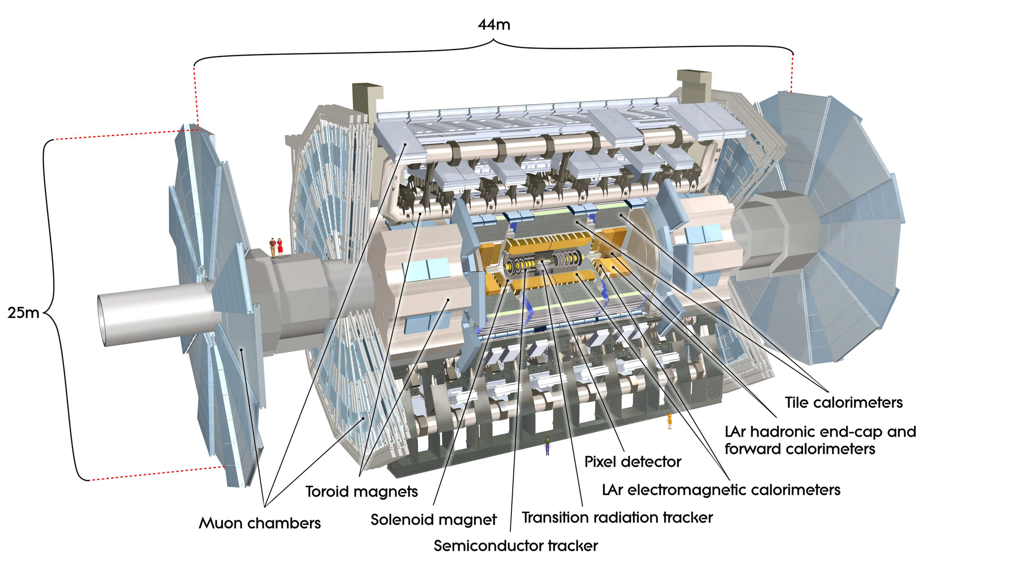
Adversarial Neural Networks for Decorrelated Jet Substructure Observables in ATLAS

Physics Motivation

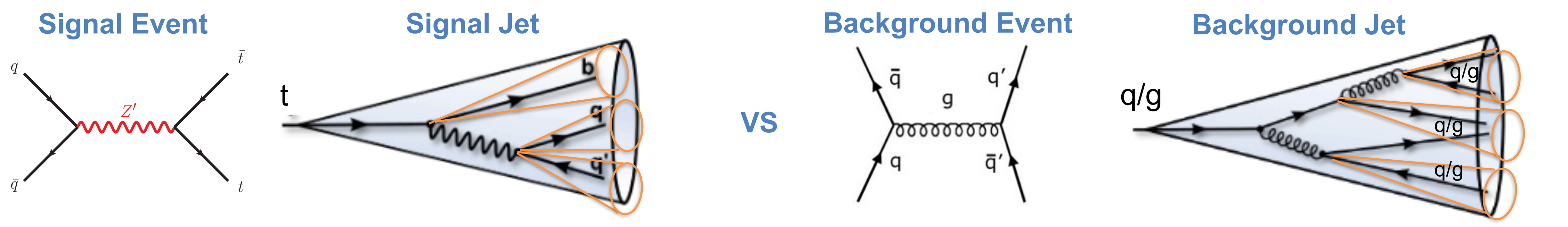
Boosted decision trees (BDTs) and deep neural networks (DNNs) are powerful techniques for classifying physics observables in the ATLAS detector known as "Jets", which are experimental representation of particles in high energy physics (HEP). Strong correlations between these techniques with the jet mass exist, spoiling new physics searches in the jet mass spectra. Adversarially trained neural networks (ANN) are presented as a way to train powerful taggers, decorrelated from the jet mass by construction. A comprehensive study in Monte Carlo simulation comparing ANN to different decorrelation algorithms in HEP is performed, which shows ANN to be the most beneficial.

