Machine Learning tools for Physics Searches at the LHC

Pietro Vischia[1]

[1] CP3 — IRMP, Université catholique de Louvain

UCLouvain
Institut de recherche
en mathématique et physique

A warm thanks to the organizers for the invitation!

Kolumbari, ICFNP2019

The purpose of this talk

UCLouvain
Institut de recherche
en mathématique et physique

- Machine learning techniques instrumental to the discovery of the Higgs boson in 2012
- Nowadays, new physics lies in the details
- Need for advanced techniques to extract “difficult” signals
  - Signal and Standard Model topologies might be very similar
  - We might not have a meaningful signal model (detect anomalies)
- Personal selection of ML topics, hopefully covering the main areas of application of ML at the LHC
  - Physics objects ID and reconstruction
  - Classification into signal and background events
  - Deal with unknown signals on top of very well known background
  - Data quality
  - Dedicated hardware for online monitoring
  - Future developments?

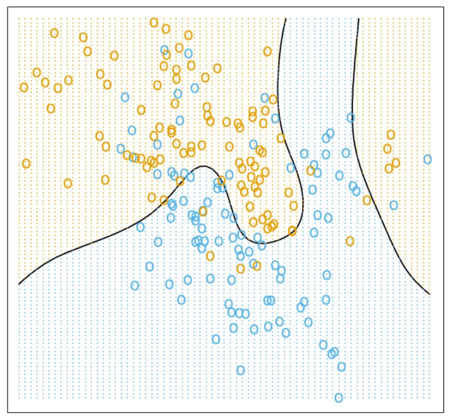
Machine learning: a black box?

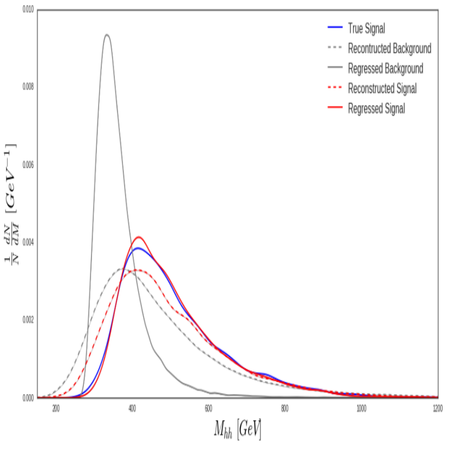
Image from Etsy listing: only 12 Australian dollars!

Brief overview of algorithms

General definition of ML

- *Vast amounts of data are being generated in many fields, and the statistician's job is to make sense of it all: to extract important patterns and trends, and understand "what the data says." We call this learning from data.* (Hastie, Tibshirani, Friedman, Springer2017)
    - Classification into classes
    - Regression of target quantities
- Well-defined mathematical problems
- Well-defined sanity checks





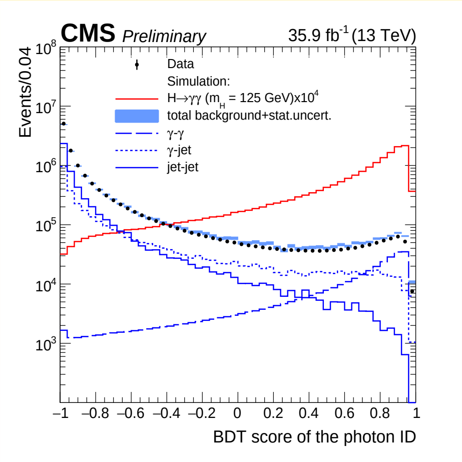
Figures from Hastie, Tibshirani, Friedman, Springer 2017, and from AMVA4NewPhysics deliverable 1.1 public report

