

Chapitre X

Reconstruction de la masse d'une résonance grâce au *Machine Learning*

Sommaire

1	Introduction	2
2	Le Machine Learning	3
2.1	Généralités	3
2.2	Le <i>Gradient Boosting</i>	3
2.3	Le <i>Deep Learning</i>	3
3	Application du Machine Learning aux événements $H \rightarrow \tau\tau$	3
3.1	Génération des événements	3
3.2	Variables d'entrées	3
3.3	Performances sur les événements de test	3
3.4	Performances sur les événements de l'analyse CMS	3
4	Prise en compte de l'empilement	3
4.1	Génération des événements	3
4.2	Performances	3
4.3	Variables d'entrées supplémentaires	3
4.4	Performances	3
5	Effets sur les résultats de l'analyse MSSM HTT	3
6	Conclusion	3

Citations incontournables :

- DELPHES 3.4.2 [1, 2] ?
- CMS Fast Simulation (FASTSIM) [3-6]
- PYTHIA 8.235 [7]
- FASTJET [8, 9]
- KERAS [10]
- TENSORFLOW [11]
- XGBOOST [12]
- [13] for an example of nn use in HEP
- [14]
- [15]
- SVFIT [16]

Citer également la thèse de Gaël :

G. TOUQUET. « Search for an additional neutral MSSM Higgs boson decaying to tau leptons with the CMS experiment ». Thèse de doct. Université Claude Bernard Lyon 1, oct. 2019. URL : <https://hal.archives-ouvertes.fr/tel-02526393>

- type of samples/events
- preselection (small HTT analysis)
- inputs
- performances : métrique ?

- mass range + plots
- METcov + plots
- PU + plots

Étapes des choix

Inputs variables

phéno, tau1 tau2 MET pT eta phi + mT 1 2 tt tot.

Inputs events

sélection des événements, target flattening. pas de PU
80-800 GeV (aller plus loin que [15])

DNN

Comme dans [15], table 1.
output activation function to linear instead of relu to not cut?
changement de la structure : 3 couches de 1000 neurones?
loss mse
optimizer adadelta
w_init_mode uniform
early stopping
Changement du mass range (biais aux bords) -> down to 50 GeV

XGB

avoid overfitting, max_depth
learning rate (eta)
num_round, early stopping (choose value of 5)
loss

1 Introduction

explosion ces dernières années de l'IA

Siri, voiture autonome, AlphaGO, ...

The Elements of statistical learning : Trees have one aspect that prevents them from being the ideal tool for predictive learning, namely inaccuracy. -> they work great with the data used to create them, but they are not flexible when it comes to classifying new samples.

IA > ML > DL

utilisation en HEP ([13, 15, 17])

cas de la masse di- τ [15], aller plus loin : au lieu de 80 à 300 GeV par pas de 5, de 50 à 800 par pas de 1

deux modèles, XGBOOST et un DNN, présentés dans la section...

application aux événements HTT section ...

Mais dans la vraie vie, empiement : prise en compte section ...

Et ajout de variables supplémentaires pour prendre en compte PU section ...

Enfin, l'apport de ces techniques à l'analyse présentée dans le chapitre ... en section ...

2 Le *Machine Learning*

2.1 Généralités

2.2 Le *Gradient Boosting*

2.3 Le *Deep Learning*

3 Application du *Machine Learning* aux événements $H \rightarrow \tau\tau$

3.1 Génération des événements

3.2 Variables d'entrées

3.3 Performances sur les événements de test

3.4 Performances sur les événements de l'analyse CMS

4 Prise en compte de l'empilement

4.1 Génération des événements

4.2 Performances

(sur ces nouveaux événements)

4.3 Variables d'entrées supplémentaires

4.4 Performances

(avec les nouvelles variables)

5 Effets sur les résultats de l'analyse MSSM HTT

(remplacement de $m_{tt\bar{t}}$ par les prédictions du meilleur modèle, nouveaux plots d'exclusion, comparaison)

