HybridKAN: Leveraging Multi-Sized Sub-MLPs for Enhanced Performance

UNIVERSITÀ DEGLI STUDI DELL'AQUILA
DISIM
Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica

Fabio D'Andreamatteo, **Andrea D'Angelo**, Giovanni Stilo

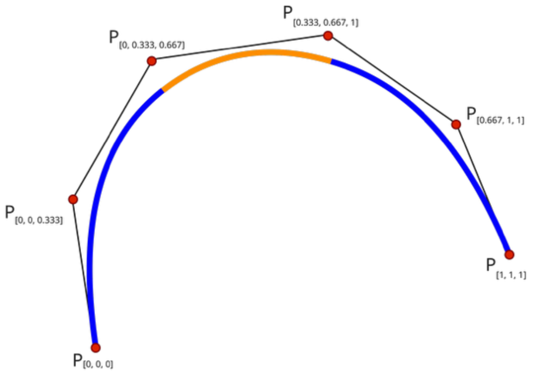
Università degli Studi dell'Aquila / Italy

Kolmogorov-Arnold Networks

- KANs were proposed in 2024 by Liu et al.

- They are an alternative architecture to MLPs built on the Kolmogorov-Arnold Representation Theorem.

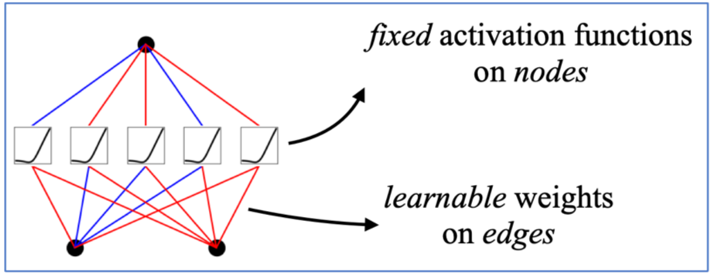
- KANs don't have linear weights: every parameter is replaced by a learnable univariate function known as a *spline*.

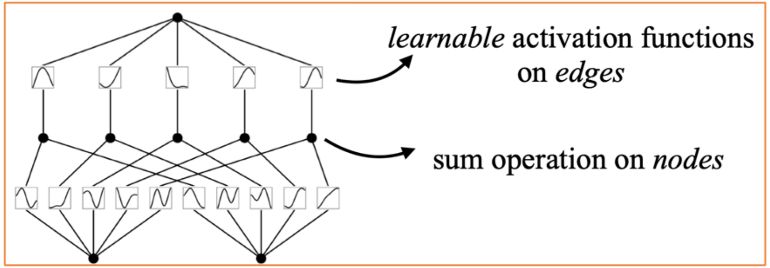


A spline is a univariate function defined piecewise by polinomials.

KANs and MLPs

MLPs	KANs
Built on the Universal Approximation Theorem: “A Neural Network with a certain structure can approximate any continuous function”	Built on the Kolmogorov-Arnold Theorem: “If f is a continuous function, it can be written as a finite addition of univariate continuous functions”



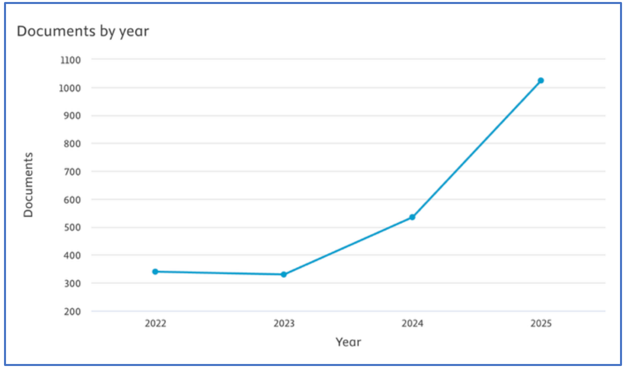


Source: KAN – Kolmogorov-Arnold Networks, Liu et al., 2024

Pros and Cons of KANs

KANs were built as a more explainable and accurate alternative to MLPs.

