Application of Kolmogorov-Arnold Networks in high energy physics

ABASOV E., BOOS E., BUNICHEV V., VOLKOV P., VOROTNIKOV G., DUDKO L., ZABORENKO A., IUDIN E., MARKINA A., PERFILOV M.

Kolmogorov-Arnold theorem

- V. I. Arnol’d, “On functions of three variables”, Dokl. Akad. Nauk SSSR, 114 (1957), pp. 679–681
- V. I. Arnol’d, “On the representation of continuous functions of three variables by superpositions of continuous functions of two variables”, Mat. Sb. (N.S.), 48(90):1 (1959), 3–74
- Kolmogorov, A. N. “On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables”, Dokl. Akad. Nauk SSSR, 108 (1956), 179–182;
- Kolmogorov, A. N. “On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition.” Dokl. Akad. Nauk SSSR 114 (1957), 953–956

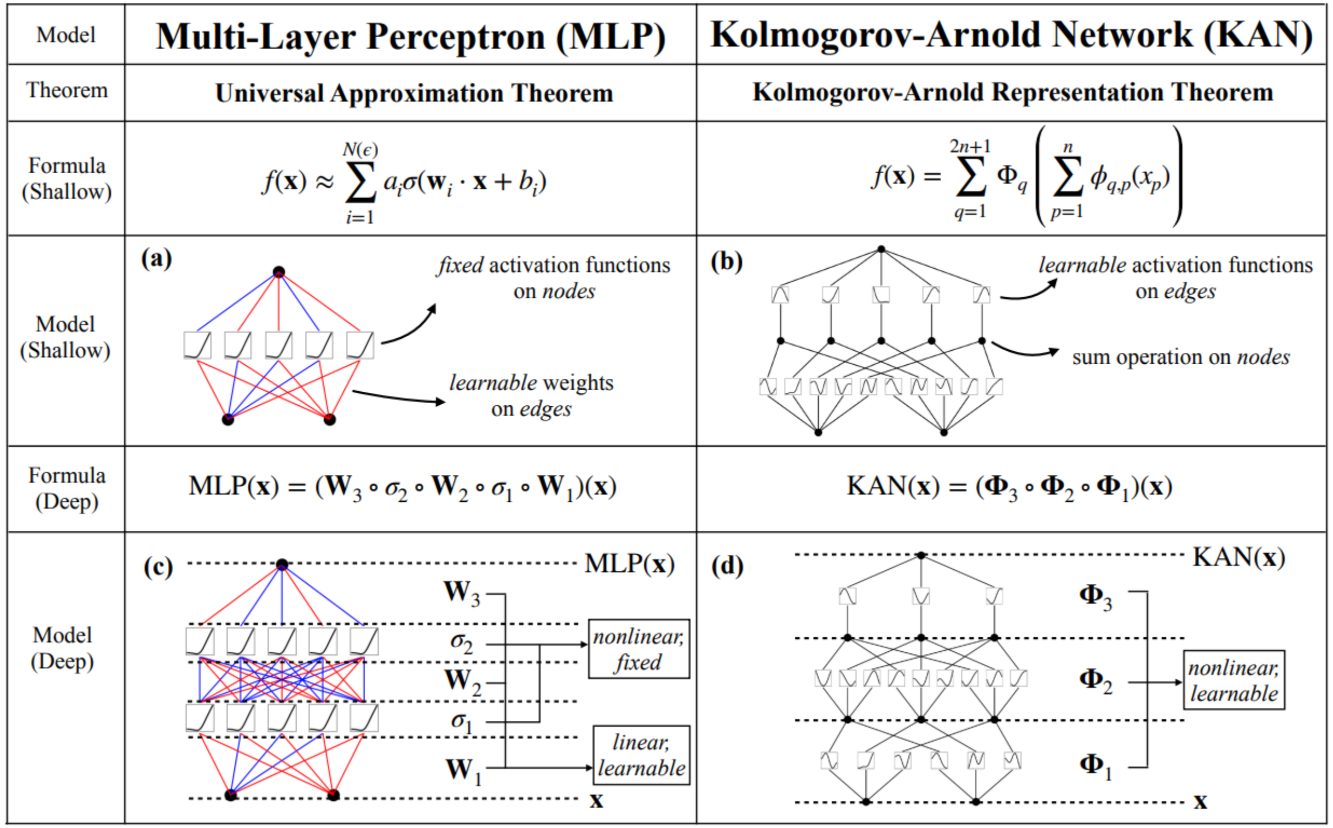
Теорема. *При любом целом $n \geqslant 2$ существуют такие определенные на единичном отрезке $E^1 = [0; 1]$ непрерывные действительные функции $\psi^{pq}(x)$, что каждая определенная на n-мерном единичном кубе E^n непрерывная действительная функция $f(x_1, \ldots, x_n)$ представима в виде*

