When Might KANs Be Preferable Over DNNs?

The Kolmogorov-Arnold Network (KAN) architecture presents a compelling alternative to conventional deep neural networks (DNNs), especially in scenarios where expressive function approximation is required with relatively shallow or compact architectures. Based on the experimental results, several key takeaways can be identified regarding the conditions under which KANs may be preferable over standard DNNs.

First, KANs offer enchanced representational flexibility through their learable activation functions. Unlike DNNs, which rely on fixed nonlinearities such as ReLU or SiLU, KANs assign a seperate trainable activation function implemented as B-splines. These spline activations are both smooth and differentiable, enabling gradient-based optimizations while offering localized flexibility to better capture complex patterns in data without the need for deeper architectures. The experiments demonstrated that KANs significantly outperformed MLP baselines in terms of mean square error on low-dimentional regression tasks, despite having relatively modest model depth.

Second, KANs may be particularly advantageous in settings where the underlying data distribution contains sharp transitions, irregular patterns, or nested nonlinear relationships. In such cases, the fixed nature of standard activation functions can limit ability of DNNs to adapt, while KANs can tailor their activation behaviour to the data. This was especially evident in the second experiment, where KAN model demonstrated better generilization and lower test error in approximating a highly nonlinear function involving exponential and trigonometric components.

However, these benefits come with trade-offs. KANs generally require more parameters and take higher training times compared to standard DNNs of similar depth, due to large number of learnable activations. In resource constrained environments or latency-sensitive environments, DNNs may remain more practical choice.

While the experimental results clearly highlight the advantages of KANs in modeling complex functions with relatively compact architectures, these findings are based on simplified version of the original KAN architecture. Further investigation is necessary to evaluate their generalization capabilities, scability, and efficiency in real-world, large-scale tasks beyond controlled synthetic settings.

In conclusion, KANs appear most suitable in scenarious where the function to be learned is highly nonlinear and tradicional DNNs struggle to achiwe low approximation error without substantial increases in network depth or capacity. The learnable activation function mechanism allows for more precise function modeling with fewer layers, making them a promising direction for tasks in regression, scientific modeling, or any domain requiring interpretable and flexible function approximation. Neverthless, their high computation cost and implementational complexity should be weighted against performance gains when deciding between KANs and DNNs.

Emirkan Burak Yılmaz