Research interest on KAN is strongly spiking:

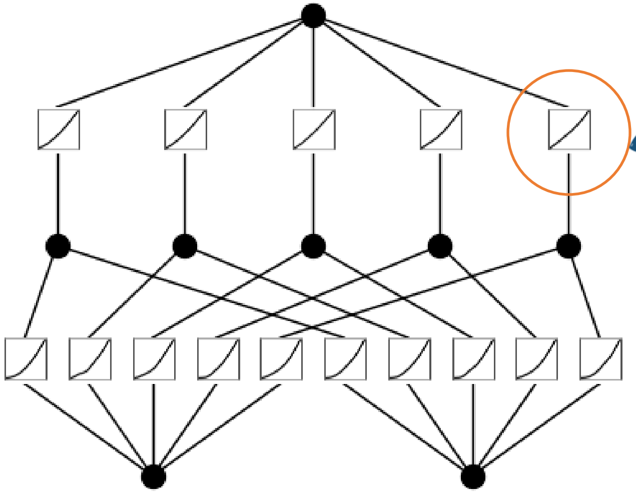


However, KANs also have several drawbacks:

- There are pathological cases where the univariate functions are irregular.
- Training times are significantly longer
- Accuracy is generally worse

Our proposal: Hybrid-KAN (1)

- Our idea stems from the Universal Approximation Theorem behind standard MLPs.
- KAN uses spline functions to approximate the objective function.



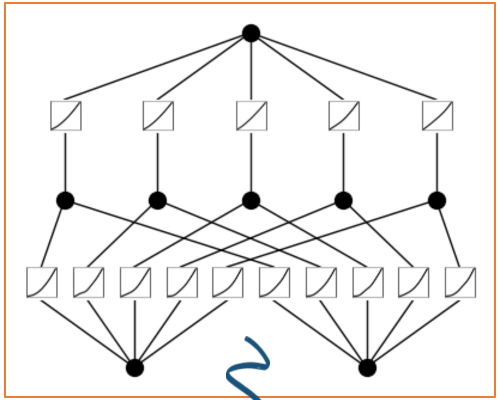
We can approximate the splines via sub-MLPs!

The final architecture is KAN where splines are replaced by sub-MLPs.

Our proposal: Hybrid-KAN (2)

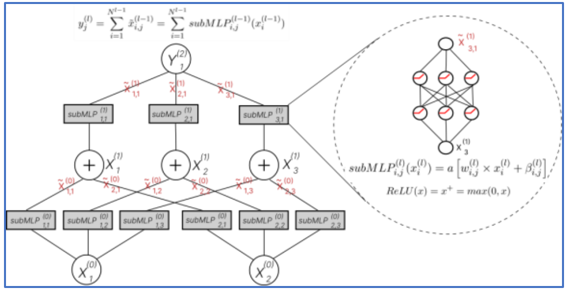
Given a function f that we want to approximate:

1



We use the KAN theorem to approximate it via combination of univariate functions

2

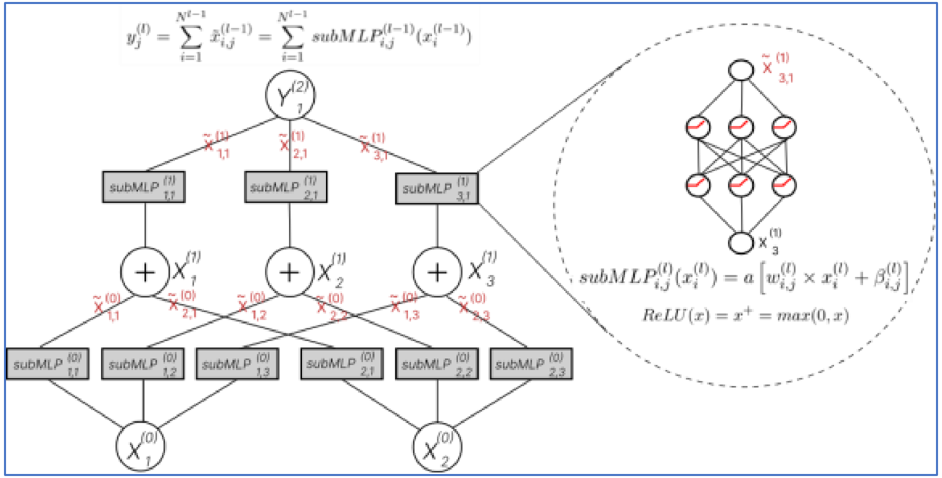


Using the UAT, we replace splines via small (sub-) MLPs.

