ML-based likelihoods

Andrea Coccaro / Riccardo Torre
INFN Genova

Eur. Phys. J. C (2020) 80:664
https://doi.org/10.1140/epjc/s10052-020-8230-1

THE EUROPEAN PHYSICAL JOURNAL C

Check for updates

Special Article - Tools for Experiment and Theory

The DNNLikelihood: enhancing likelihood distribution with Deep Learning

Andrea Coccaro[1], Maurizio Pierini[2], Luca Silvestrini[2,3], Riccardo Torre[1,2,a]

[1] INFN, Sezione di Genova, Via Dodecaneso 33, 16146 Genoa, Italy
[2] CERN, 1211 Geneva 23, Switzerland
[3] INFN, Sezione di Roma, P.le A. Moro, 2, 00185 Rome, Italy

INFN

LHC legacy

- presentation of physics searches and measurements are usually condensed in plots or tables
- long-standing efforts for reducing the tremendous loss of information,
  e.g. efficiency parametrization, HEPData, RECAST, etc.
- also various discussions and actions on data preservation, open data, public release of experimental likelihoods
- LHC provided us with exceptional high-quality data, which has been scrutinised in countless ways for looking for new physics and for measuring SM properties
- need to ensure legacy for this treasure

  https://opendata.cern.ch/docs/about

CERN Open Data Portal

Documentation About

The CERN Open Data portal is the access point to a growing range of data produced through the research performed at CERN. It disseminates the preserved output from various research activities and includes accompanying software and documentation needed to understand and analyse the data.

The portal adheres to established global standards in data preservation and Open Science: the products are shared under open licenses; they are issued with a Digital Object Identifier (DOI) to make them citable objects.

Experimental likelihoods

- likelihoods capture the experimental information of physics analyses
- trade-off between information and compression
- the proposal in arXiv 1911.03305 is to *encode the experimental likelihood with all the dependence on elementary nuisance parameters into a deep neural-network function*

The approach of the proposal allows to

1. encode both binned and unbinned likelihoods
2. re-sample with different priors
3. ease combination of different likelihoods whenever correlations are known
4. adopt different statistical approaches
5. distribute likelihoods with a platform-independent at-large supported format such as ONNX

The DNNLikelihood

$$\boxed{(\vec{\mu}, \vec{\theta}) \to \mathcal{L}(\vec{\mu}, \vec{\theta})} \quad \Longrightarrow \quad \boxed{\mathcal{L}_{\mathrm{DNN}}(\vec{\mu}, \vec{\theta})}$$

Application

DNNLikelihood from the toy LHC-like likelihood considered in [JHEP 04 (2019) 064] consisting of 1 physical parameter (signal strength), 94 non-gaussian nuisance parameters (90 fully uncorrelated, 2 fully correlated, 2 normalisations)