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{q=2n+1} \chi_q \left[\sum_{p=1}^{n} \psi^{pq}(x_p) \right], \qquad (1)$$

где функции $\chi_q(y)$ действительны и непрерывны.

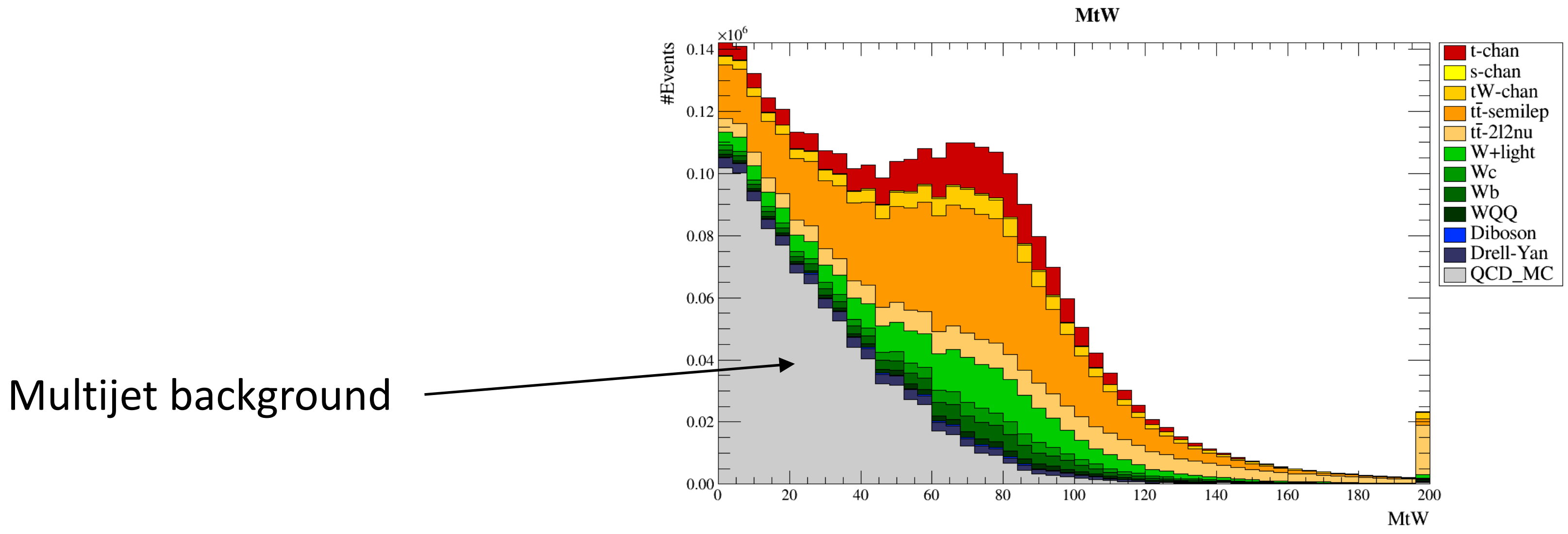
Introduction to KANs

- Learnable splines instead of linear functions
- Splines can be fine-grained by changing the grid
- Require fewer neurons than MLP, but are more computationally complex
- Splines can be manually fixed as any analytical function



