KAN

Kolmogorov–Arnold Networks

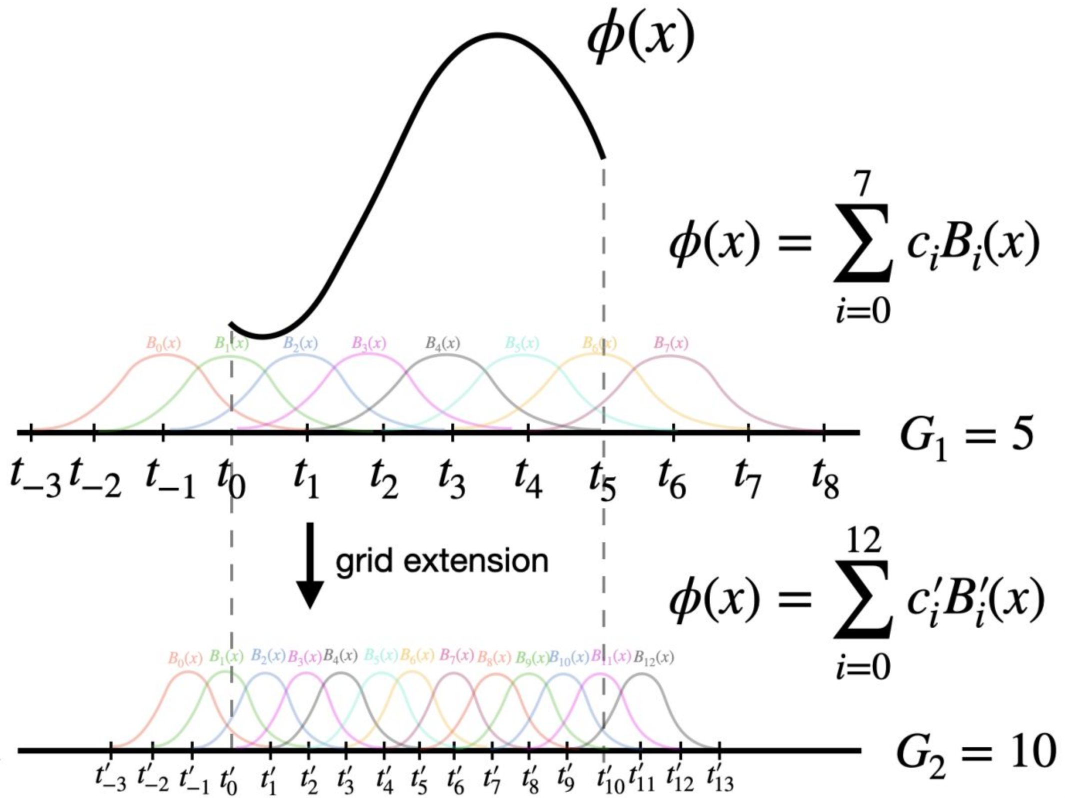
Summary

- Komolgorov-Arnold Theorem
- How B-Splines works?
- A Brief Context
- What is KAN?
- Advantages and Challenges
- Variations of KAN
- Applications
- Conclusion

Komolgorov-Arnold Theorem

*"Kolmogorov-Arnold Representation Theorem establishes that if **f** a **multivariate continuous function** on a bounded domain, then **f** can be written as a **finite composition** of **continuous functions** of a **single variable** and the binary operation of addition."*

Komolgorov-Arnold Theorem



Komolgorov-Arnold Theorem

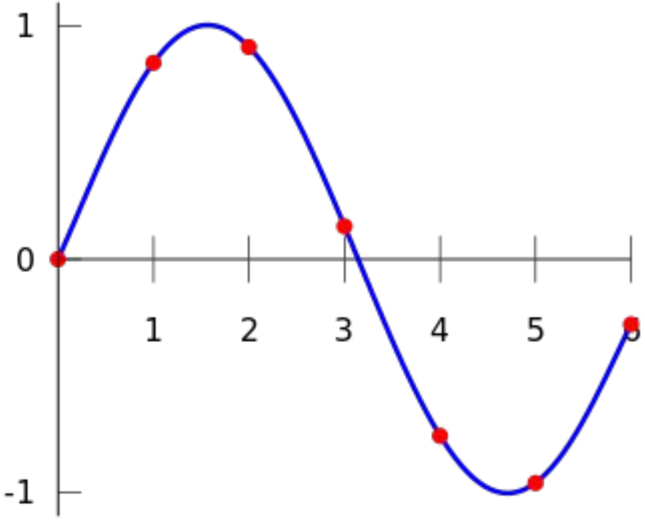
$$f(x_1, \ldots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right)$$