Object ID

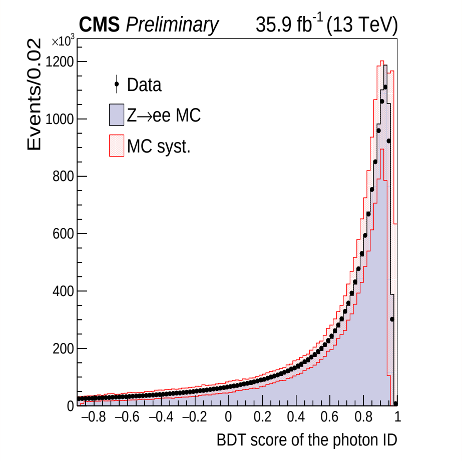
BDTs for object ID: the $\mathrm{H} \to \gamma\gamma$ case

- Object identification routinely done with ML techniques since the Higgs discovery
- Classification problem (e.g. true photon vs object misidentified as such)

γ identification score (lowest-scoring photon)



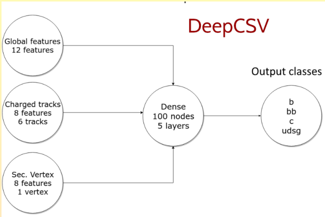
Validation in $\mathrm{Z} \to \mathrm{ee}$ events

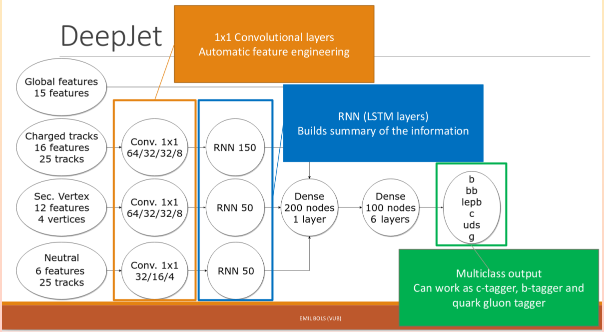


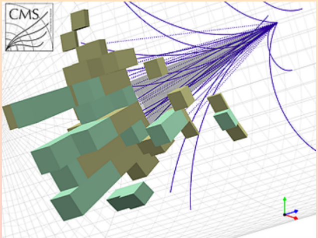
Plots from CMS-PAS-HIG-16-040

Object ID enters the era of mathematical representations

- b tagging at CMS
  - CSV (Run I and early Run II): BDT sensitive to secondary vertexes
- DeepCSV: similar inputs, generic DNN
- Domain knowledge can inform the representation used!
  - Leading criterion for choice of technique for the classifier
- What is the best representation for jets?
  - Convolutional networks for images
  - Particle-based structure



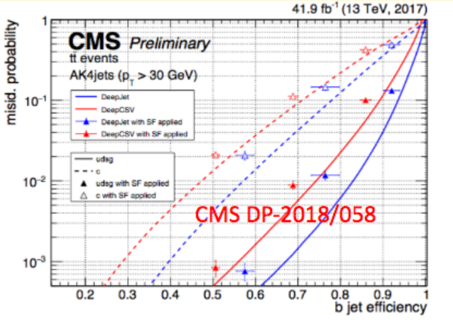


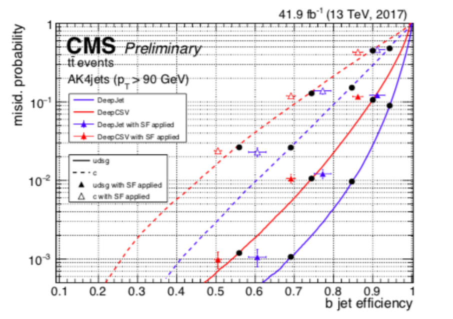


CMS DeepJet, plot from Emil Bols' talk at IML workshop

Object ID enters the era of mathematical representations

- Clear gains even with respect to generic DNN approach (DeepCSV)