HybridKAN!

Our proposal: Hybrid-KAN (3)

Hybrid-KAN shifts the underlying theorem back to UAT, by mimicking KAN's architecture and splines.



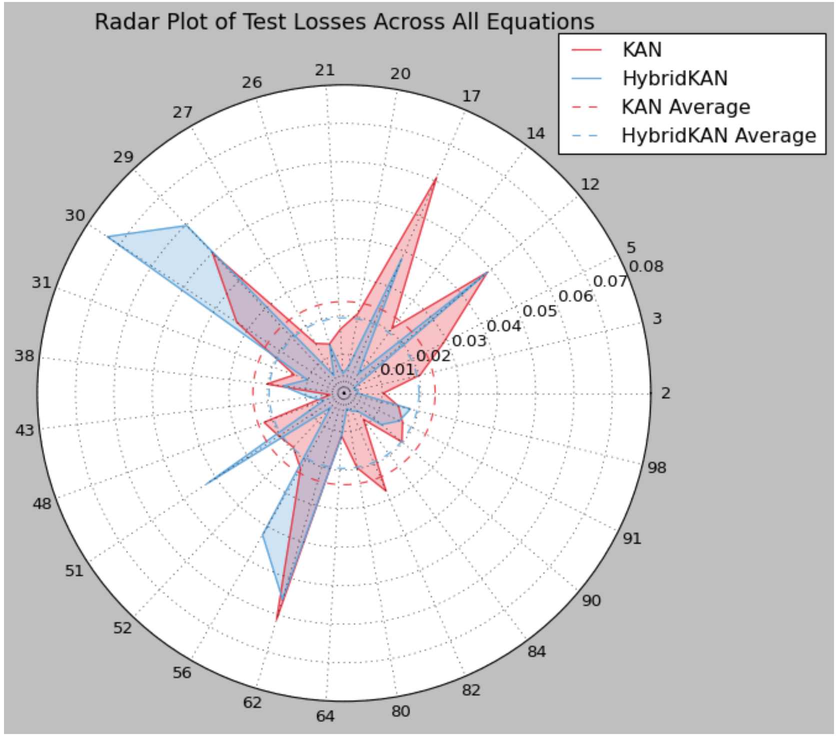
Hybrid-KANs are customizable:

- The number of subMLPs per layer should mimick the number of splines in KANs;
- However, the size and architecture of each subMLP can be customized.
- In our study, we performed a parameter study on the size and shape of subMLPs, selecting the best performing ones.