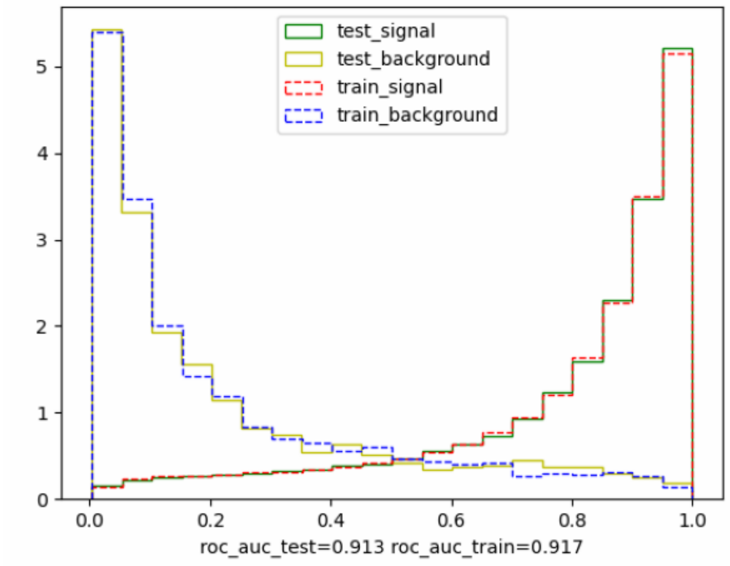
Classification task: multijet processes separation

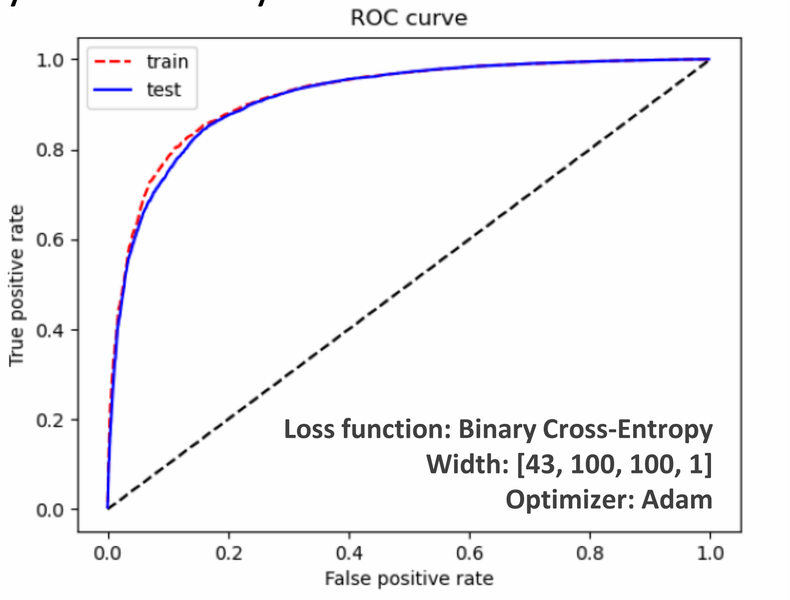
- One of the main tasks in the HEP analysis is to separate signal events from the background
- The most prominent at low energies and easiest to separate is the multijet background
- Pipeline of using MLPs for this tasks is already in use in the current analyses