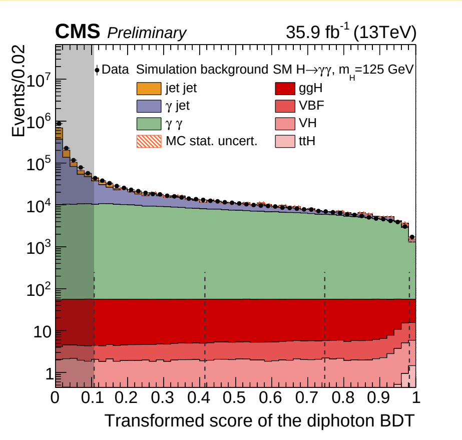


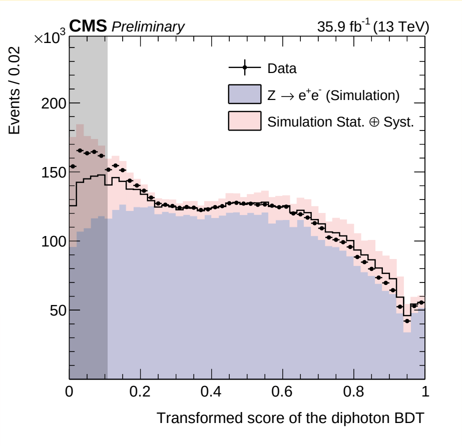


CMS DeepJet

Combining MVA IDs for object ID

- Dedicated per-event BDT classifier for the diphoton mass resolution
  - Photon ID BDT output used as input
  - High score to events with photons showing signal-like kinematics, good mass resolution, and high photon identification score
- Validation in $Z \to ee$ events where electrons are reconstructed as photons



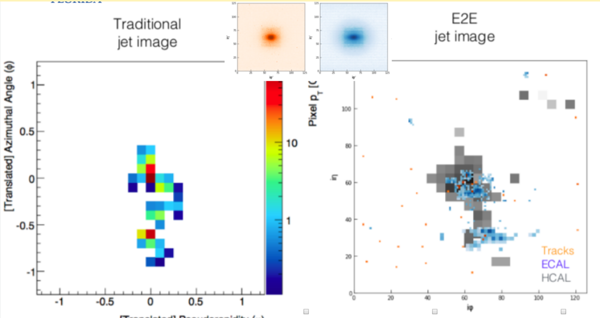


Plots from CMS-PAS-HIG-16-040

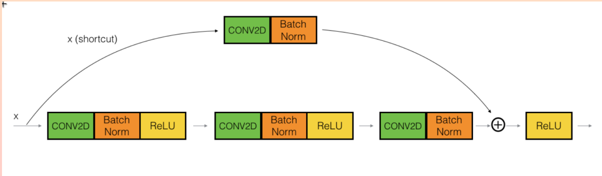
End-to-end jet reconstruction

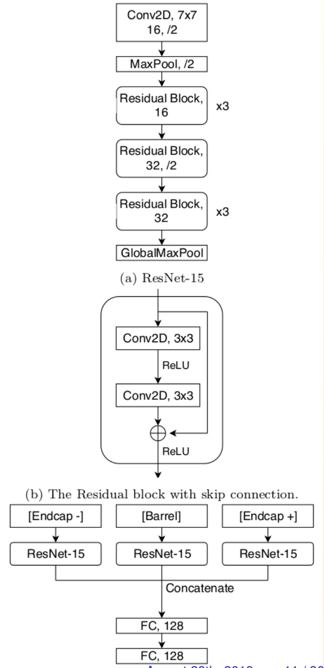
UCLouvain
Institut de recherche
en mathématique et physique

- Build images by projecting different layers into a single one
- Treat the result as an image with Res(idual)Net(works)
- Role of tracks in jet reco from network matches physics we know