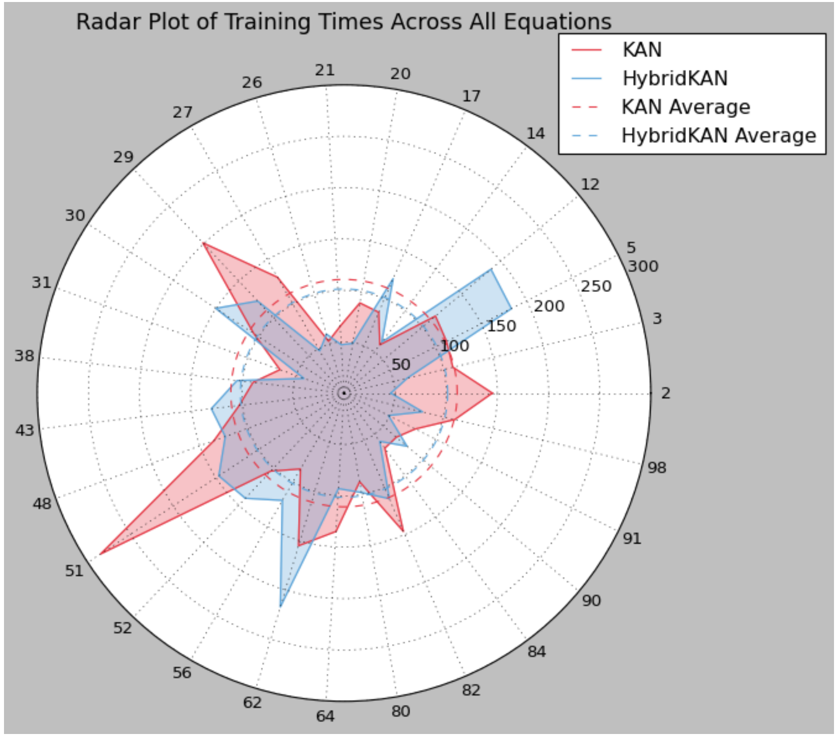
Experimental Setting

- We assess HybridKANs on the same set of equations that were used in the original KAN paper.
- These are 27 equations from the Feynman dataset, with varying numbers of parameters per function.
- We selected settings that are favorable to KAN, as these equations are free of noise and have been used to validate their accuracy.

Results (1)

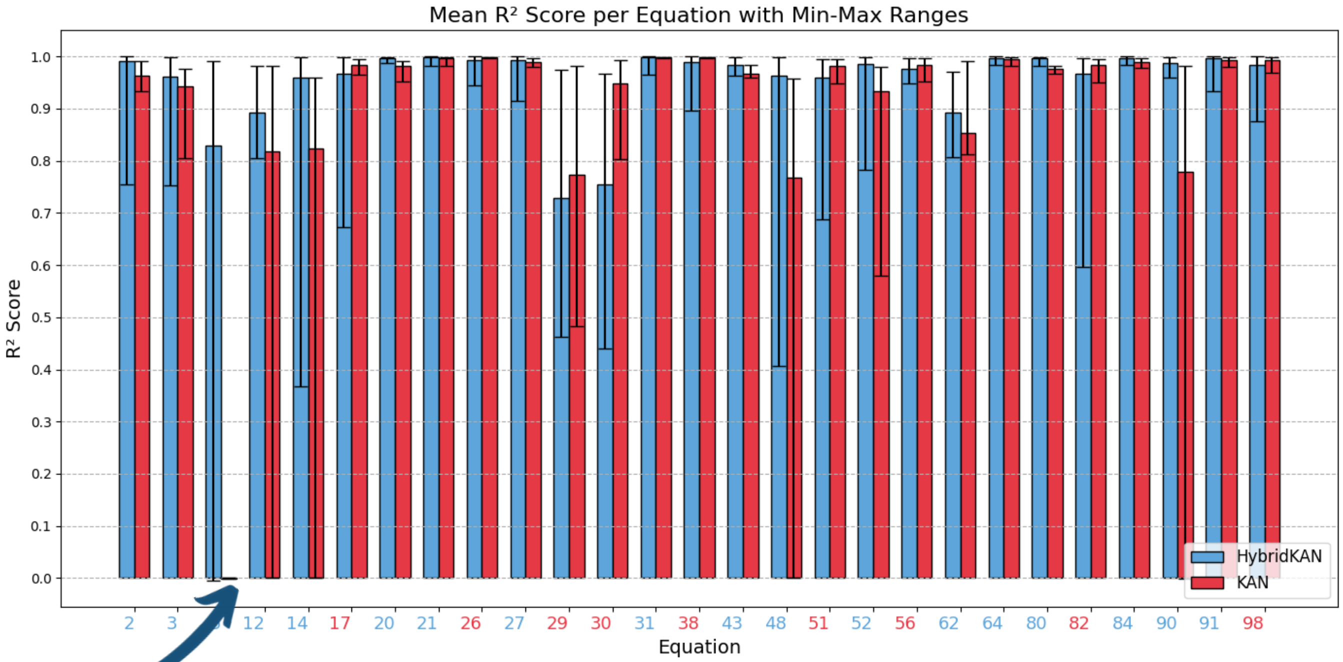


Loss for each equation
(KAN in red, HybridKan in blue)