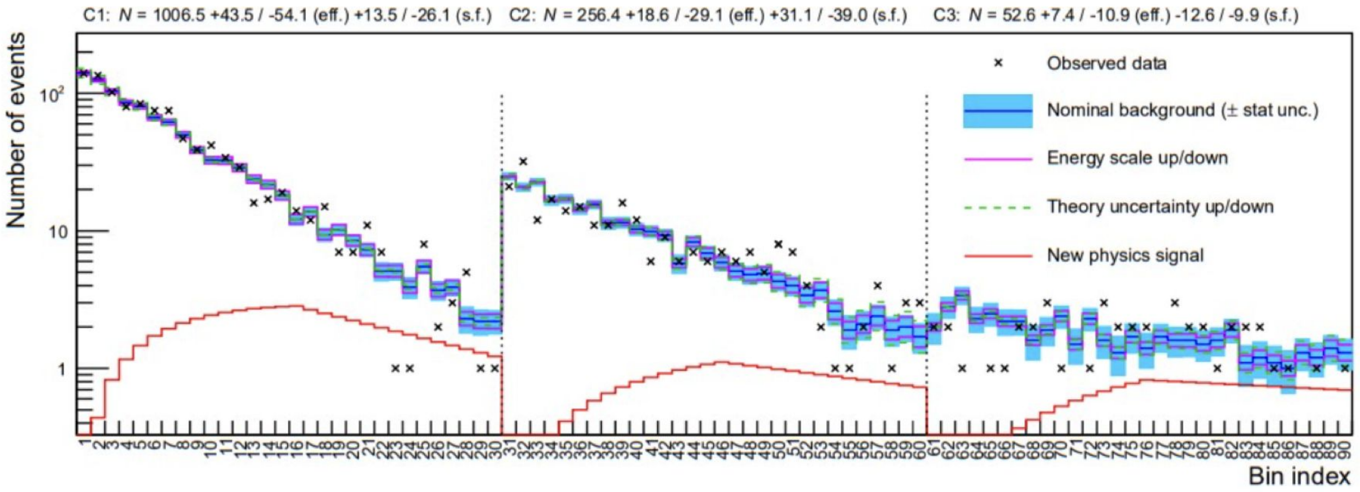
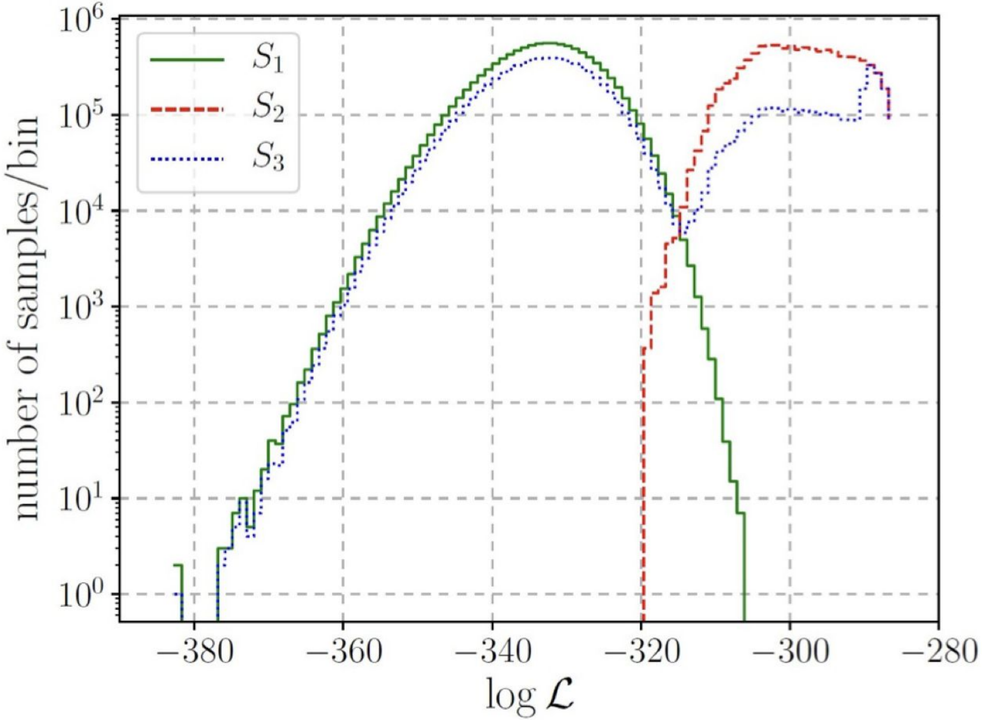


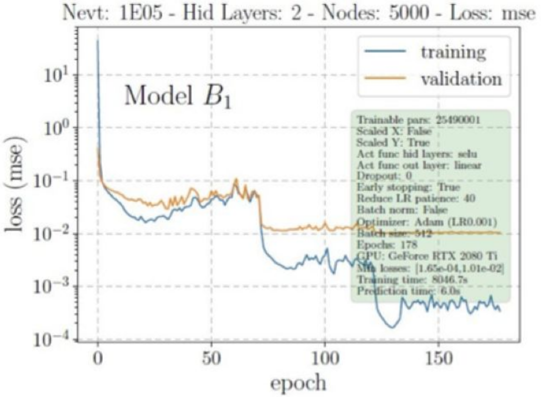
Figure 2. LHC-like search for new physics (mockup). The search is performed across three event categories, each divided into 30 bins to make a total of 90 search regions. The nominal expected contribution in each bin from the background and from the new physics signal is shown by the blue and red lines, respectively. The solid and dashed lines show the $\pm 1\sigma$ correlated variation in each bin expected due to an experimental and theoretical uncertainty while the blue shaded band shows the uncorrelated uncertainty in each bin due to limited MC simulation. The "observed" number of events in data in each bin is indicated by the black points.