In this expression:

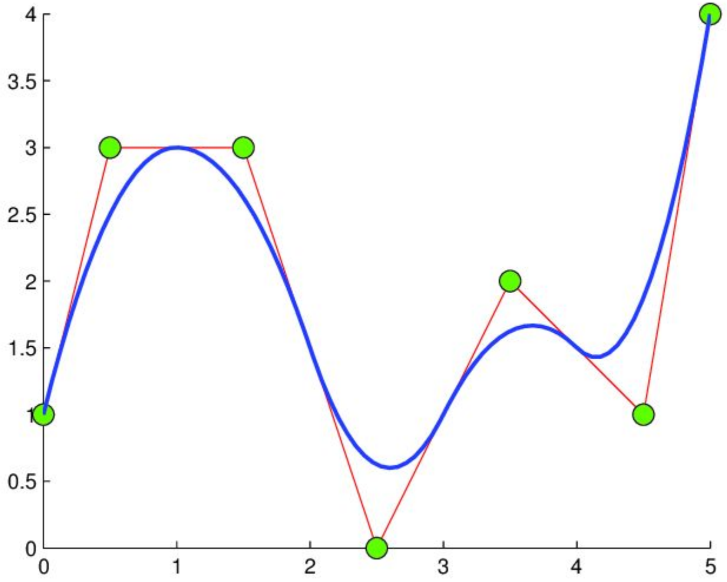
- $f(x_1, \ldots, x_n)$ is the original multivariate continuous function.
- $\phi_{q,p}$ are univariate continuous functions that map each input variable x_p.
- Φ_q are continuous functions that combine the outputs of the univariate functions.

How B-Splines works?

- **Spline:** a **piecewise-defined polynomial** function used in **interpolation**.
- **B-Splines:** a type of spline function used to create **smooth curves**.



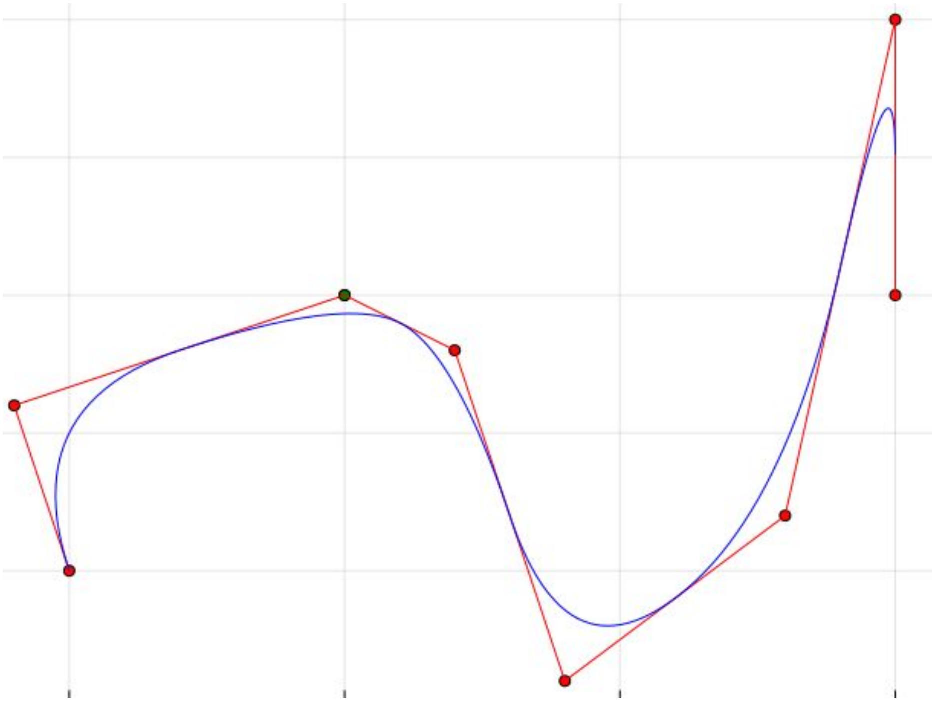
a) Spline



b) B-Spline

How B-Splines works?

