PROCEEDINGS OF SCIENCE

PoS(LHCP2022)242

Constraining Deep Neural Network classifiers' systematic uncertainty via input feature space reduction

Andrea Di Luca,[a,b,*] **Marco Cristoforetti**[b,c] **and Roberto Iuppa**[a,b]

[a] *Dipartimento di Fisica, Università di Trento, Via Sommarive 14, 38123 Trento, Italy*
[b] *TIFPA, Via Sommarive 14, 38123 Trento, Italy*
[c] *FBK, Via Sommarive 18, 38123 Trento, Italy*

E-mail: andrea.diluca@unitn.it

In this work, we show how using a sub-optimal set of input features can lead to higher systematic uncertainty associated with Deep Neural Network classifier predictions. For this study, we considered the case of highly boosted di-jet resonances produced in pp collisions decaying to two b-quarks to be selected against an overwhelming QCD background. Results from a Monte Carlo simulation with HEP pseudo-detectors are shown.

The Tenth Annual Conference on Large Hadron Collider Physics - LHCP2022
16-20 May 2022
online

[*] Speaker

1. Introduction

In current and future high-energy physics experiments, the sensitivity of selection-based analysis will increasingly depend on the choice of the set of high-level features determined for each collision. In this context, Deep Learning approaches are widely used to improve the selection performance in physics analysis. A crucial aspect is that the results of a model based on a large number of input variables are more difficult to explain and understand. This point becomes relevant for Deep Neural Network (DNN) models since they do not provide uncertainty estimation and are often treated as perfect tools, which they are not.

In this work, we make the relationship between input feature space reduction and the effect on the DNN classifier's performance systematic uncertainty explicit.

2. Boosted $H \to bb$ tagging

As a benchmark application, we developed a fully connected DNN to classify pp collision events where a Higgs boson with very high transverse momentum decays to two b-quarks. In this regime, the decay products of the Higgs boson are very collimated and it is challenging to resolve the di-jet structure [1]. A single large and massive jet containing both the b-quarks originated jets are more likely to be reconstructed. Recognizing these events in a pp collision experiment represents a challenging task, mainly because of the huge irreducible background of QCD multi-jet production.

2.1 Simulated data and object reconstruction

The dataset used to develop the classifier is produced using a framework developed by combining Pyhtia8 [2], to generate high-energy physics events, Delphes [3], to simulate the detector response and RAVE [4] for secondary vertex reconstruction. Large-radius anti-k_t jets [5] (large-R jets) with $R = 1$ are reconstructed together with variable-radius track jets [6] with $R_{\text{MAX}} = 0.4$, $R_{\text{MIN}} = 0.02$ and $\rho = 30$ GeV (ρ parameter determines how fast the effective jet size decreases with the transverse momentum of the jet). For large-R jets, we defined kinematic variables plus jet substructure variables. For the variable R track jets, we defined kinematic variables plus the b-tagging information and variables connected to the secondary vertex. We selected large-R jets with $p_T > 450$ Gev$/c^2$ and $\eta < 2$. Then we look for the 2 highest p_T track jets contained in a selected large-R jet. The total number of initial features is 39. Then, the features are importance-sorted using the automated feature selection, described in [7].

3. Input feature space reduction effect on systematic uncertainty

Reducing the size of the input parameter space is relevant to estimate the uncertainty of the output of the classifier. This hypothesis is validated by looking at the weights associated with each feature in the input layer. The same model was trained 100 times by changing the random seed, which controls the initialization of the random weight. The weight distribution for each of the features was considered. Figure 1 summarizes the steps to achieve the distribution of each of the weights. Given the size of the first hidden layer L (in this case $L = 128$), each of the input features

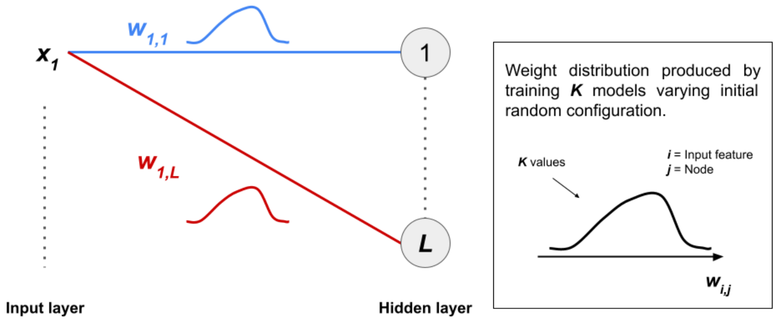

Figure 1: By training a model K times (in this work $K = 100$) by varying the initial random configuration, different weight values $w_{i,j}$, where i runs over the input features and j over the hidden layer's nodes, can be achieved. For each of the distributions, it is possible to have an estimate of the $\mu_{i,j}$ and σ_{i_j}.

i will be associated with $j = 0, \ldots, L$ weight distributions, each with a mean $\mu_{i,j}$ and σ_{i_j}. Figure 2 shows a box plot produced by looking at the normalized $\sigma_{i_j}/|\mu_{i,j}|$ for each feature. Then each of the boxes is produced using $L = 128$ estimates of the standard deviations of the weight distributions. The boxes are sorted using the feature ranking. By looking at the plot, we can observe how the position in the ranking is linked to the variability of the weight of the feature in the network: the lower the normalized variability, the higher the ranking, i.e. the greater the importance of the considered input feature. This fact suggests that changes in the irrelevant feature weights can lead to fluctuations in the model performance.

