Sept. 2022, Milan, Italy

Learning new physics efficiently with nonparametric methods

Letizia M.[1,2], **Losapio G.[1]**, Rando M.[1], Grosso G.[2,3,4], Wulzer A.[3], Pierini M.[4], Zanetti M.[2,3], Rosasco L.[1,5,6]

[1] MaLGa, DIBRIS, U. Genoa - [2] INFN - [3] U. Padua - [4] CERN - [5] MIT - [6] IIT

Motivation

- The recent work [1] provides physicists a useful tool to search for new physics with neural networks on low-dimensional datasets at the cost of a huge computational demand.

Contributions

- The theoretical framework introduced in [1] has been generalized to different loss functions and machine learning models
- Starting from the algorithm proposed in [1], a new, efficient algorithm based on kernel methods has been introduced with a hyperparameters tuning procedure
- a Python module named `ml4hep` has been created to train neural networks and kernel methods on local machines or a farm of computers

Problem setting

$$\{x_i\}_{i=1}^{\mathcal{N}_0} \sim p(x|0), N(0)$$

$$\{x_j\}_{j=1}^{\mathcal{N}_1} \sim p(x|1), N(1)$$

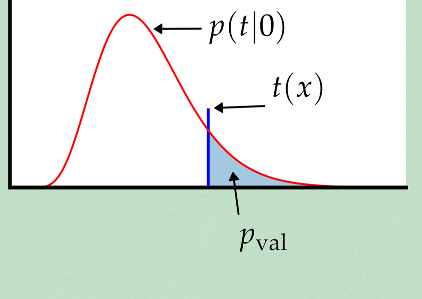
Reference Model

Experimental Data

$$\mathbf{p(x|0) = p(x|1)\ ?}$$

Idea hypothesis testing

1. $f_n(x) \approx \log \dfrac{p(x|1)}{p(x|0)} \longrightarrow t(x)$
2. Toy experiments: $p(t|0)$
3. p-value: $p_{\text{val}} \longrightarrow z_{\text{obs}}$



Remarks

- Inspired by [1]
- Unbalanced classification problem with $\mathcal{N}_0 \gg \mathcal{N}_1$
- Multiple training with $N_{\text{toy}} = \mathcal{O}(100)$

Data

Figure 1 - CMS experiment, Geneva, Switzerland

- LHC-like data of increasing dimensionality (1 to 28 features) produced by software simulations
- Benchmarked with different types of putative new physics signals

Experiments

- Neural networks vs LogFalkon*

Custom loss

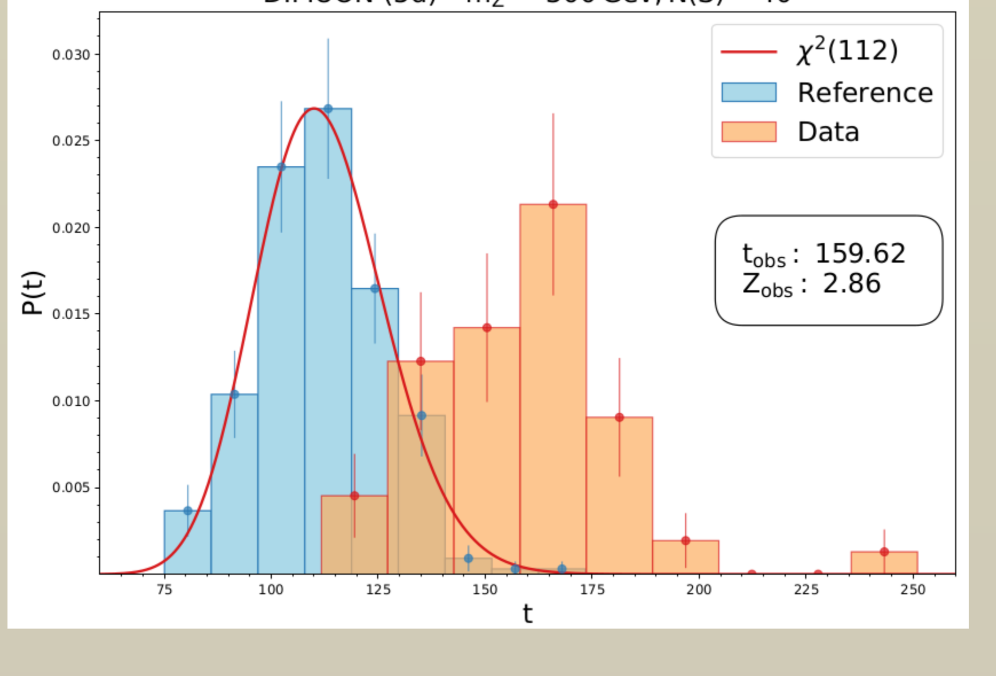
$$\ell(f(x), y) = (1-y)\frac{N(R)}{\mathcal{N}_{\mathcal{R}}}\sum_{x\in\mathcal{R}}(e^{f(x)}-1) - y f(x)$$

W.B.C.E.

$$\ell(f(x), y) = y\,\beta_1 \log\left(1+e^{-f(x)}\right) + (1-y)\,\beta_0 \log\left(1+e^{f(x)}\right)$$