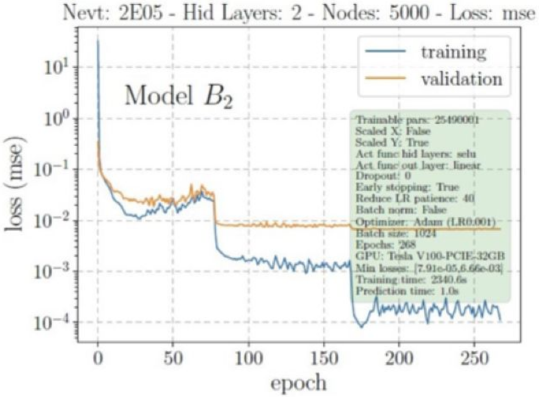
Sampling

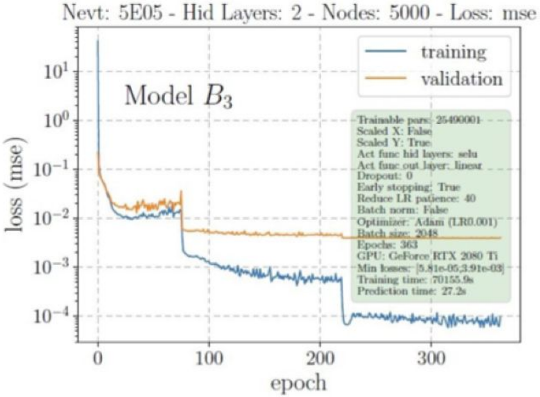
Likelihood function needs to be sampled for training the DNN. Allowing frequentist and bayesian inference means sampling accurately in different regions, i. e. close to maxima and where prior volume is large