- **B-Splines properties:**
    - **Local Control:** The influence of each control point is limited to a specific region of the curve.
    - **Continuity:** Ensure a certain level of continuity at the knots
    - **Stability:** Numerically stable and less prone to oscillations compared to other polynomial interpolation methods.



A Brief Context

- **Most functions** in **science** and **daily life** are **often smooth** and have **sparse compositional** structures.

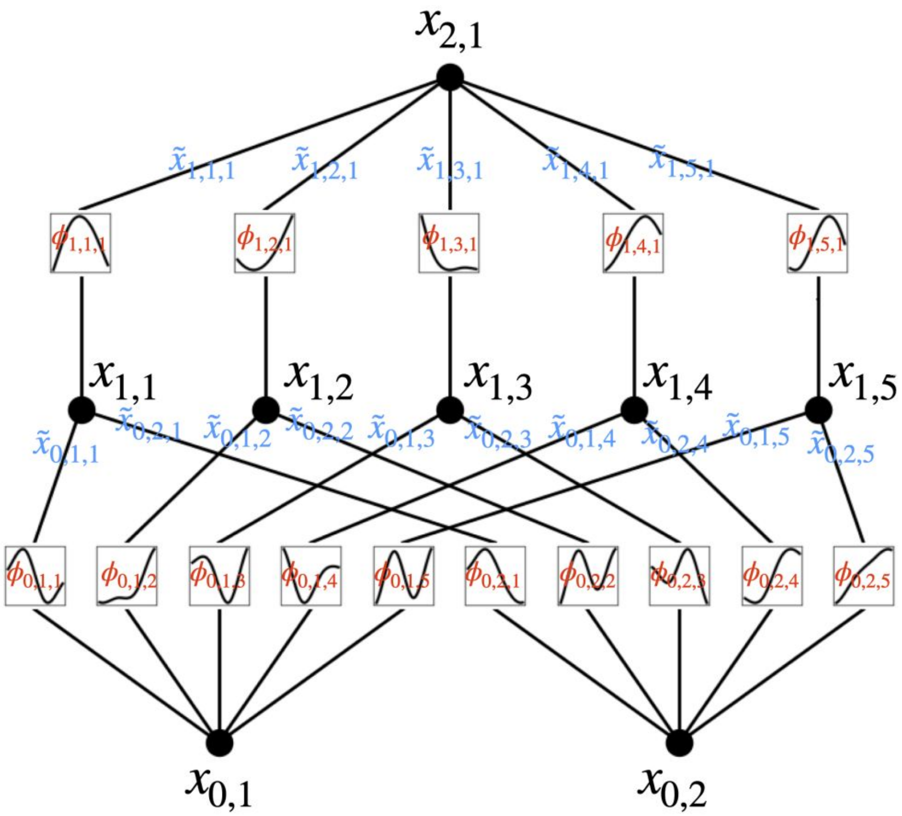
- **MLP** require **many parameters**, being **inefficient** in memory and computation.

- **MLPs** use **predefined activation functions**, **limiting** their ability to **model complex non-linear relationships**.

- **NNs** are considered **"black boxes"**, making it **difficult to understand** the **decision-making process**.

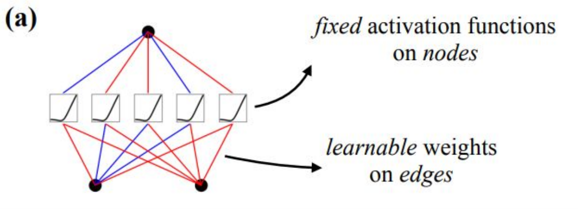
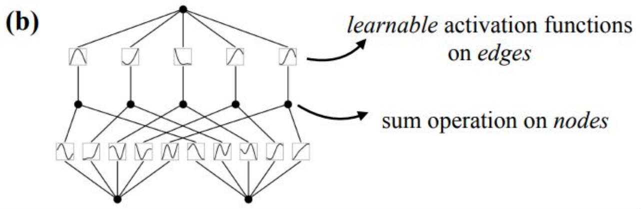
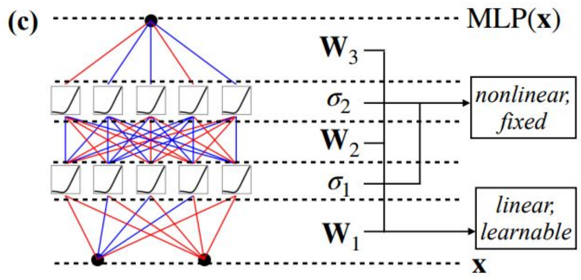
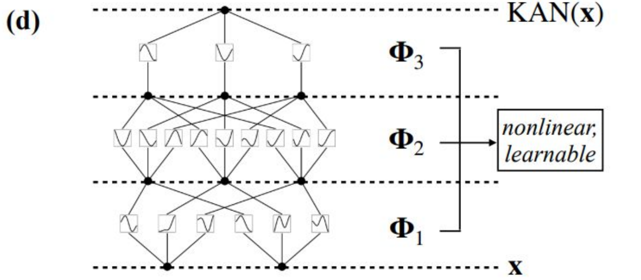
What is KAN?