The ATLAS Detector



Jet Classification

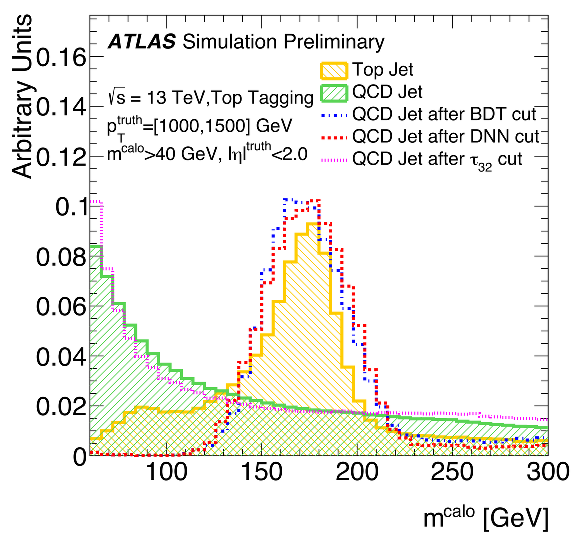
Theoretical models predict that heavy exotic particles can yield boosted top quarks (and W-bosons), whose decay is contained at detector level within a single "signal" jet. It is fundamental for new physics searches to distinguish these signal jets from (the more common) background jets generated by QCD fragmentation of gluons or light quarks.



Standard BDTs and DNNs Classifiers

BDTs and DNNs taggers result in a single discriminant $\mathscr{W}$ which is highly correlated with the jet mass.

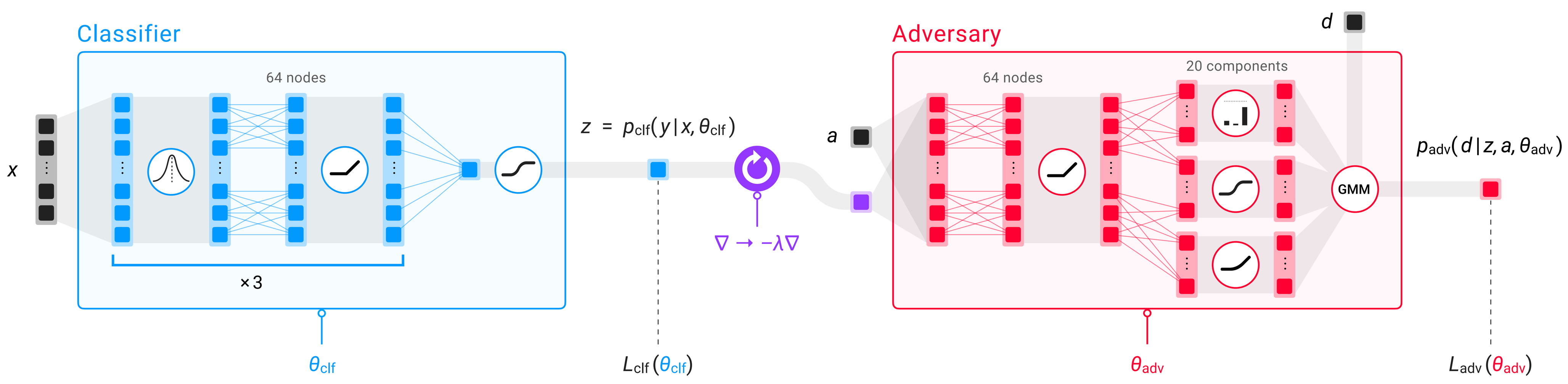
After a cut in $\mathscr{W}$, the background shape turns out similar to the signal one. Thus introducing fake peaks and so spoiling new physics searches in the jet mass spectra.



https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-004/

Novel Proposal

Use of adversarial neural networks to leverage classification power while ensuring mass-decorrelation. Technique inspired from Generative Adversarial Networks (GANs) developed by Ian Goodfellow et. al.



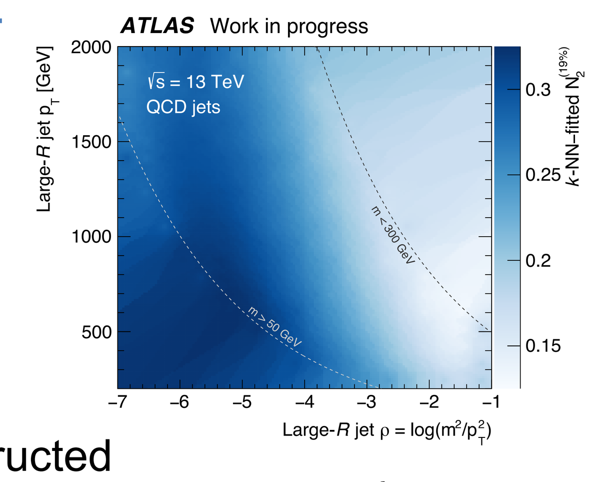
Auxiliary decorrelated learning algorithms

k-NN regression

1. The N_2 distribution (jet substructure observable) is computed in a two dimensional profile for a given bkg. eff.

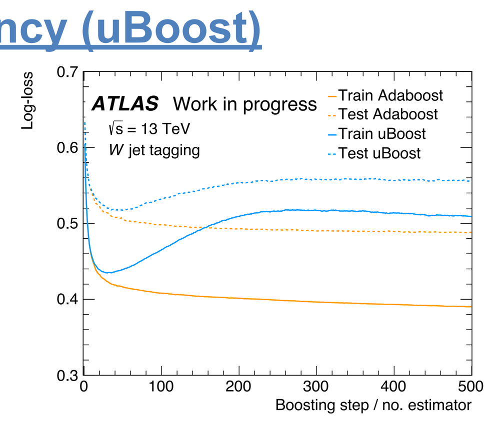
2. A two-dimensional regression fit using k-NN algorithm is performed yielding the N_2 functional dependence.

