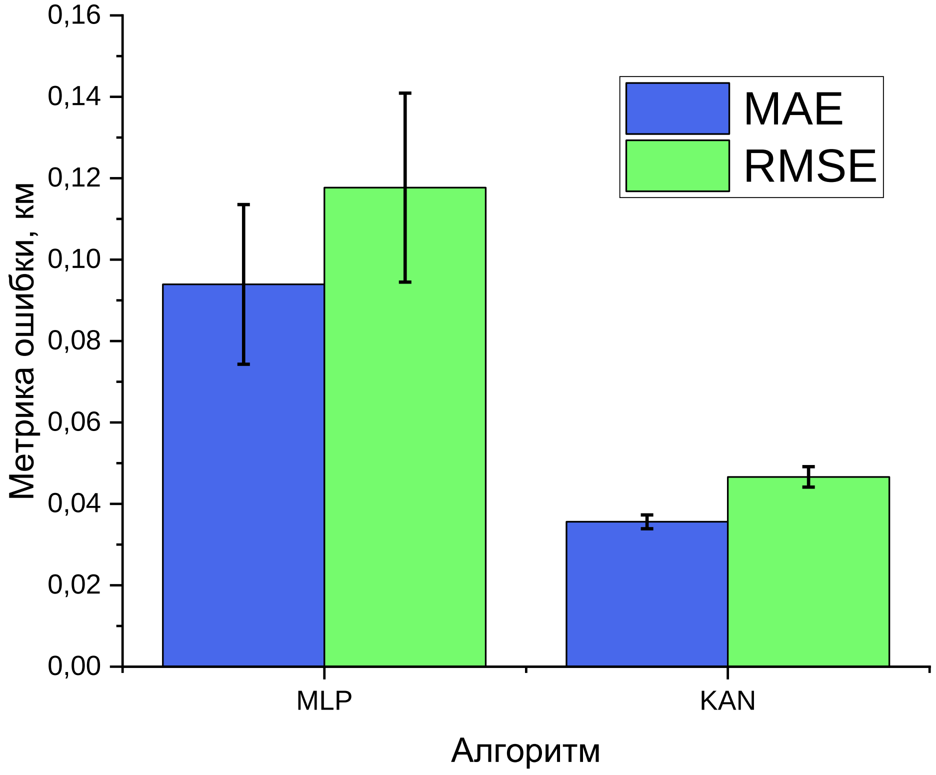


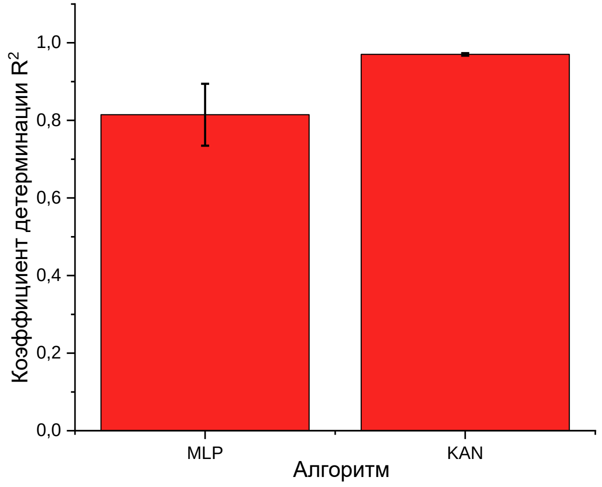




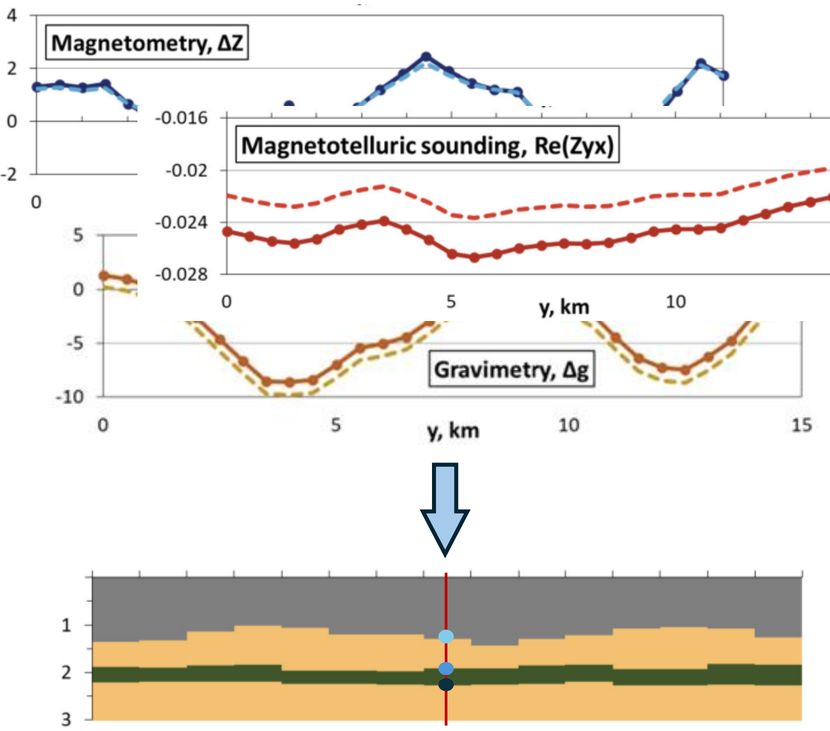
Сравнение персептрона и КАН
Максимальный набор входных данных