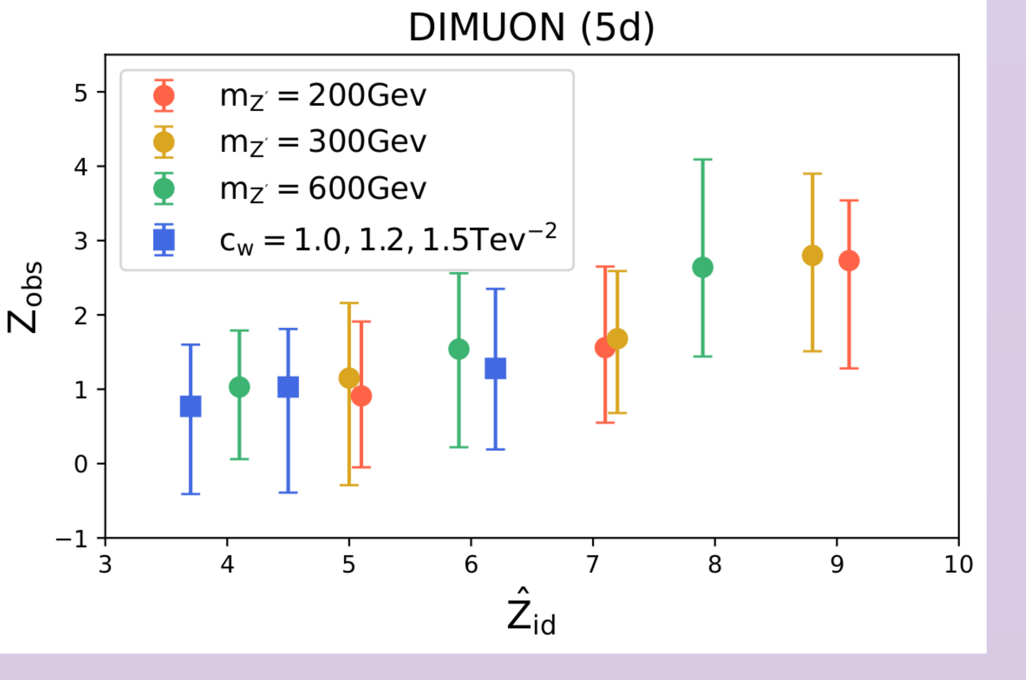
Physics-informed weights

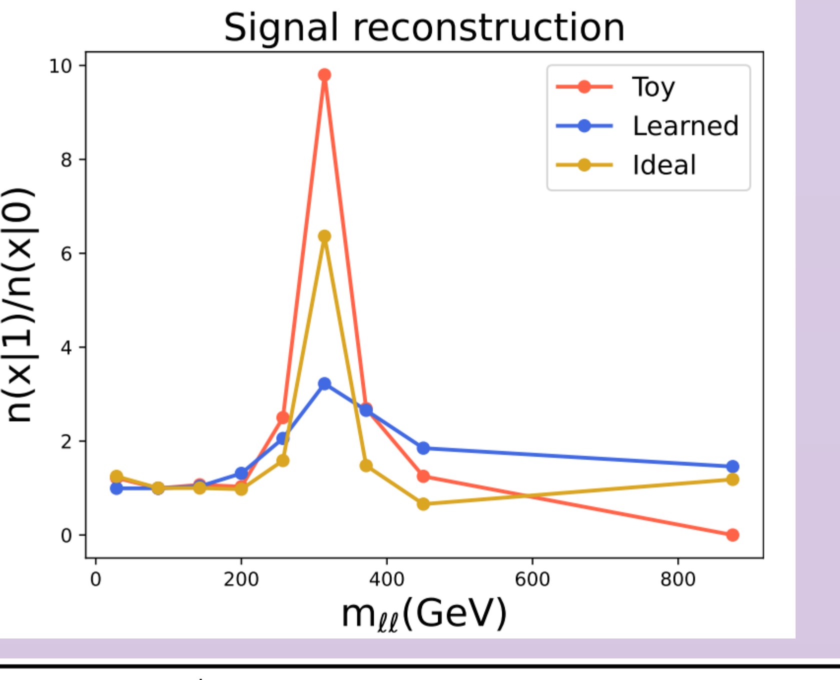
$$\beta_1 = \frac{\mathcal{N}_0}{N(0)},\ \beta_0 = 1 \longrightarrow f_n \approx \log\left[\frac{n(x|1)}{n(x|0)}\right]$$



- Hyperparameter tuning (see [2]) + performance assessment on different datasets

Results





Model	Training time	Resources
Our approach	$\mathcal{O}(s)$	**Single GPU machine**
Neural networks	$\mathcal{O}(h)$	Farm of CPUs

Conclusion

Figure 2 - Farm of computers vs a single laptop

- Delivered comparable performance to [1] with orders of magnitude gain in terms of training time and computational requirements

Future work

- Complete pipeline for data analysis
- A theory-grounded hyperparameter tuning procedure
- Improved Nyström centers selection
- Treatment of systematic uncertainties
- Improved parallelization on multiple GPU machines
- Extended strategy to other applications and domains

References

[1] d'Agnolo, R. T., et. al. (2021). Learning multivariate new physics. *The European Physical Journal C*, *81*(1), 1-21.

[2] Letizia, M., et. al. (2022). Learning new physics efficiently with nonparametric methods. *arXiv preprint arXiv:2204.02317.*

Credits

Figure 1 - home.cern

Figure 2 - (Left) Sashkin - stock.adobe.com, (Right) notebookcheck.it

*Library available at https://falkonml.github.io/falkon/

Contacts

Gianvito Losapio

gvlosapio@gmail.com, gvlosapio.netlify.com