3. For each jet, a new observable is constructed by subtracting the predicted values from N_2: $N_2^{k-NN} = N_2 - N_2^{pred}$



Adaptive boosting for uniform efficiency (uBoost)

uBoost technique gives larger (smaller) weights in event regions of m where the selection efficiency is lower (higher) than the mean.



Training Procedure

1. Classifier is **pre-trained** to perform binary classification of the jet labels y based in input features x using a binary cross-entropy loss $L_{clf}(\theta_f)$

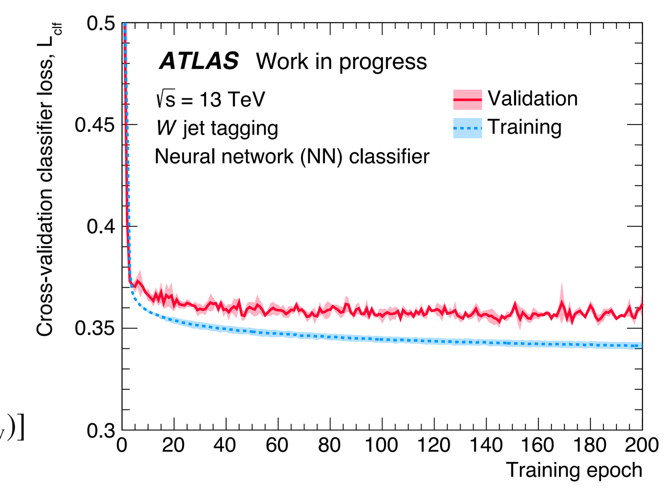
2. Adversarial is **pre-trained** for a fixed L_{clf}

$$L_{adv}(\theta_{adv}) = \mathbb{E}_{z\sim p_{clf}(X,\theta_{clf})}\mathbb{E}_{d\sim D}\mathbb{E}_{a\sim A}[-\log p_{adv}(d\,|\,z,a,\theta_{adv})]$$

A Gaussian Mixture Model is used to predict the jet mass called the adversary posterior:

$$P_{adv}(m|z,a,\theta_{clf})$$

Classifier pre-training

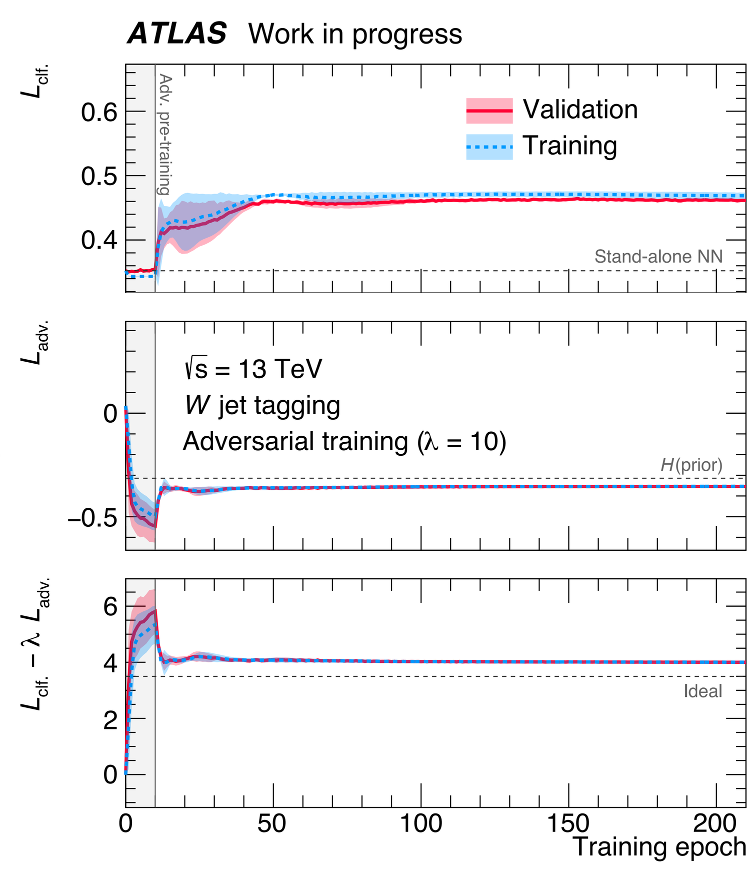


3. **Overall training** is computed by the effective loss:

