KAN – Kolmogorov-Arnold Networks

Nasy

Jun 07, 2024

Table of Contents

Introduction

Motivation

- Design for interpretable AI for Science, Physics

Kolmogorov-Arnold Representation Theorem (KART)

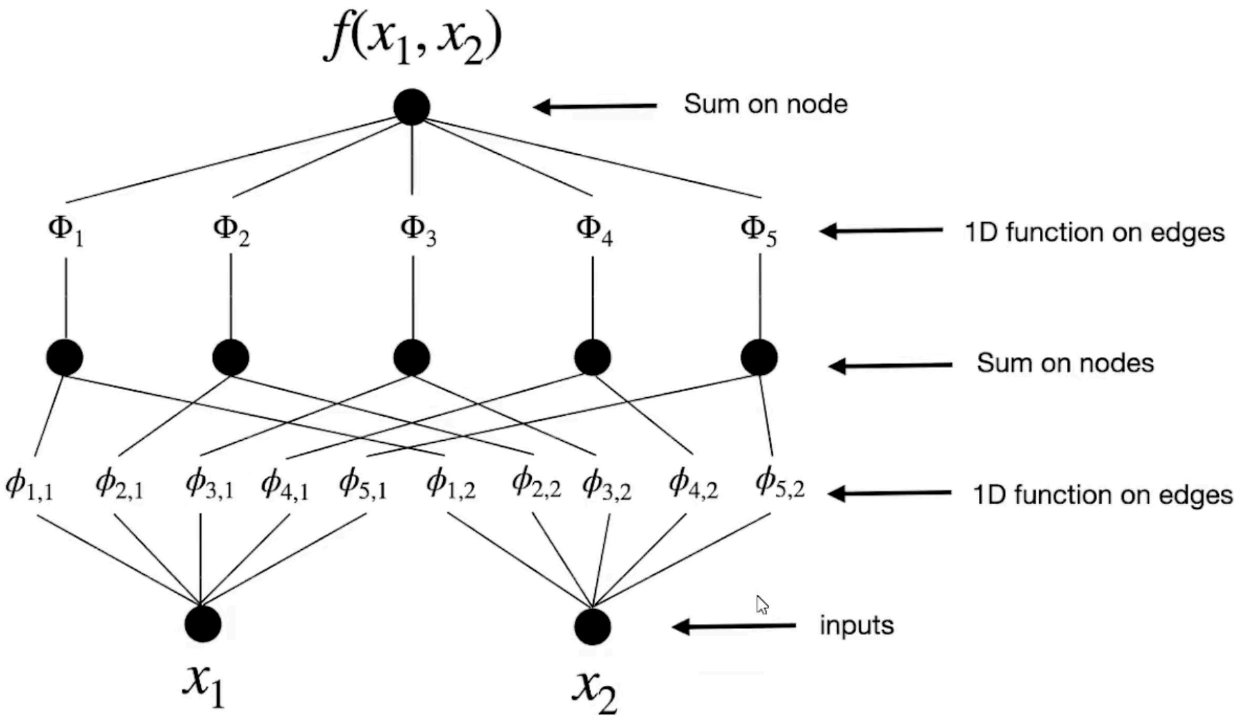
[From wikipedia]

If f is a multivariate continuous function, then f can be written as a finite composition of continuous functions of a single variable and the binary operation of addition.