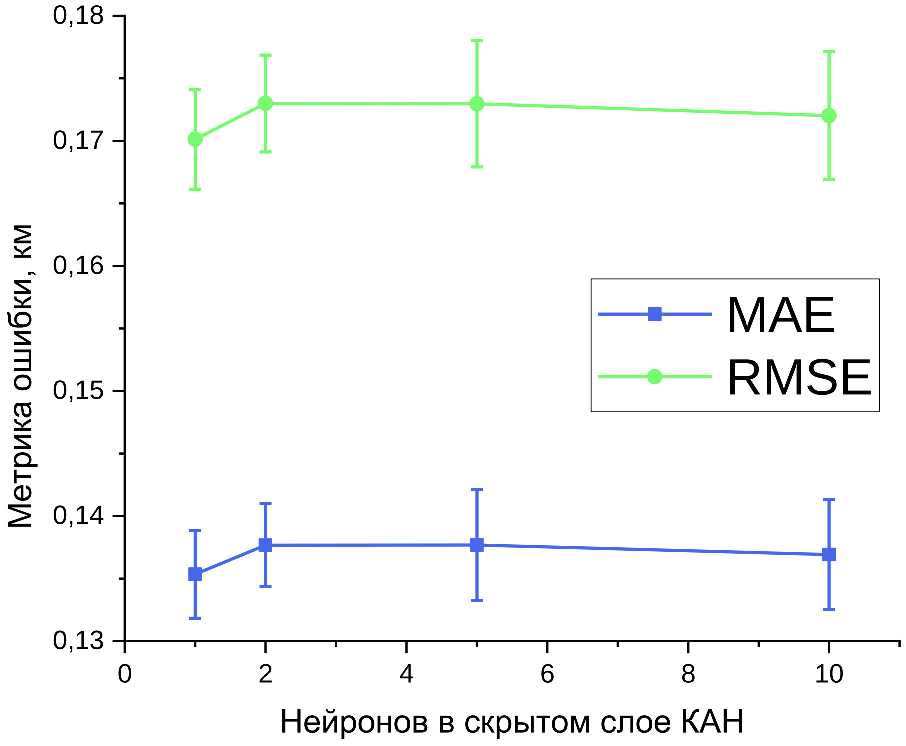


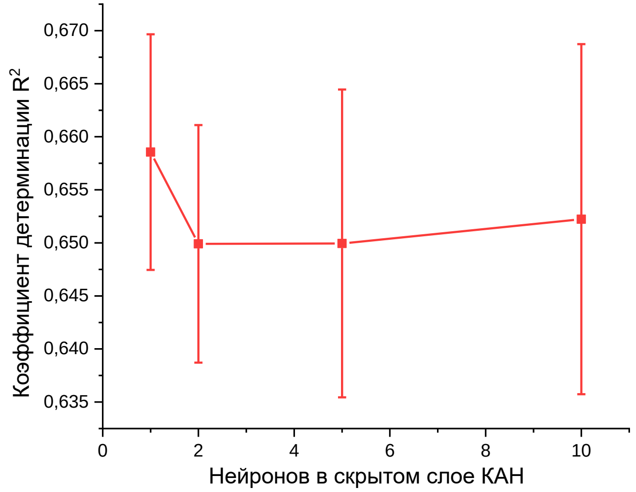




КАН с разным количеством нейронов в скрытом слое



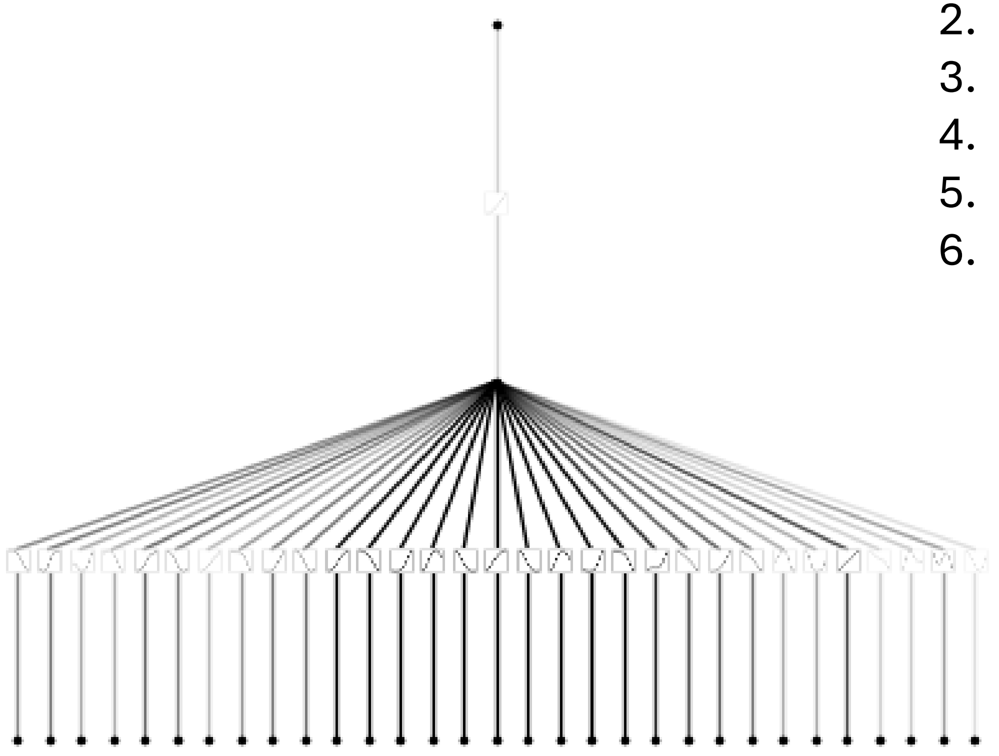




Выводы

1. Сети Колмогорова-Арнольда (КАН) на обратной задаче георазведки превзошли персептрон как при ограниченном, так и при полном наборе входных данных.
2. Оптимальные количество нейронов в скрытом слое у КАН – 1 нейрон.

Спасибо за внимание!



Детали реализации КАН:

1. Используемая библиотека: pykan
2. Порядок сплайнов: K=3
3. Размер сетки сплайнов: grid=3
4. Функция потерь: MSE
5. Оптимизатор: LBFGS
6. Остановка обучения: по callback на валидационном наборе (tol=0.001, iter_no_change=25)