$$\min_{\theta_{clf}}\max_{\theta_{adv}} L_{classifier} = L_{clf}(\theta_{clf}) - \lambda L_{adv}(\theta_{adv})$$

using gradient reversal layer scaled by λ

Overall training

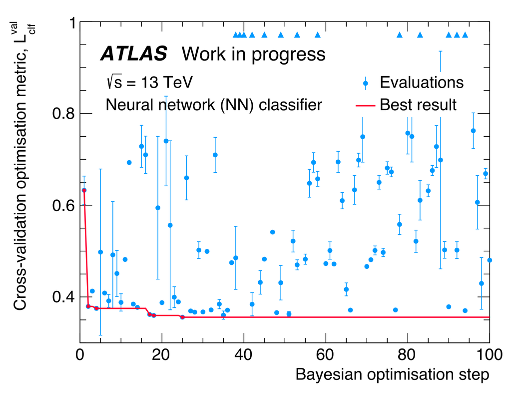


Hyperparameters Optimisation

3-fold stratified CV bayesian optimisation technique with Spearmint.

Pros: Capable of probing a large parameter space with few iterations.

Cons: No guarantee to find the optimal configuration.



Evaluation Metrics

Classification

Selection signal efficiency ϵ_{sig} and associated background rejection $1/\epsilon_{bkg}$ for cuts on the classifier variable

$$\varepsilon_{sig} = \frac{N_{sig}^{baseline\ \&\ pass}}{N_{sig}^{baseline}} \qquad \frac{1}{\varepsilon_{bkg}} = \frac{N_{bkg}^{baseline}}{N_{bkg}^{baseline\ \&\ pass}}$$

Mass-decorrelation

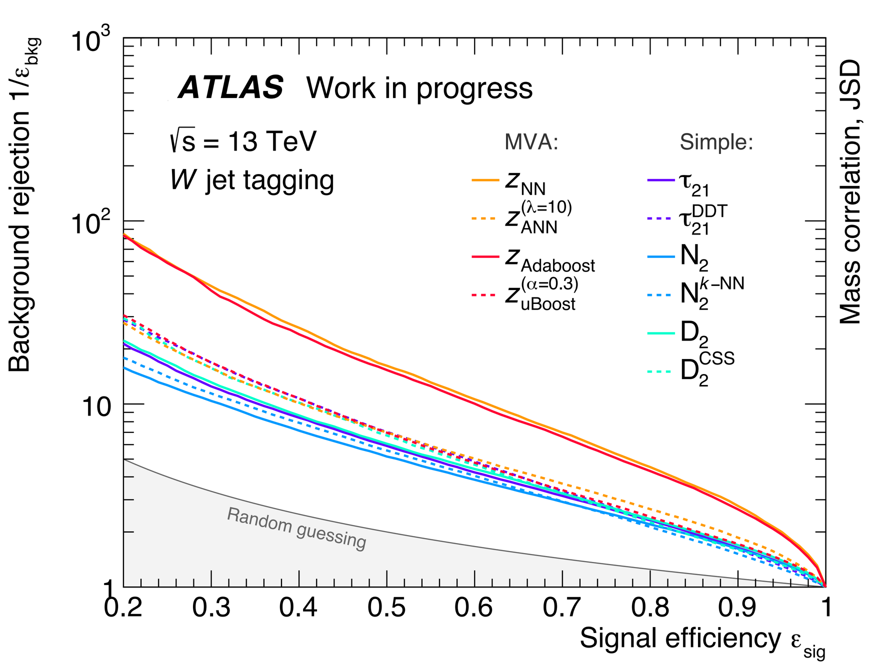
Jensen-Shannon divergence for discrete probability distributions P, Q was chosen as able to express highly non-linear correlations:

$$JSD(P||Q) = \frac{1}{2}\Big(-\sum_i P_i \log_n\Big(\frac{M_i}{P_i}\Big) - \sum_i Q_i \log_n\Big(\frac{M_i}{Q_i}\Big)\Big), \quad \text{with } M = \frac{P+Q}{2}$$