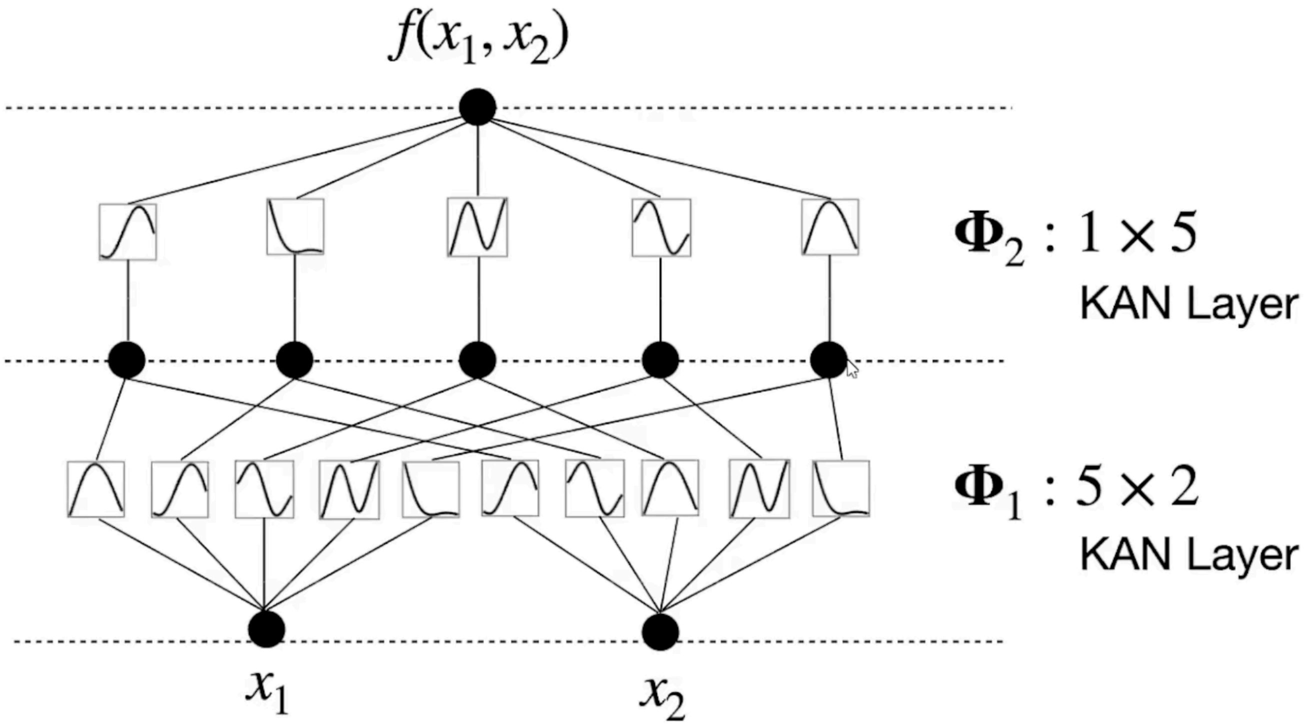
$$f(X) = f(x_1, ..., x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \varphi_{q,p(x_p)} \right) \tag{1}$$

where $\varphi_q, p : [0, 1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$

Intuitive Picture of KART



KART to KAN

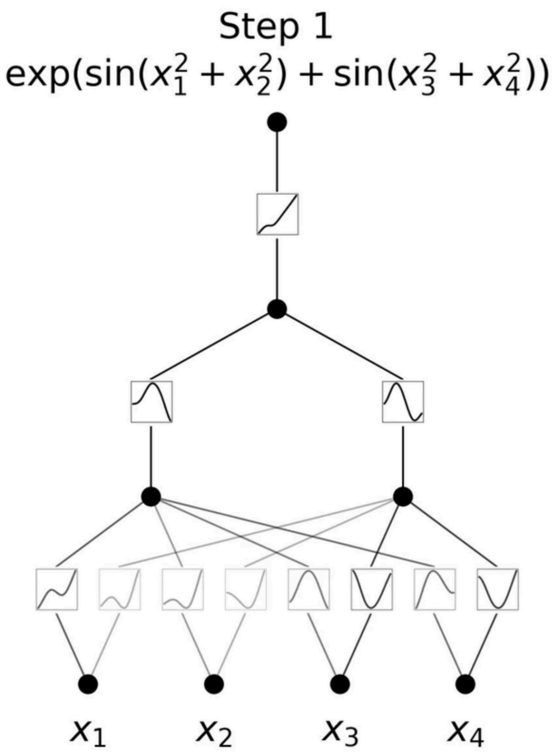


An example of KAN

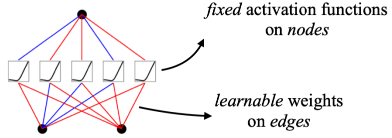
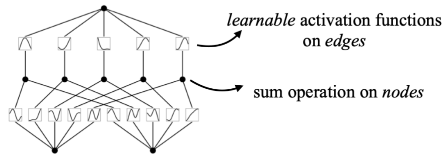
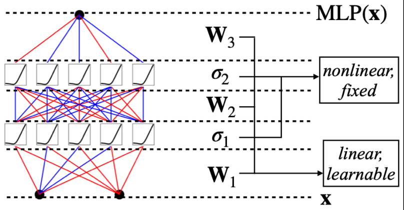
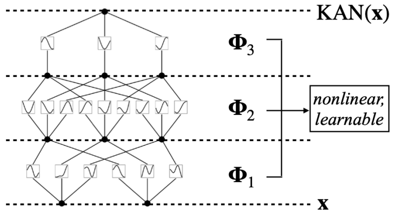
Fit function