Classification task: Baseline

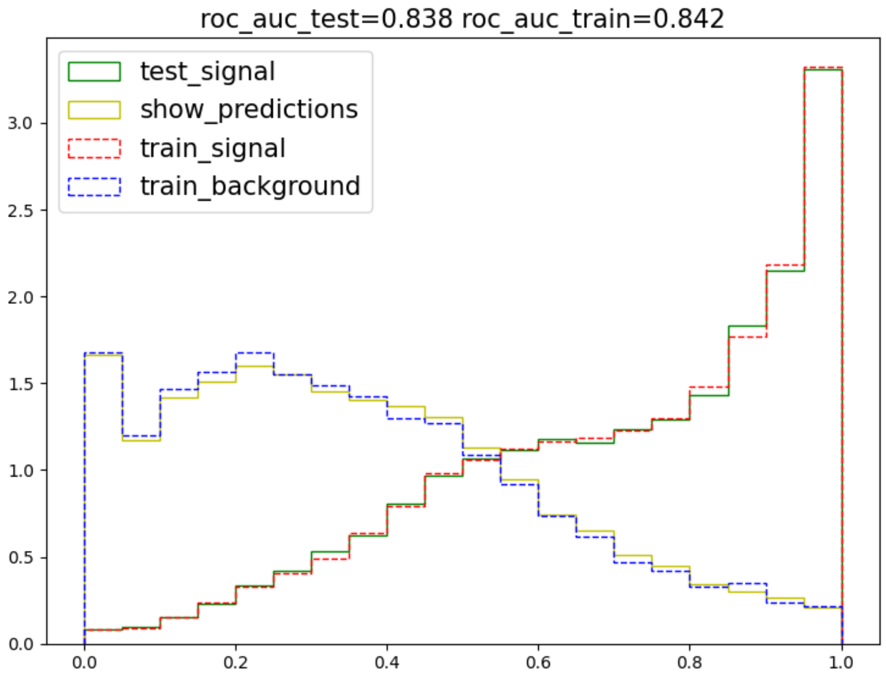
- 43 input variables
- Baseline - MLP with 2 layers with 100 neurons.
- This approach has demonstrated good efficiency in real analysis.