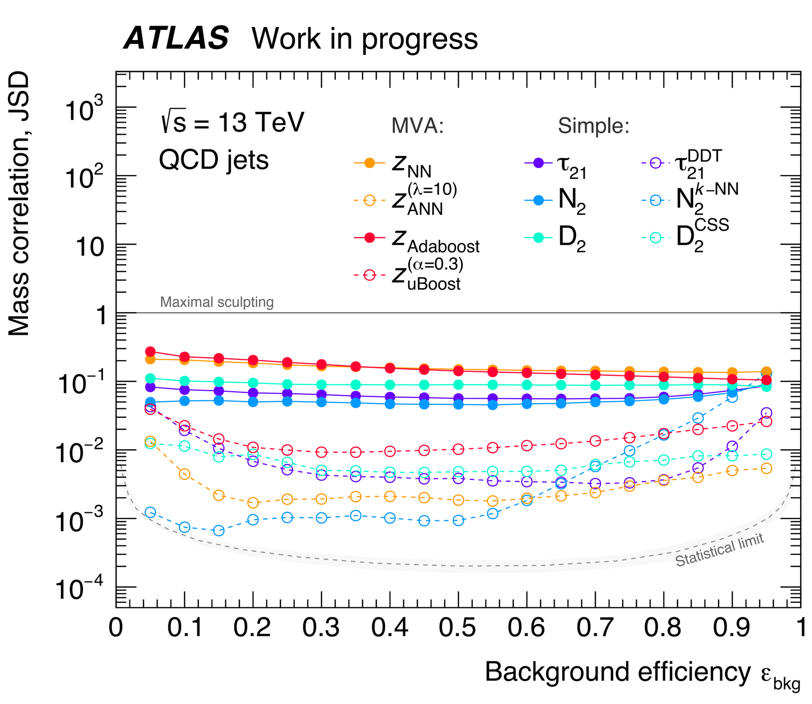
Where $P = 1/N_{bkg}^{pass}dN_{bkg}^{pass}/dm$ and $Q = 1/N_{bkg}^{fail}dN_{bkg}^{fail}/dm$

Performance Results

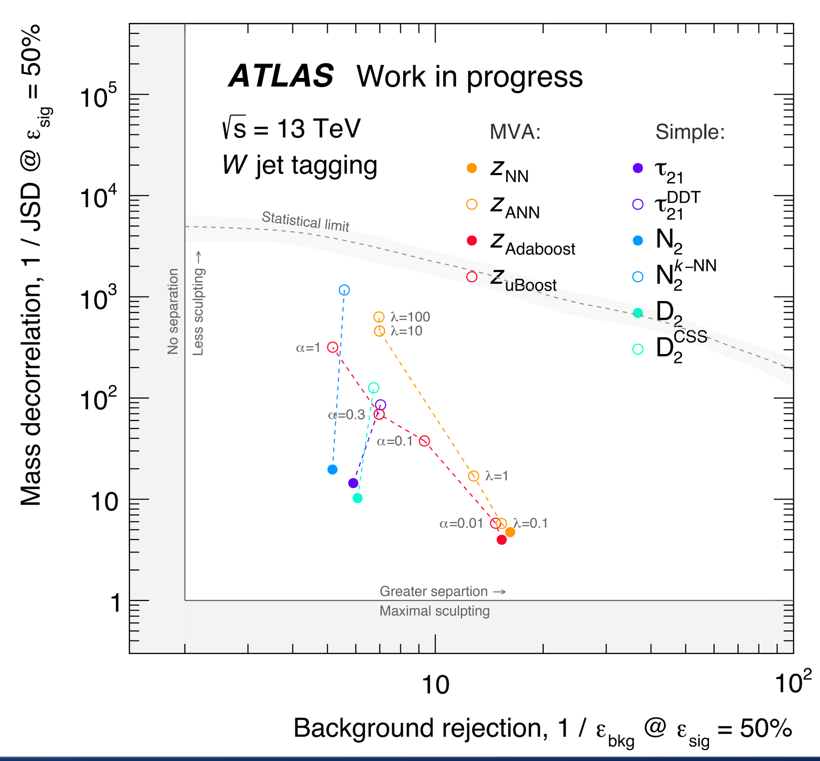
ROC



JSD



Combined



Conclusions

- Powerful learning algorithms for jet identification play a crucial role for new physics searches in HEP.
- Standard BDT and DNN techniques are strongly correlated with the jet mass, yielding a reduction of search sensitivity.
- ANN inspired by GANs model was described together with other decorrelated learning algorithms.
- ANN clearly outperforms standard taggers. Able to reach high classification power while maintaining mass decorrelation.

Gino Marceca (Universidad de Buenos Aires)