Andrews et al., 2019





arXiv:1902.08276, S. Gleyzer's talk at 3rd IML workshop, Priya Dwivedi

Signal extraction

UCLouvain
Institut de recherche
en mathématique et physique

Separating signal from background: cuts

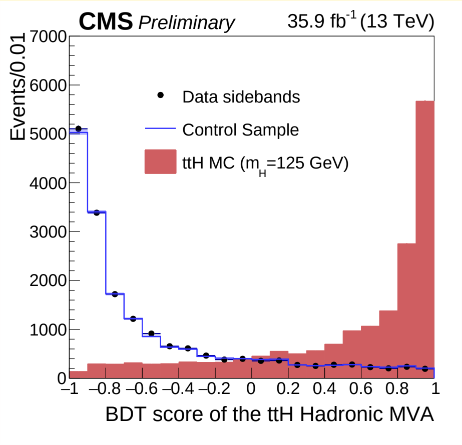
- Obtain high purity in $t\bar{t}H$ categories by removing events from the dataset
- Delicate: if using events by cutting on classifier output, dependence on training MC
  - Dangerous, e.g. prevents from using eventual unfolding results for BSM comparisons
- For both channels, events with low diphoton-BDT score are excluded
  - Threshold optimized jointly with $\gamma\gamma$-ID score: maximize expected precision on signal strength

$t\bar{t}H$ leptonic

- ≥ 1 light lepton
- ≥ 2 jets
- ≥ 1 btagged jet

$t\bar{t}H$ hadronic

- ≥ 3 jets
- ≥ 1 btagged jet
- 0 light leptons
- BDT classifier (input: N_{jets}, $p_{\mathrm{T}}^{leadjet}$, lead and sublead btag scores)

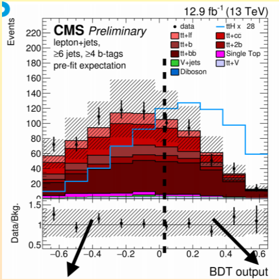


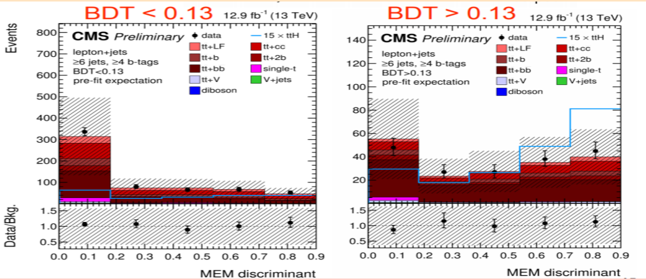
Evt Cat.	SM 125 GeV Higgs boson expected signal										
	Total	ggH	VBF	ttH	bbH	tHq	tHW	WH lep	ZH lep	WH had	ZH had
ttH Had.	5.85	10.99 %	0.70 %	77.54 %	2.02 %	4.13 %	2.02 %	0.09 %	0.05 %	0.63 %	1.82 %
ttH Lep.	3.81	1.90 %	0.05 %	87.48 %	0.08 %	4.73 %	3.04 %	1.53 %	1.15 %	0.02 %	0.02 %

Plots from CMS-PAS-HIG-16-040

Separating signal from background: full shape exploitation

- Increase sensitivity by retaining full classifier shape, plus further classification
  - Different signal/background fractions
  - Constrain background normalization or uncertainties in bkg-dominated regions



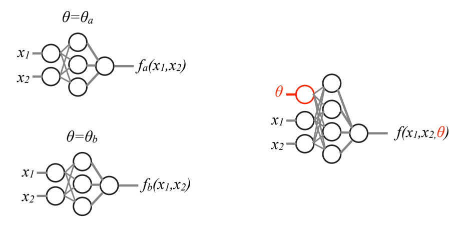


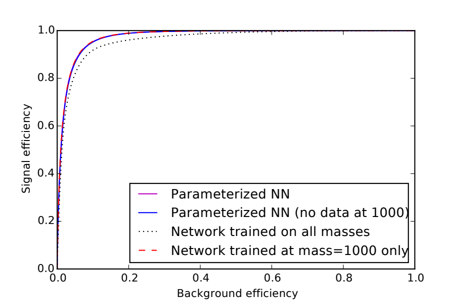
From $t\bar{t}H$ (bb), CMS−PAS−HIG−16−004

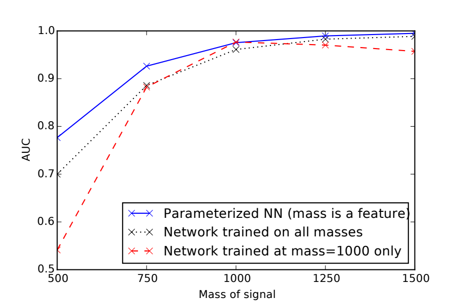
Unknown parameter? Parametrized machine learning helps you!