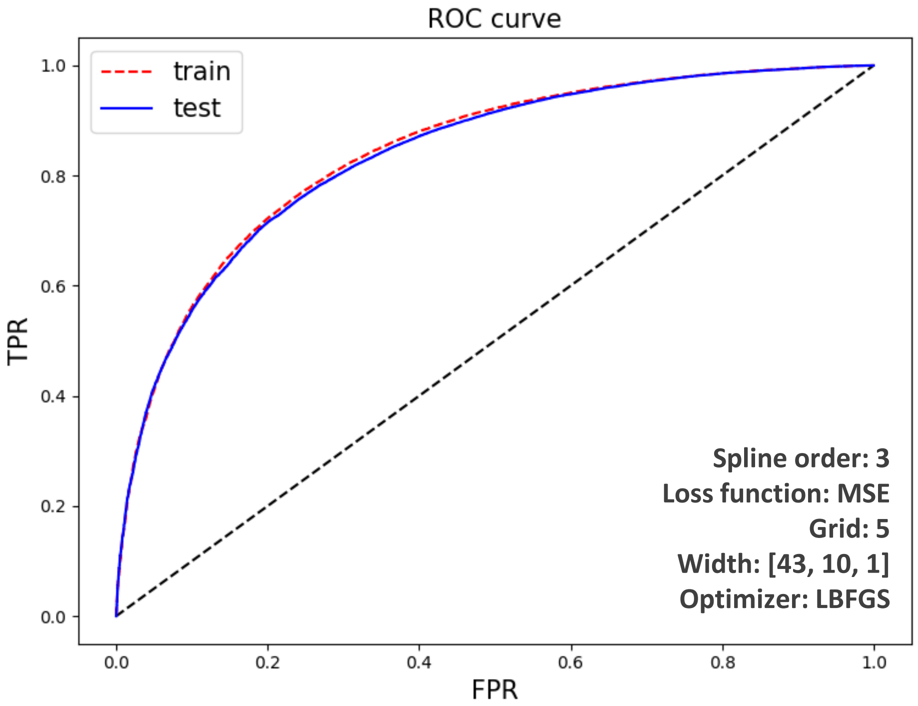




Classification task: KAN

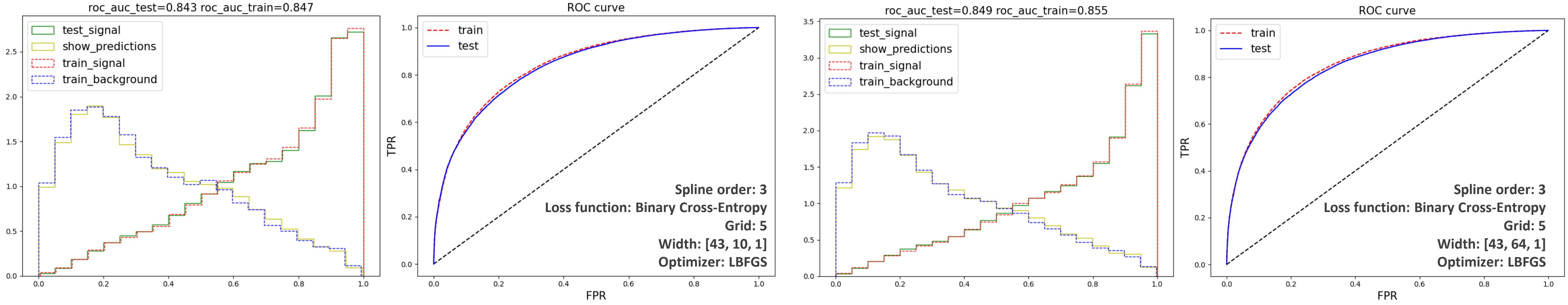
- Original KAN implementation suffers from various bugs.
- The only working solution for this task is using the simplest network with MSE loss instead of Cross-entropy.





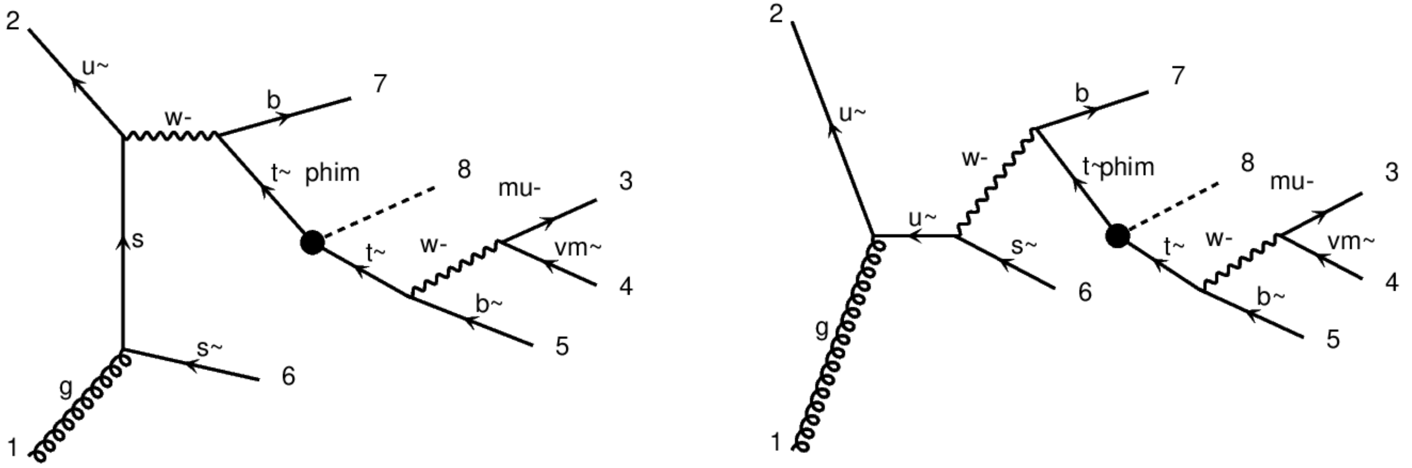
Classification task: Efficient KAN

- Efficient KAN is a standalone barebones implementation of KAN
- No pruning or symbolic functions, only basic training
- Does not exhibit aberrant behavior typical to original KANs
- Still does not beat the baseline MLP
- Hidden layer size does not affect performance



Regression task: Kinematics reconstruction

- Simplified Model of Dark Matter
- Mediator and neutrino cannot be observed in the detector
- Task: reconstruct 4-momenta of 2 invisible particles.