- Kolmogorov-Arnold Networks (KANs) are a **neural network architecture** that uses **learnable univariate spline** as **activation functions**, to capture complex, non-linear relationships more efficiently and interpretably.



What is KAN?

MLPs treat **linear transformations and nonlinearities separately** as W and σ, while **KANs treat them all together** in Φ

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)		
Formula (Deep)	$\mathrm{MLP}(\mathbf{x}) = (\mathbf{W}_3 \circ \sigma_2 \circ \mathbf{W}_2 \circ \sigma_1 \circ \mathbf{W}_1)(\mathbf{x})$	$\mathrm{KAN}(\mathbf{x}) = (\mathbf{\Phi}_3 \circ \mathbf{\Phi}_2 \circ \mathbf{\Phi}_1)(\mathbf{x})$
Model (Deep)		

What is KAN?

In matrix form, this reads

$$\mathbf{x}_{l+1} = \underbrace{\begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \cdots & \phi_{l,2,n_l}(\cdot) \\ \vdots & \vdots & & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix}}_{\mathbf{\Phi}_l} \mathbf{x}_l,$$

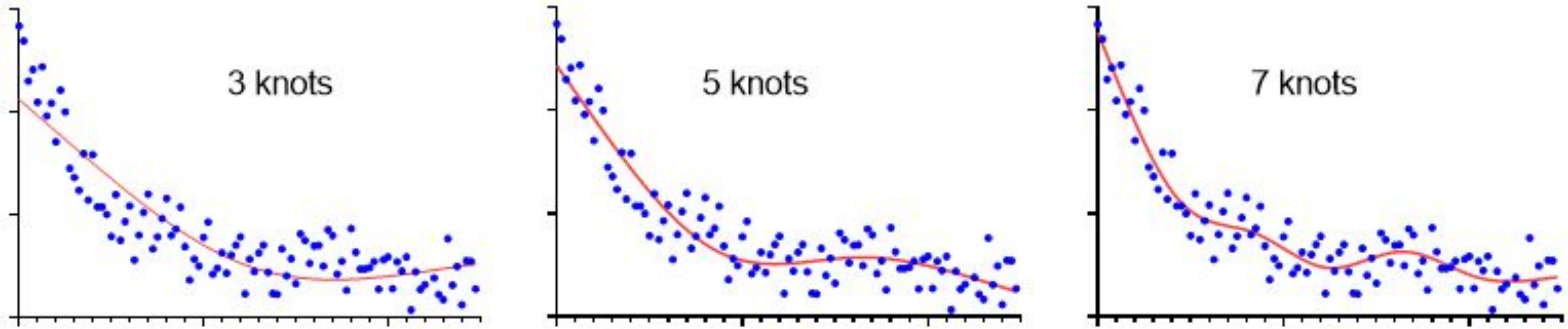
where $\mathbf{\Phi}_l$ is the function matrix corresponding to the l^{th} KAN layer. A general KAN network is a composition of L layers: given an input vector $\mathbf{x}_0 \in \mathbb{R}^{n_0}$, the output of KAN is

$$\text{KAN}(\mathbf{x}) = (\mathbf{\Phi}_{L-1} \circ \mathbf{\Phi}_{L-2} \circ \cdots \circ \mathbf{\Phi}_1 \circ \mathbf{\Phi}_0)\mathbf{x}.$$

What is KAN?