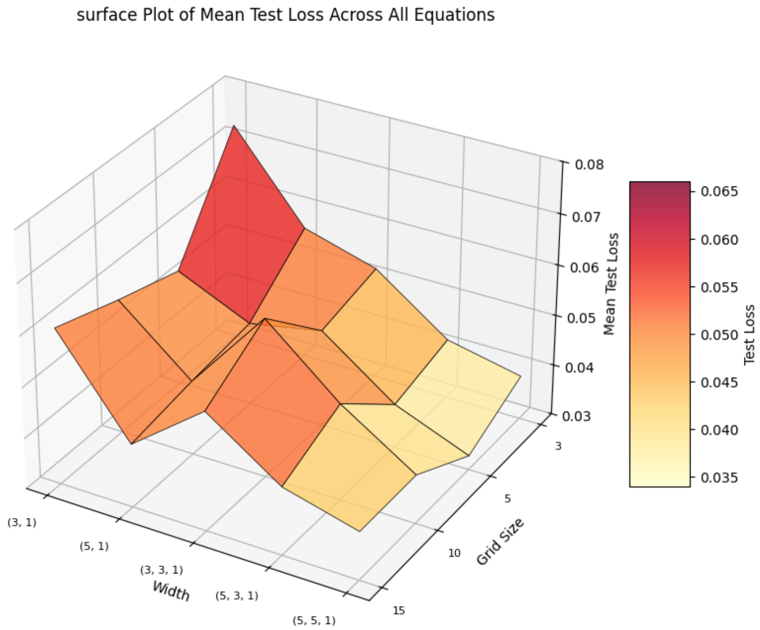
Training time for each equation
(KAN in red, HybridKan in blue)

Results (2)

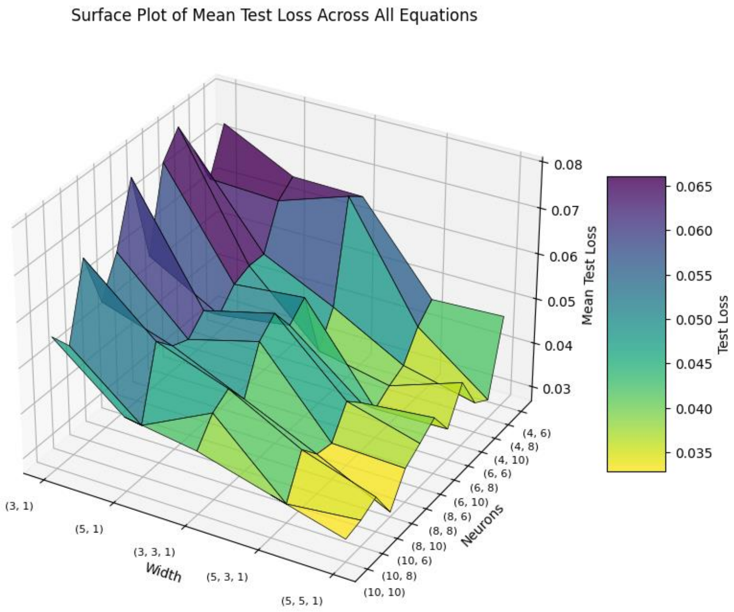


During our experiments, we also found empirical evidence of pathological behavior Of KAN for certain functions, where the univariate functions are irregular and the network collapses.

Results (3)



Loss surface - KAN



Loss surface - HybridKAN

HybridKAN’s loss surface is smoother w.r.t. architecture parameters.

Summing up

- HybridKAN achieves more accurate function approximation on experimental settings that are favorable to KAN;
- HybridKAN takes generally less time to train;
- The loss function of HybridKAN is more predictable w.r.t. to model parameters.

Future Work:

- While promising, HybridKAN needs to be tested on real-world datasets with noise (although we expect an even bigger gap!)
- We need to understand how explainable the subMLPs are w.r.t. splines.

Thank you!

UNIVERSITÀ
DEGLI STUDI
DELL'AQUILA

DISIM
Dipartimento di Ingegneria
e Scienze dell'Informazione
e Matematica