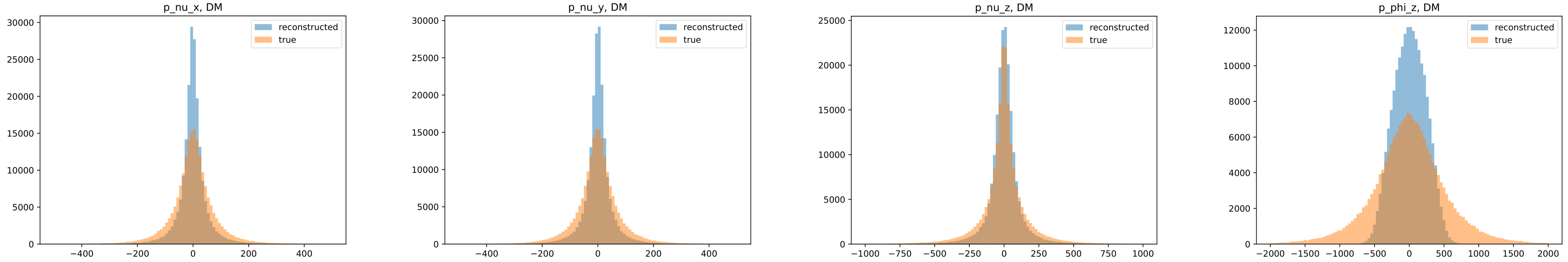
Typical Feynman diagrams for associated top-quark and DM production

Regression task: MLP baseline

Task : reconstruct 4-momenta of 2 invisible particles – neutrino and Dark Matter mediator.

Baseline – MLP, 3 hidden layers, 500 neurons per layer.

Best MAE on test – 0.28



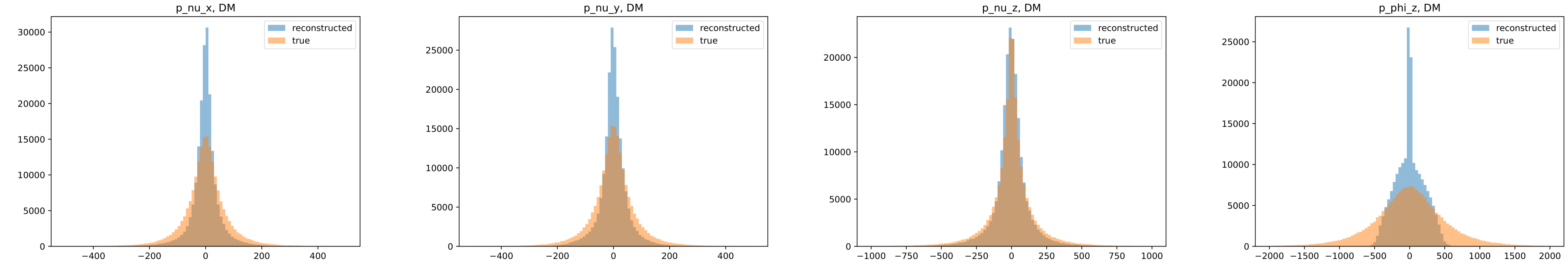
Neutrino and DM mediator 4-momenta reconstruction performance: MLP with MAE loss

Regression task: KAN application

Task : reconstruct 4-momenta of 2 invisible particles – neutrino and Dark Matter mediator.

KAN network: 2 hidden layers, 30 nodes per layer.