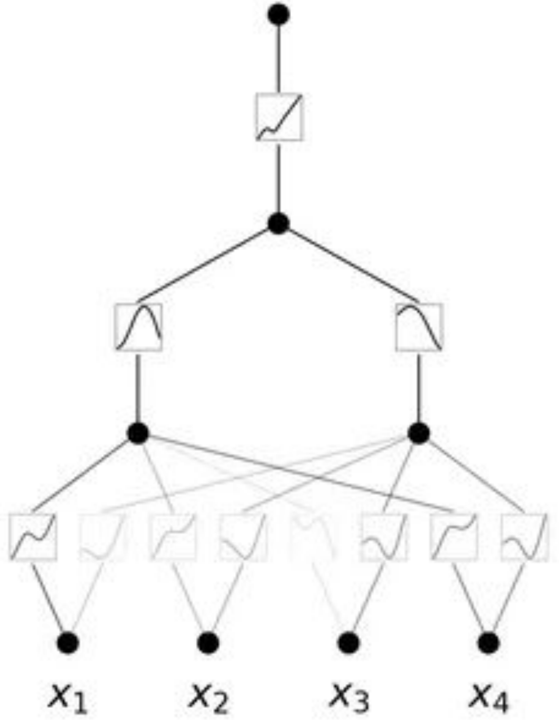
- **KANs Concepts:**
  - **Internal DOF:** Grid points of the Spline (Activation Function).
    - Responsible for learning univariate functions.
  - **External DOF:** The edges of the KAN layer.
    - Responsible for learning compositional structures of multiple variables.

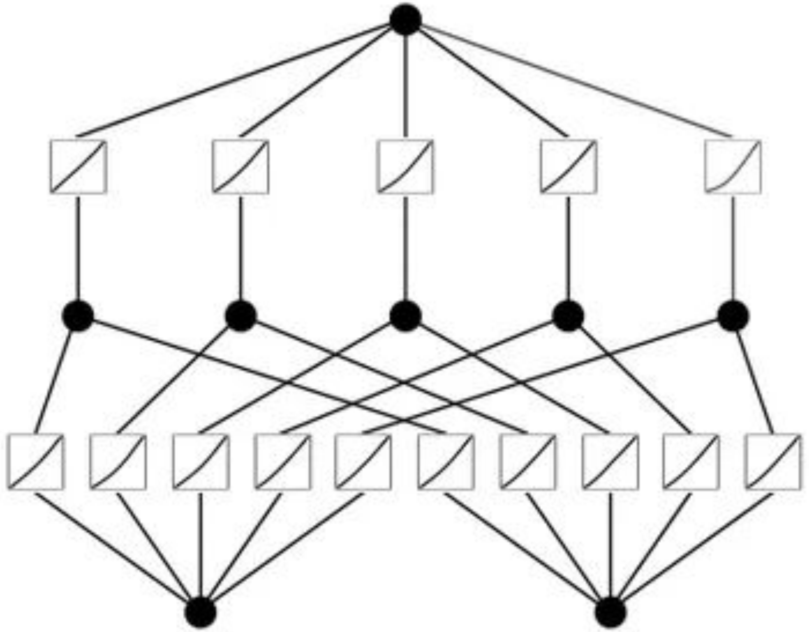
What is KAN?



Internal DOF

What is KAN?