6 Conclusion

Références

- [1] J. de FAVEREAU & coll. « DELPHES 3 : a modular framework for fast simulation of a generic collider experiment ». *Journal of High Energy Physics* **2** (fév. 2014). DOI : [10.1007/jhep02\(2014\)057](https://doi.org/10.1007/jhep02(2014)057).
- [2] A. MERTENS. « New features in DELPHES 3 ». *Journal of Physics : Conference Series* **608.1** (2015). Sous la dir. de L. FIALA, M. LOKAJICEK & N. TUMOVA. DOI : [10.1088/1742-6596/608/1/012045](https://doi.org/10.1088/1742-6596/608/1/012045).
- [3] S. ABDULLIN & coll. « The Fast Simulation of the CMS Detector at LHC ». *Journal of Physics : Conference Series* **331.3** (déc. 2011). DOI : [10.1088/1742-6596/331/3/032049](https://doi.org/10.1088/1742-6596/331/3/032049).
- [4] A. GIAMMANCO. « The Fast Simulation of the CMS Experiment ». *Journal of Physics : Conference Series* **513.2** (juin 2014). DOI : [10.1088/1742-6596/513/2/022012](https://doi.org/10.1088/1742-6596/513/2/022012).
- [5] M. KOMM. « Fast emulation of track reconstruction in the CMS simulation ». *Journal of Physics : Conference Series* **898** (oct. 2017). DOI : [10.1088/1742-6596/898/4/042034](https://doi.org/10.1088/1742-6596/898/4/042034).
- [6] S. SEKMEN. *Recent Developments in CMS Fast Simulation*. 2017. arXiv : [1701.03850](https://arxiv.org/abs/1701.03850).
- [7] T. SJÖSTRAND & coll. « An Introduction to PYTHIA 8.2 ». *Computer Physics Communications* **191** (2015), p. 159-177. DOI : [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024). arXiv : [1410.3012](https://arxiv.org/abs/1410.3012) [hep-ph].

- [8] M. CACCIARI, G. P. SALAM & G. SOYEZ. « FASTJET user manual ». *European Physical Journal* **C72** (nov. 2012). DOI : [10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2). arXiv : [1111.6097](https://arxiv.org/abs/1111.6097) [hep-ph].
- [9] M. CACCIARI & G. P. SALAM. « Dispelling the N^3 myth for the k_T jet-finder ». *Physics Letters* **B641.1** (sept. 2006), p. 57-61. DOI : [10.1016/j.physletb.2006.08.037](https://doi.org/10.1016/j.physletb.2006.08.037).
- [10] F. CHOLLET & coll. KERAS. <https://keras.io>. 2015.
- [11] M. ABADI & coll. TENSORFLOW : *Large-scale machine learning on heterogeneous distributed systems*. Software available from tensorflow.org. 2015. URL : <https://www.tensorflow.org/>.
- [12] T. CHEN & C. GUESTRIN. « XGBOOST : A Scalable Tree Boosting System ». *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (août 2016). DOI : [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [13] D. GUEST & coll. « Jet flavor classification in high-energy physics with deep neural networks ». *Physical Review* **D94.11** (déc. 2016). DOI : [10.1103/physrevd.94.112002](https://doi.org/10.1103/physrevd.94.112002).
- [14] W. SARLE. « Neural Networks and Statistical Models ». 1994. URL : https://people.orie.cornell.edu/davidr/or474/nn_sas.pdf.
- [15] P. BÄRTSCHI & coll. « Reconstruction of τ lepton pair invariant mass using an artificial neural network ». *Nuclear Instruments and Methods in Physics Research* **A929** (2019), p. 29-33. DOI : [10.1016/j.nima.2019.03.029](https://doi.org/10.1016/j.nima.2019.03.029). URL : <http://www.sciencedirect.com/science/article/pii/S0168900219303377>.
- [16] L. BIANCHINI & coll. « Reconstruction of the Higgs mass in $H \rightarrow \tau\tau$ Events by Dynamical Likelihood techniques ». *Journal of Physics : Conference Series* **513.2** (juin 2014). DOI : [10.1088/1742-6596/513/2/022035](https://doi.org/10.1088/1742-6596/513/2/022035).
- [17] G. TOUQUET. « Search for an additional neutral MSSM Higgs boson decaying to tau leptons with the CMS experiment ». Thèse de doct. Université Claude Bernard Lyon 1, oct. 2019. URL : <https://hal.archives-ouvertes.fr/tel-02526393>.