$$\exp(\sin(x_1^2 + x_2^2) + \sin(x_3^2 + x_4^2)) \quad (2)$$

Which may need three layers of KAN



MLP vs KAN

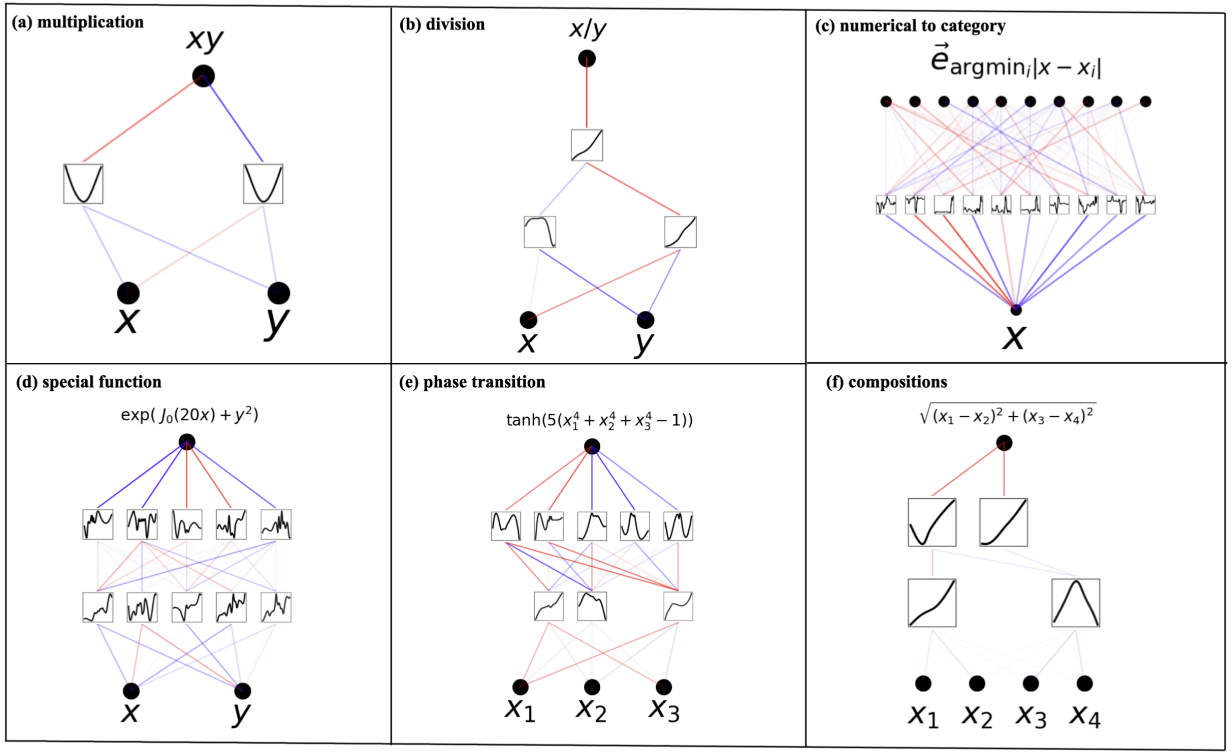
Model	Multi-Layer Perceptron (MLP)	Kolmogorov-Arnold Network (KAN)
Theorem	**Universal Approximation Theorem**	**Kolmogorov-Arnold Representation Theorem**
Formula (Shallow)	$f(\mathbf{x}) \approx \sum_{i=1}^{N(\epsilon)} a_i \sigma(\mathbf{w}_i \cdot \mathbf{x} + b_i)$	$f(\mathbf{x}) = \sum_{q=1}^{2n+1}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

```
MLP = einsum("ij,j->i", w1, sigma(input))

KAN = einsum("ijk,jk->i", w2, phi(input))

input = (d,); w1 = (out, d); w2 = (out, d, 1 + m); phi(input) = (d, 1 + m)
```