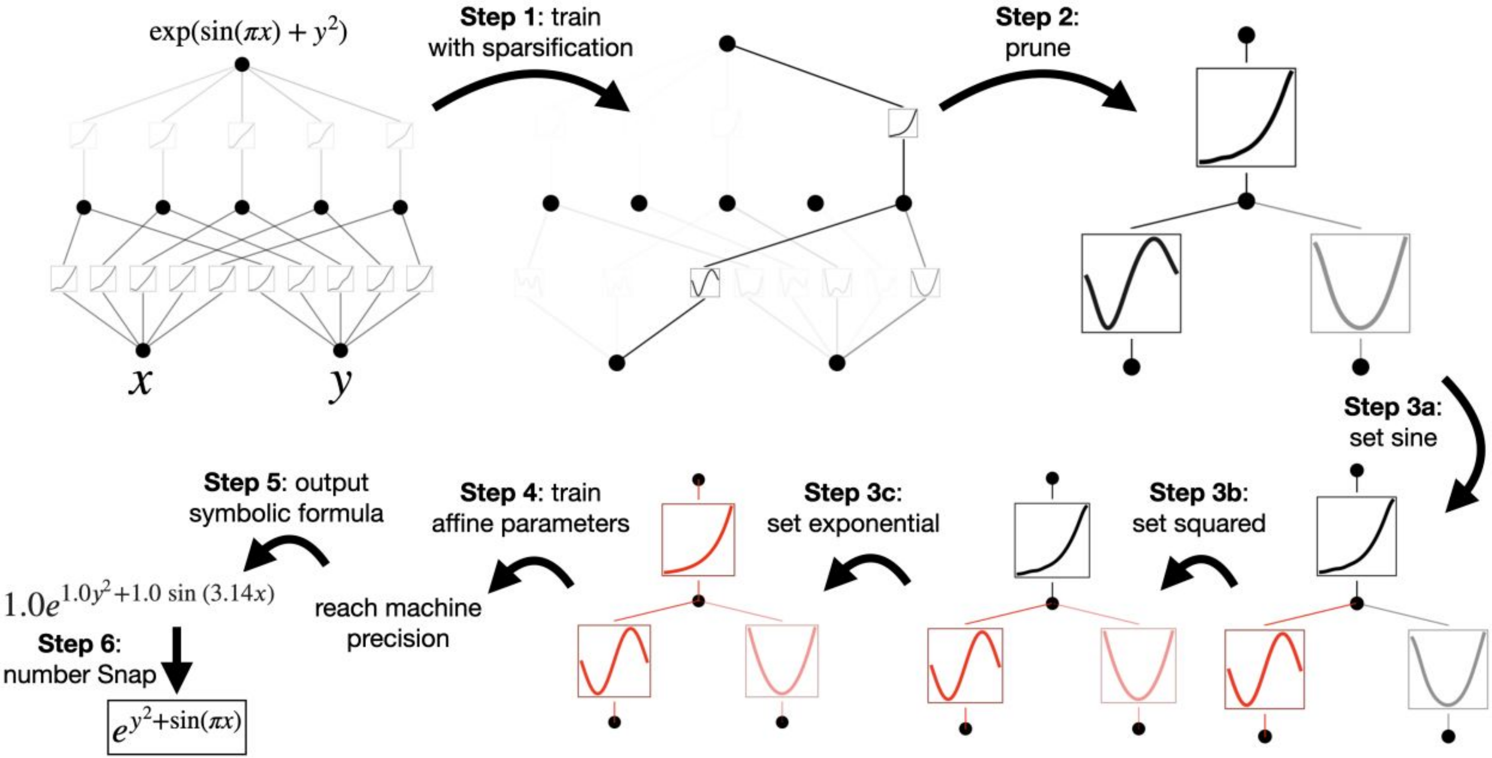


External DOF

What is KAN?

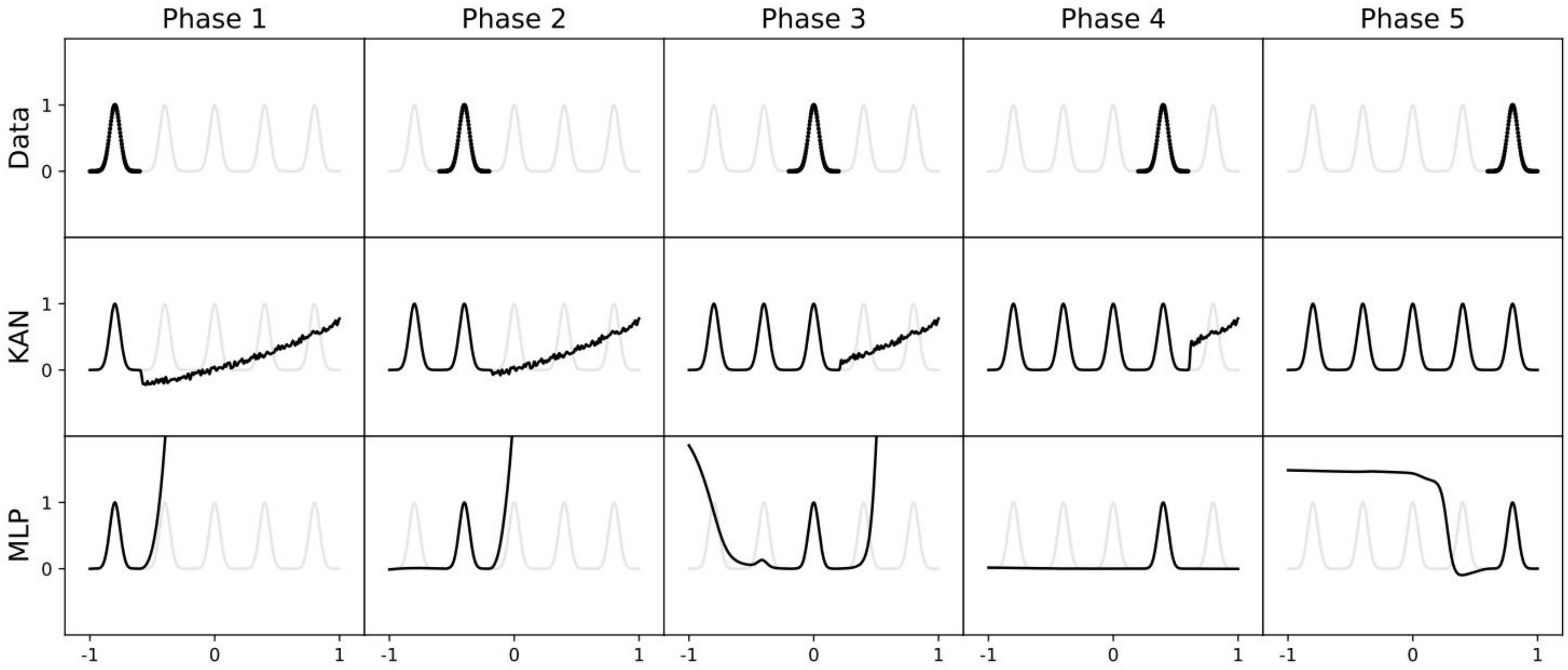
- **KANs Concepts:**
  - **Interpretability: Smaller KANs** can be **interpreted** by analyzing their **activation functions**.
    - **Sparsity** allow prune and reduce KANs.
    - *The way we visualize KANs is like displaying KANs' "brain" to users, and users can perform "surgery" (debugging) on KANs.*
  - **Learning: KANs** can learn both the **compositional structure** and the **univariate functions**.
    - By internal and external DOFs.

