KAN Tutorial Slides

Daniel Precioso Garcelán

October 3, 2025

Kolmogorov-Arnold Representation Theorem

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and let $f : \Omega \to \mathbb{R}$ be a continuous function; i.e. $f \in C(\Omega)$.

Then there exist continuous univariate functions

$$\Phi_q : \mathbb{R} \to \mathbb{R}, \quad q = 1, \ldots, 2d + 1;$$

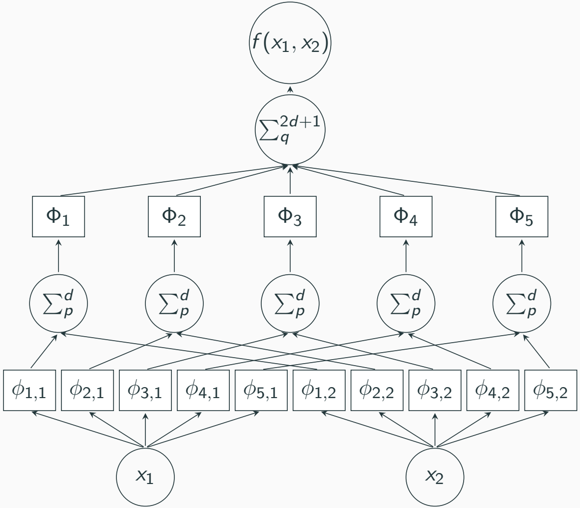
and continuous univariate functions

$$\phi_{pq} : \mathbb{R} \to \mathbb{R}, \quad p = 1, \ldots, d; \quad q = 1, \ldots, 2d + 1;$$

such that for every $\mathbf{x} = (x_1, \ldots, x_d) \in \Omega$,

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

Kolmogorov–Arnold Representation Theorem

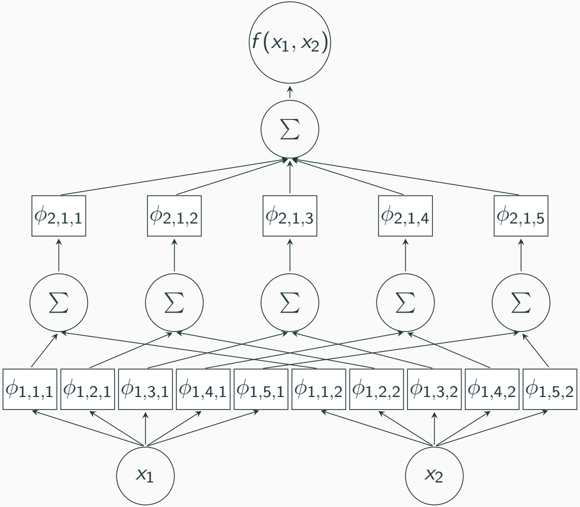


The theorem states that any $f(x_1, x_2)$ can be written as a sum of univariate compositions.

The diagram shows this expression visually: each block represents a component of the decomposition.

Together, they form a **Kolmogorov–Arnold Network (KAN)**.

Kolmogorov–Arnold Networks



In a network setting, each univariate function is written as ϕ_{lpq}, where:

- l: layer depth
- p: output node index
- q: input node index

This network is a **KAN [2,5,1]**: it has 2 inputs, one hidden layer with 5 nodes, and 1 output.

B-Splines

Φ_q and $\phi_{p,q}$ can be chosen from any family of continuous univariate functions. A common choice is the **B-spline** family.

A B-spline of degree k is defined as:

$$B_k(x) = \sum_{i=0}^{n-k-1} P_i N_{i,k}(x)$$

where n is the number of control points (length of the knot vector),
$N_{i,k}$ are the basis functions of degree k,
and P_i are the basis function weights.

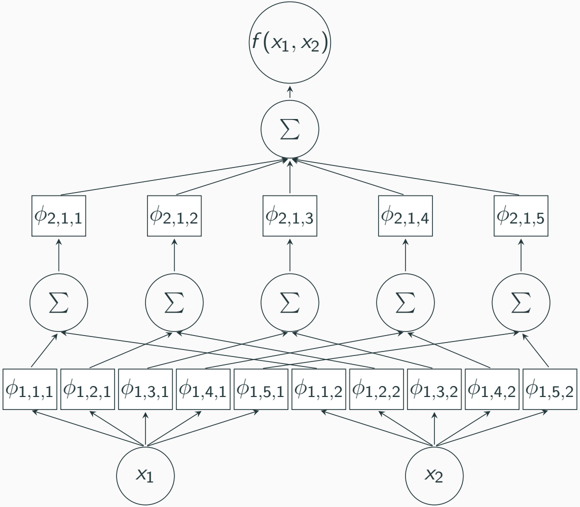
B-Splines

The basis functions follow the standard **Cox–de Boor recursive definition**:

$$N_{i,0}(x) = \begin{cases} 1, & t_i \leq x < t_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

$$N_{i,k}(x) = \frac{x - t_i}{t_{i+k} - t_i} N_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} N_{i+1,k-1}(x), \quad k > 0$$

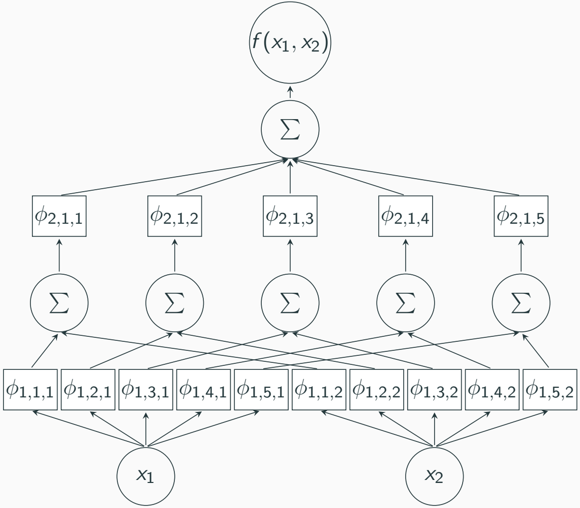
where $t_i \in [t_1, t_n]$ is the **knot vector**, a non-decreasing sequence of real numbers.



We choose the same k-degree and n knot vector length for all B-splines. The univariate functions become:

$$\phi_{pq} = w_{pq} B_{pq}(x) + b_{pq}$$

Kolmogorov–Arnold Networks



Hyperparameters

- n: number of control points.
- k: B-spline degree.

Learnable parameters

(for each edge)

- t_i: knot vectors, $i \in [1, n]$.
- P_i: basis weights, $i \in [1, n - k - 1]$.

Grid Extension

Continuous Learning

B-splines are made of basin functions that operate in small bounds of X. Learning new information in a part of X does not alter other regions of X. In contrary, traditional MLP risk **catastrohpic forgetting**.

Symbolic Regression

Sparsibility