Kolmogorov-Arnold Networks Theoretical Background, current approaches, and potential next steps

Tatiana Boura and Stasinos Konstantopoulos

28 Nov 2024

Abstract

In this talk we will present the Kolmogorov-Arnold Network (KAN), a new and promising approach for machine learning. KAN shifts the objective of training from parameterizing how fixed activation functions contribute to the next layer, to parameterizing activation functions that are then fed to the next layer under simple summation.

We will first discuss the mathematical background on multivariate functions and in particular the Kolmogorov-Arnold Representation Theorem that underlies KANs. We will also discuss works that use the Representation Theorem and subsequent mathematical results as a tool for understanding deep learning and how it differs from the perceptron. We will then focus on recent ML articles and present the original KAN article as well as relevant literature on its extensions and applications. We will close with presenting and discussing our plans in this field, exploring opportunities to collaborate.

References

Jonas Actor. Computation for the Kolmogorov superposition theorem. Master's thesis, Rice University, May 2018.

Vladimir I. Arnold. On functions of three variables (in Russian). *Dokl. Akad. Nauk SSSR*, 114(4), 1957. URL `https://doi.org/10.1007/978-3-642-01742-1_2`. English translation in Givental et al., Vladimir I. Arnold, Collected Works, vol 1. Springer, 2009.

Divesh Basina, Joseph Raj Vishal, Aarya Choudhary, and Bharatesh Chakravarthi. KAT to KANs: A review of Kolmogorov-Arnold Networks and the neural leap forward. arXiv:2411.10622 [cs.LG], November 2024. URL `https://arxiv.org/abs/2411.10622`.

Zavareh Bozorgasl and Hao Chen. Wav-KAN: Wavelet Kolmogorov-Arnold Networks. arXiv:2405.12832 [cs.LG], 2024. URL `https://doi.org/10.48550/arXiv.2405.12832`.

Aaron R. Bragg. On the Kolmogorov-Arnold representation theorem for continuous functions. Master's thesis, Wichita State University, December 2016.

Jürgen Braun and Michael Griebel. On a constructive proof of Kolmogorov's superposition theorem. *Constr Approx*, 30, 2009. URL https://doi.org/10.1007/s00365-009-9054-2.

Carl de Boor. A practical guide to spline. *Applied Mathematical Sciences*, 27, 1978. doi: https://doi.org/10.1109/10.2307/2006241.

Ziqin Feng. *Hilbert's 13th Problem*. PhD thesis, University of Pittsburg, January 2010.

Remi Genet and Hugo Inzirillo. SigKAN: Signature-weighted Kolmogorov-Arnold Networks for time series. arXiv:2406.17890 [cs.LG], 2024a. URL https://doi.org/10.48550/arXiv.2406.17890.

Remi Genet and Hugo Inzirillo. TKAN: Temporal Kolmogorov-Arnold Networks. arXiv:2405.07344 [cs.LG], 2024b. URL https://doi.org/10.48550/arXiv.2405.07344.

Remi Genet and Hugo Inzirillo. A temporal kolmogorov-arnold transformer for time series forecasting. arXiv:2406.02486 [cs.LG], 2024c. URL https://doi.org/10.48550/arXiv.2406.02486.

Federico Girosi and Tomaso Poggio. Representation properties of networks: Kolmogorov's theorem is irrelevant. *Neural Computation*, 1(4), December 1989. URL https://doi.org/10.1162/neco.1989.1.4.465.

Robert Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem. In *IEEE First International Conference on Neural Networks*, 1987.

David Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 8, 1902. URL https://doi.org/10.1090/S0002-9904-1902-00923-3.

Benjamin C. Koenig, Suyong Kim, and Sili Deng. KAN-ODEs: Kolmogorov-Arnold Network ordinary differential equations for learning dynamical systems and hidden physics. arXiv:2407.04192 [cs.LG], 2024. URL https://doi.org/10.48550/arXiv.2407.04192.

Andrey N. Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk SSSR*, 114(5), 1957. URL https://www.mathnet.ru/eng/dan22050.

Věra Kůrková. Kolmogorov's theorem is relevant. *Neural Computation*, 3(4): 617–622, 1991. URL doi:10.1162/neco.1991.3.4.617.

Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993. URL https://doi.org/10.1016/S0893-6080(05)80131-5.

Xing Liu. *Kolmogorov's Superposition Theorem and Its Applications*. PhD thesis, Imperial College of London, September 2015.

Ziming Liu, Pingchuan Ma, Yixuan Wang, Wojciech Matusik, and Max Tegmark. KAN 2.0: Kolmogorov-Arnold Networks meet science. arXiv:2408.10205 [cs.LG], August 2024a. URL https://doi.org/10.48550/arXiv.2408.10205.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov-Arnold Networks. arXiv:2404.19756 [cs.LG], June 2024b. URL https://arxiv.org/abs/2404.19756.

Kaihuai Qin. General matrix representations for b-splines. In *Proceedings Pacific Graphics '98. Sixth Pacific Conference on Computer Graphics and Applications (Cat. No.98EX208)*, pages 37–43, 1998. doi: https://doi.org/10.1109/PCCGA.1998.731996.

Johannes Schmidt-Hieber. The Kolmogorov-Arnold representation theorem revisited. *Neural Networks*, 137, 2021. URL https://doi.org/10.1016/j.neunet.2021.01.020.

David A. Sprecher. On the structure of continuous functions of several variables. *Transactions of the American Mathematical Society*, 115, 1965. URL https://doi.org/10.1090/S0002-9947-1965-0210852-X.

David A. Sprecher. An improvement in the superposition theorem of Kolmogorov. *Journal of Mathematical Analysis and Applications*, 38(1), April 1972. URL https://doi.org/10.1016/0022-247X(72)90129-1.

Cristian J. Vaca-Rubio, Luis Blanco, Roberto Pereira, and Màrius Caus. Kolmogorov-Arnold Networks (KANs) for time series analysis. arXiv:2405.08790 [eess.SP], 2024. URL https://doi.org/10.48550/arXiv.2405.08790.

Anatoliy Georgievich Vitushkin. On Hilbert's thirteenth problem and related questions. *Russian Mathematical Surveys*, 59(1), 2004. URL https://doi.org/10.1070/RM2004v059n01ABEH000698.