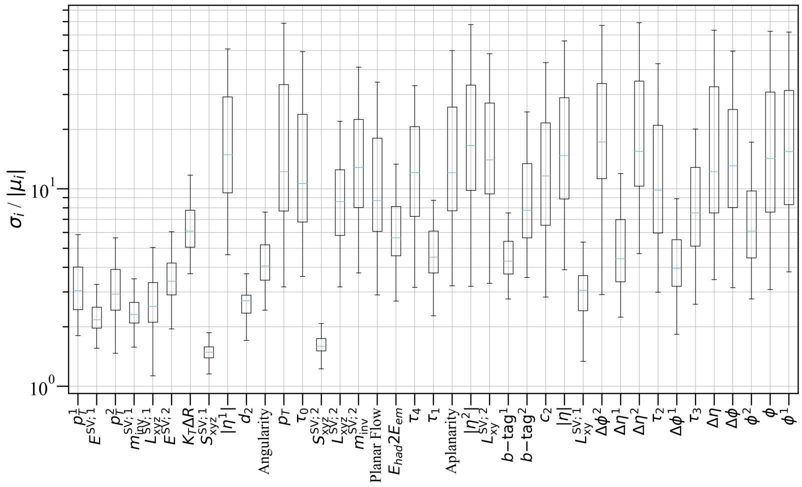

Figure 2: Box plot of normalized standard deviations of weights associated with each input variable.

This hypothesis is tested by considering different models that are trained by choosing a fixed number of input features. The order used to select the features is the one defined by the CatBoost feature ranking. Each model is trained 100 times by changing the random seed. Then, we considered the performance of each model. The left-side plot in Figure 3 shows the box plot obtained by looking at the background rejection rate for a fixed working point $WP = 0.85$ for the Higgs boson tagging efficiency, which is a common WP for physics analysis. The performance of the models grows by increasing the number of input features to the network up to 24 features where a plateau is reached.

A remarkable behavior is found by looking at the mean value and standard deviation of the performance for each box, that are shown in center and right-side plots of Figure 3. The normalized standard deviation starts at low values. This is expected since the model is not properly working and no large performance fluctuations are expected in this regime by considering different training setups that start from different random configurations. As the model performance improves, larger standard deviations are observed up to a point where a plateau is reached at 24-30 used features. When new features are added, the mean values of the background rejection rate stay almost unchanged, while larger standard deviations are observed. Using irrelevant quantities among input features has the major drawback of increasing the standard deviation of the expected performance. This fact translates into higher systematic uncertainties in physics analysis when the output of a DNN model is used for event selection.

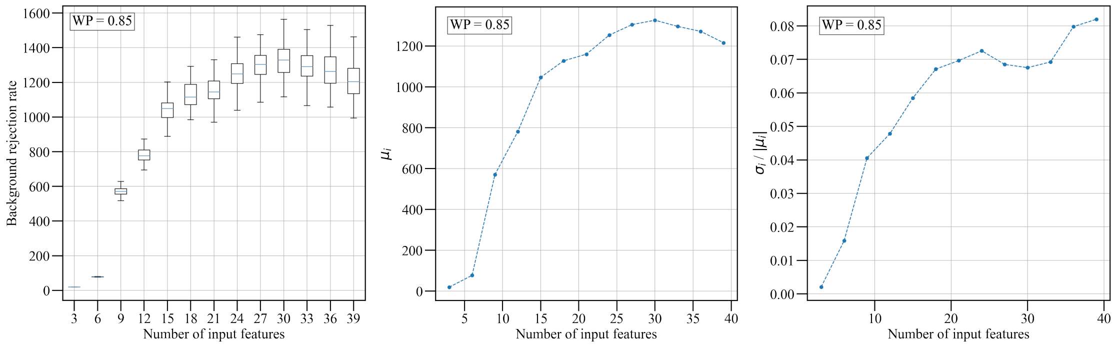

Figure 3: (Left) Box plot of background rejection rate measured at a fixed working point $WP = 0.85$ by varying the number of used input feature. Mean (Center) and normalized standard deviation (Right) of the box plot.

4. Conclusion

In this work, we shown how a proper choice of the input features in a DNN model can affect the model prediction, avoiding the inclusion of undesired systematic uncertainties due to irrelevant features. We used a classification task with 39 features as a case study: events with a boosted large and massive jet containing both of the b quarks originating from H boson decay are discriminated from the background.

References

[1] ATLAS Collaboration collaboration, *Performance of large-R jets and jet substructure reconstruction with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2012-065, CERN, Geneva (Jul, 2012).

[2] T. Sjöstrand et al., *An introduction to pythia 8.2*, *Computer Physics Communications* **191** (2015) 159–177.

[3] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [1307.6346].

[4] W. Waltenberger, *Rave—a detector-independent toolkit to reconstruct vertices*, *Nuclear Science, IEEE Transactions on* **58** (2011) 434 .

[5] M. Cacciari et al., *The anti-ktjet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008) 063–063.

[6] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* **06** (2009) 059 [0903.0392].

[7] A. Di Luca, M. Cristoforetti, F.M. Follega and R. Iuppa, *Automatic selection of observables for the analysis of high-energy particle jets*, *Nuovo Cim. C* **44** (2021) 42.