Best MAE on test – 0.3

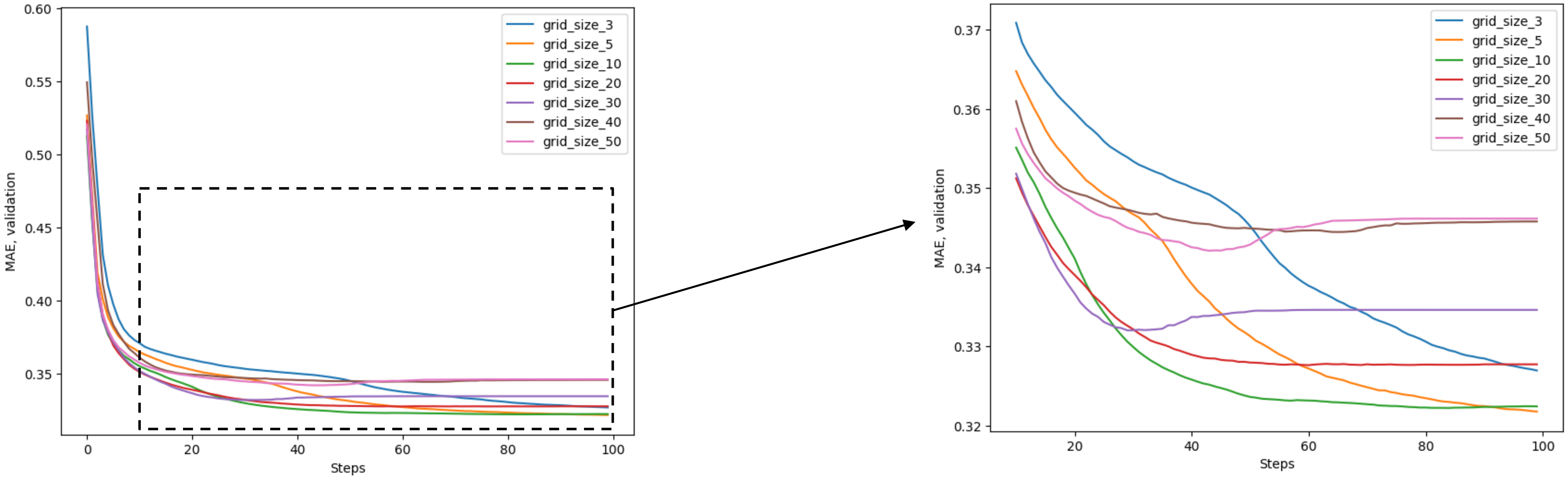


Neutrino and DM mediator 4-momenta reconstruction performance: KAN with MAE loss

Grid size dependency

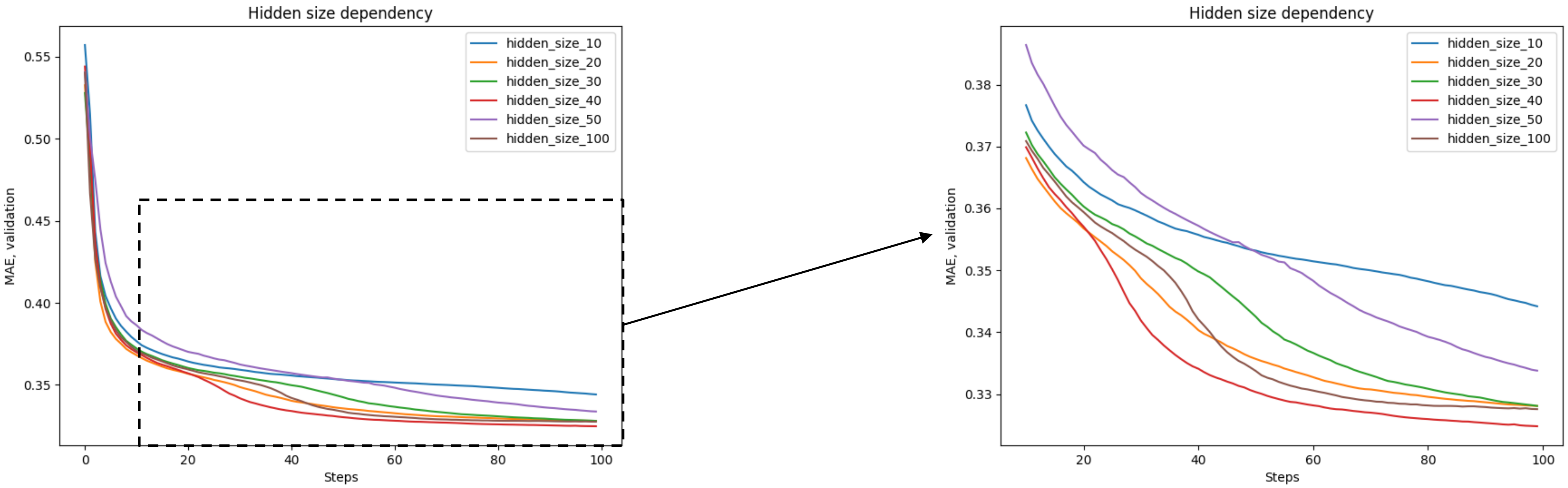
One of the main hyperparameters for KANs is the size of the grid, on which the splines are derived.