Results of bayesian DNNLikelihood





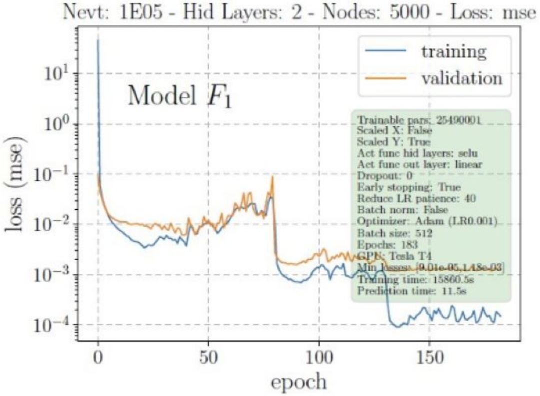


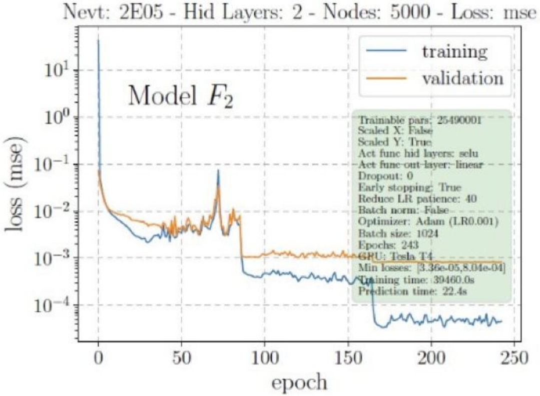
Metrics

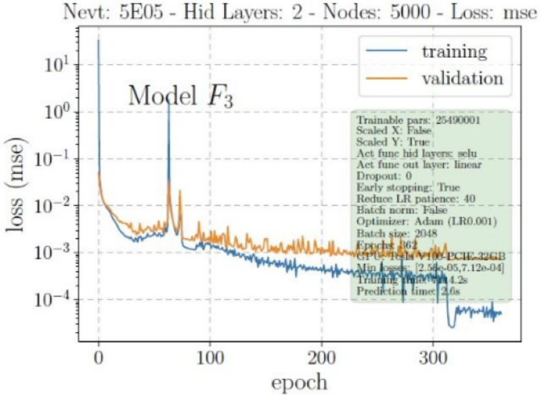
Name	B_1	B_2	B_3
Sample size ($\times 10^5$)	1	2	5
Epochs	178	268	363
Loss train (MSE) ($\times 10^{-3}$)	0.14	0.088	0.054
Loss val (MSE) ($\times 10^{-3}$)	10.11	6.66	3.9
Loss test (MSE) ($\times 10^{-3}$)	10.02	6.64	3.9
ME train ($\times 10^{-3}$)	0.47	0.53	0.28
ME val ($\times 10^{-3}$)	5.44	2.58	1.76
ME test ($\times 10^{-3}$)	4.91	2.31	1.72
Median p-value of 1D K-S test vs pred. on train	0.41	0.46	0.39
Median p-value of 1D K-S test vs. pred. on val.	0.24	0.33	0.43
Median p-value of 1D K-S val vs. pred. on test	0.24	0.40	0.34
Training time (s)	1007	2341	8446
Prediction time (μs/point)	11.5	10.4	14.5

HPDI	$\mu > 0$	B_1	B_2	B_3
68.27%	0.48	0.49	0.49	0.49
95.45%	0.86	0.92	0.91	0.88
99.73%	1.22	1.35	1.34	1.29

Results of full DNNLikelihood





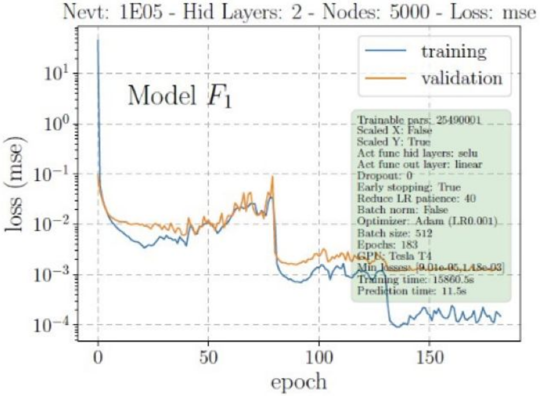


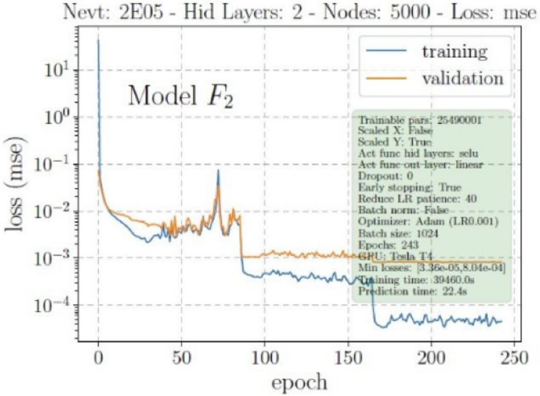
Metrics

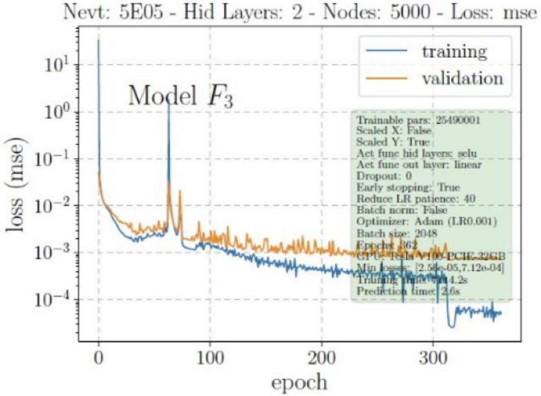
Name	F_1	F_2	F_3
Sample size ($\times 10^5$)	1	2	5
Epochs	183	243	362
Loss train (MSE) ($\times 10^{-3}$)	0.092	0.026	0.030
Loss val (MSE) ($\times 10^{-3}$)	1.18	0.80	0.71
Loss test (MSE) ($\times 10^{-3}$)	1.17	0.80	0.72
ME train ($\times 10^{-3}$)	3.07	0.47	1.1
ME val ($\times 10^{-3}$)	1.78	0.87	0.82
ME test ($\times 10^{-3}$)	1.50	0.68	0.86
Median p-value of 1D K-S test/pred-train	0.53	0.48	0.44
Median p-value of 1D K-S test/pred-val	0.15	0.27	0.20
Median p-value of 1D K-S val/pred-test	0.13	0.31	0.33
Mean error on $t_\mu(\mu)$	0.11	0.12	0.032
Training time (s)	1236	2819	7114
Prediction time (μs/point)	11.1	10.8	10.5

