Rec-KAN: Recursive KAN of KANs

Preprint · October 2024

DOI: 10.13140/RG.2.2.30664.12806

Gary Nan Tie

Mu Risk LLC

Recursive KAN: KAN of KANs

Gary Nan Tie, Oct 3, 2024

Abstract
Schematically, the architecture of a Kolmogorov-Arnold neural network (KAN) has sum operations on nodes and learnable activation functions on edges. We introduce Rec-KAN, recursive KANs, by having learnable sparse KANs on edges. Rec-KANs are interpretable, parallelizable, enjoy compositional sparsity, and have sub-quadratic runtime.

Let KAN have shape $[n_0, n_1, \ldots, n_L]$

$$x_{l+1,j} \triangleq \sum_{i=1}^{n_l} \varphi_{l,j,i}(x_{l,i}), \quad \varphi_{l,j,i} : \mathbb{R} \to \mathbb{R} \quad \text{and}$$

$$l = 0, \ldots, L-1 \qquad i = 1, \ldots, n_l \qquad j = 1, \ldots, n_{l+1}$$

in matrix form:

$$x_{l+1} = \Phi_l \, x_l, \quad (n_{l+1} \times n_l) \; \Phi_l = [\varphi_{l,j,i}]$$

Schematically, the KAN architecture has sum operations on the nodes and learnable activation functions φ on the edges. We propose Rec-KAN, recursive KANs, by having learnable sparse KANs on edges.

$$\text{Rec-KAN}(x) = (\Phi_{L-1} \circ \ldots \circ \Phi_0)\, x$$

where $\Phi_l \triangleq \Psi_1^l + \ldots + \Psi_M^l$

with KAN layer $\Psi_k^l = [\psi_{k,j,i}^l]$, $\psi_{k,j,i}^l : \mathbb{R} \to \mathbb{R}$

$$i = 1, \ldots, n_l \qquad j = 1, \ldots, n_{l+1} \qquad k = 1, \ldots, M$$

Without loss of generality, let Rec-KAN have shape

$[n_0, N, \ldots, N, n_L]$ (otherwise let $N = \max\{n_1, \ldots, n_{L-1}\}$).

Note that the intermediate layers $l = 1, \ldots, L-2$

$\Phi_l = \Psi_1^l + \ldots + \Psi_M^l$ have MN^2 activation functions

to learn! To make Rec-KAN tractable we sparsify

the Φ_l by defining $\Psi_k^l = D(f_k^l)\, C(h_k^l)\, D(g_k^l)$

where $D(f)$ is a $N \times N$ diagonal matrix with diagonal

$f = (f_1, \ldots, f_N)$ and $C(h)$ is a $N \times N$ circulant matrix

with generator $h = (h_1, \ldots, h_N)$.

In terms of number of activation functions to learn:

- sparse Rec-KAN Φ_l has $M3N$
- Rec-KAN Φ_l has MN^2
- unstructured KAN Φ_l has N^2

If $N > 3$ and $N > 3M$ then sparse Rec-KAN Φ_L have the fewest activation functions to learn.

Note that $\Phi_L = \Psi_1^L + \ldots + \Psi_M^L$ can be computed in parallel and that $\Psi_K^L(x) = \left(D(f_K^L)\, C(h_K^L)\, D(g_K^L)\right) x$ can be computed using a Fast Fourier Transform.

Case by case, there is a tradeoff between accuracy and runtime (sparsity) to be determined.

Sparse recursive KAN, $(\Phi_{L-1} \circ \ldots \circ \Phi_0)\, x$

with $\Phi_L = \sum_{K=1}^{M} D(f_K^L)\, C(h_K^L)\, D(g_K^L)$ are:

1) interpretable

2) parallelizable

3) compositionally sparse with sub-quadratic runtime.

Layers L, nodes N, and number of DCDs M are hyperparameters to be chosen.

References

KAN:Kolmogorov–Arnold Networks
Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson,
Marin Soljaˇci'c, Thomas Y. Hou, Max Tegmark
arXiv:2404.19756v4 [cs.LG] 16 Jun 2024

Sparse Kolmogorov-Arnold Networks
Gary Nan Tie, Aug 25, 2024
ResearchGate DOI 10.13140/RG.2.2.10535.97448/1