Lower values of the grid size can lead to poor accuracy, while higher values lead to overfitting.



Mean Absolute Error on validation for various grid sizes

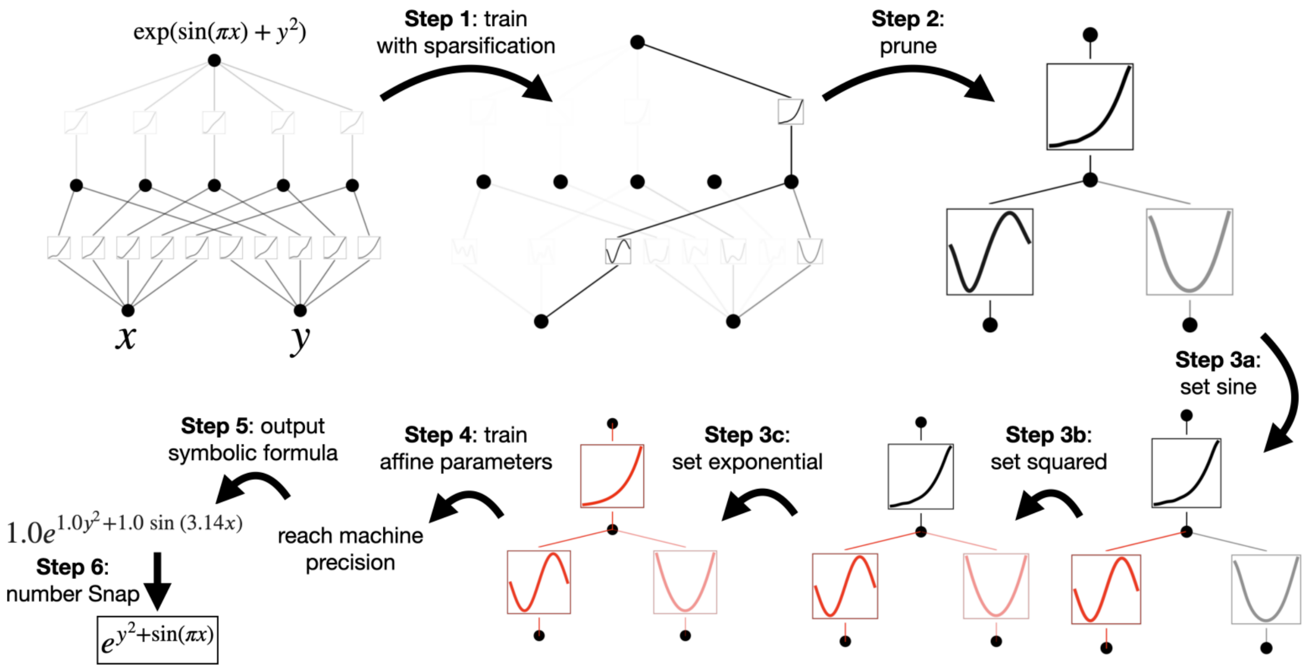
Hidden size dependency



Mean Absolute Error on validation for various hidden layer sizes

Perspectives: Symbolic functions

If the most critical bugs are fixed, the logical next step will be to apply the symbolic functionality of KANs to the analysis to obtain analytical formula for the solution.



Conclusion

1. While potentially intriguing, original KAN implementation suffers from many critical bugs, such as poor GPU memory utilization and computational errors, which do not allow one to realize it’s full potential.
2. Efficient KAN is a solid framework for initial KAN implementation, but many original features are missing and contains only a barebones KAN layer.
3. In both classification and regression tasks eKAN achieves similar results as baseline MLP without any significant improvements.
4. If bugs in original KAN are fixed, next steps will be implementing pruning and symbolic functions in the analysis.