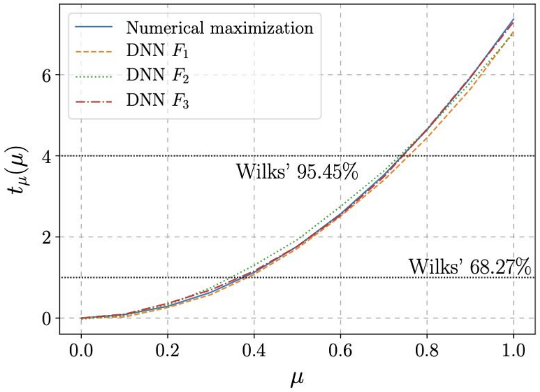
HPDI	$\mu > 0$	F_1	F_2	F_3
68.27%	0.48	0.50	0.50	0.48
95.45%	0.86	0.93	0.93	0.88
99.73%	1.22	1.35	1.37	1.28

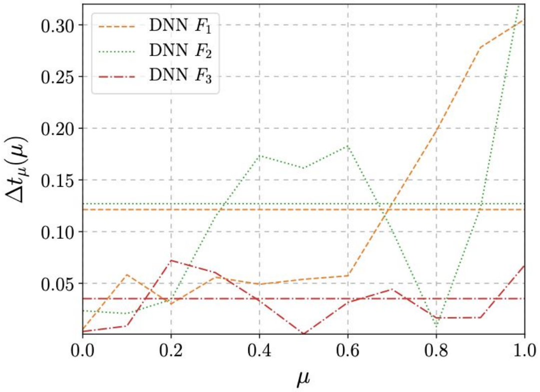
Results of full DNNLikelihood











DNN approach for flavour likelihood

- Example of complicated experimental likelihood
    a. No analytical formulation, multi-modal dimensions, complex correlation among nuisance parameters
    b. Obtained from the Bayesian fit in [arXiv:2011.01212](https://arxiv.org/abs/2011.01212)
    c. 83 parameters, of which 6 parameters of interest (SMEFT Wilson coefficients)
- DNN able to correctly predict the likelihood, with a 10^3 / 10^4 timing improvement

The DNNLikelihood Framework

What is this framework useful for?

- Encoding the likelihood function
- Sampling the likelihood object with an API to the emcee python package
- Encoding the DNN-version of the likelihood function
- Storing the data for constructing the DNN-version of the likelihood
- Managing ensemble of likelihoods for hyper-parameter optimisation
- Interfacing with HistFactory class used by ATLAS to encode likelihoods
- A suite of routines for performing statistical analysis and producing plots

Full documentation

- http://rtorre.web.cern.ch/rtorre/DNNLikelihood_doc/

Discussion

- Further improvements to the methodology
    a. Supervised vs unsupervised DNN approach
    b. Other comments on the bayesian vs frequentist statistical approaches in relation to the NN representation of the likelihood?
    c. NN representation of the full statistical model
- Scaling with complexity
    a. Successfully trained real experimental multi-modal flavour likelihood
    b. More complex yet realistic scenarios?
    c. Application to unbinned likelihoods
- Implementation
    a. Beta-version of a framework and related documentation for constructing ensemble of NN likelihoods and studying their properties is in place
    b. Can this approach be used within experiments as is? And outside such as fitting groups?

Contacts

You are welcome to contact us in case of curiosities

- andrea.coccaro@ge.infn.it
- riccardo.torre@ge.infn.it