Functions Represented by KAN



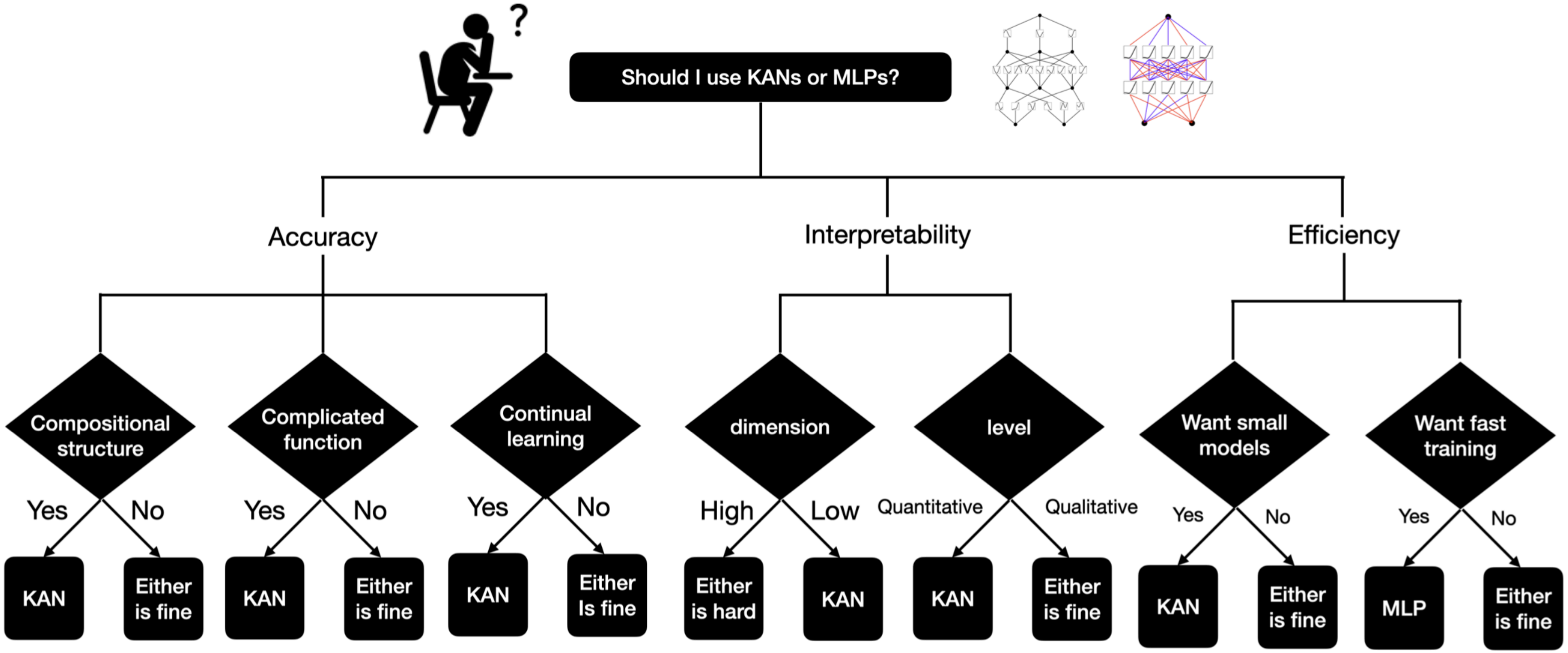
Application

KAN for scientific discoveries

Theory

- Proof and algorithm detail please see the section 2.2 in the paper
- φ is Basic Spline

When to Use It?


?
Should I use KANs or MLPs?
Accuracy
Interpretability
Efficiency
Compositional structure
Complicated function
Continual learning
dimension
level
Want small models
Want fast training
Yes
No
Yes
No
Yes
No
High
Low
Quantitative
Qualitative
Yes
No
Yes
No
KAN
Either is fine
KAN
Either is fine
KAN
Either Is fine
Either is hard
KAN
KAN
Either is fine
KAN
Either is fine
MLP
Either is fine

Conclusion

- Kolmogorov-Arnold Representation Theorem (KART)
- KART to KAN
- MLP vs KAN
- Examples