What is KAN?



The idea is to start from a large enough KAN and train it with sparsity regularization followed by pruning.

What is KAN?



Catastrophic forgetting is a serious problem in current machine learning. KANs have **local plasticity** and can avoid catastrophic forgetting by leveraging the locality of splines.

Advantages and Challenges

Advantages

- **Interpretability:** KANs are easier to understand and visualize.
- **Efficiency:** KANs achieve comparable or superior results with fewer parameters than MLPs.
- **Precision:** Learnable activation functions enable KANs to model complex data relationships with high accuracy.

Challenges

- **Training Speed:** KANs train slowly because different activation functions cannot leverage batch computation.
- **Implementation Complexity:** Using splines and complex functions makes KANs harder to implement and optimize.
- **Further Research Needed:** More studies are needed to explore the robustness of KANs across diverse datasets and their compatibility with other deep learning architectures.

Variations of KAN

- **U-KAN:** Integrates KANs into the U-Net architecture, commonly used for image segmentation tasks.
- **Diffusion U-KAN:** Applies U-KAN for noise prediction in diffusion models.
- **FourierKAN-GCF:** Combines KANs with Fourier transforms for Graph Collaborative Filtering (GCF).
- **KA2NCD:** Uses KANs in Cognitive Diagnostic Networks (CDNs).
- **iKAN:** Designed for Human Activity Recognition (HAR) using wearable sensor data.
- **T-KAN:** Temporal KAN for univariate time series prediction.

Variations of KAN

- **Wav-KAN:** Integrates wavelets into KANs, replacing traditional splines.
- **MT-KAN:** Multivariate Temporal KAN for modeling interactions between multiple time series.
- **ReLU-KAN:** Uses ReLU activation functions instead of splines.
- **DeepOKAN:** Combines KANs with Deep Operator Networks (DeepONets).
- **Chebyshev Polynomial-Based KAN:** Uses Chebyshev polynomials as an alternative to splines.
- **TKAT (Temporal Kolmogorov-Arnold Transformer):** Integrates KANs with Long Short-Term Memory (LSTM) networks.

Applications

- Time Series Prediction
- Image Segmentation and Generation
- Image Classification
- Human Activity Recognition (HAR)
- Signal Analysis
- Function Approximation

Conclusions

- **KANs** offer **significant benefits** in precision, **interpretability**, and efficiency **compared** to **MLP**.

- They show **potential** across **various applications**, with improved **interpretability** being a **key advantage**.

- However, **KANs** face **challenges** like **slower training speeds** and **implementation complexity**.

- **Future research** is **essential** to fully **unlock** their **potential** and address current **limitations**.