UCLouvain
Institut de recherche
en mathématique et physique

- Classifier sensitive to the value of the parameter
  - Train with true (sig) or random (bkg) value of parameter as input
  - Evaluate in slices at set parameter value
- Better than single training, and interpolates as well!
- Already used!
  - First CMS application: `CMS-HIG-17-006`
  - On the arxiv since yesterday: CMS-HIG-18-004, arXiv:1908.09206 ☺





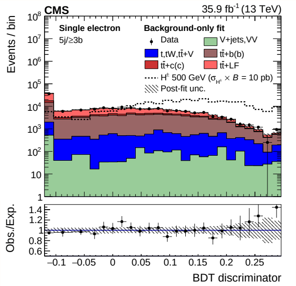


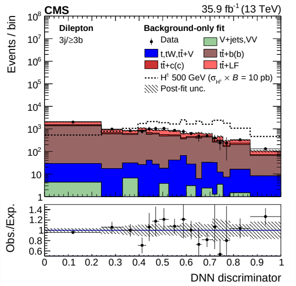
From Baldi *et al.* arXiv:1601.07913

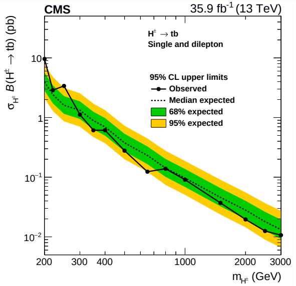
Different techniques good in different ways

UCLouvain
Institut de recherche
en mathématique et physique

- Each classification problem is a problem of its own
  - Choice of algorithm dictated by e.g. the inputs, the complexity of the problem (network capacity)
- Sometimes not trivial: CMS-HIG-18-004, arXiv:1908.09206 ☺
  - 20–40% improvement over H_T-based limits by using BDT (single lepton) and parameterized DNN (dilepton)
  - DNN: better sensitivity at low mass, where BDT not enough capacity to learn similar $\mathrm{t\bar{t}}$ vs $\mathrm{H}^{\pm}$ kinematics)



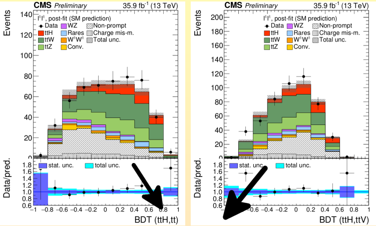




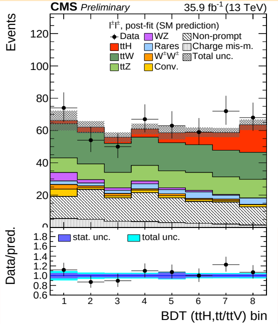
Advanced use cases: how many BDTs do you have?

- $t\bar{t}H$ multilepton: dedicated classifiers
  - BDT1: $t\bar{t}H$ vs $t\bar{t}$
  - BDT2: $t\bar{t}H$ vs $t\bar{t}V$
- Finely partition the 2D plane (BDT1, BDT2)
  - Use a training sample to compute binning
  - Apply to the application sample used for inference
- Define target N_{bins} with clustering techniques (k-means)
- Finally split regions based on empirical likelihood
  - Likelihood ratio approximated with $\frac{S}{B}$
  - Ordering from Neyman-Pearson lemma
  - Quantiles-based binning

BDT classifier output (2LSS)



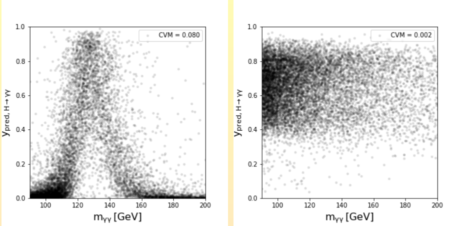
Final 1D discriminator (2LSS)

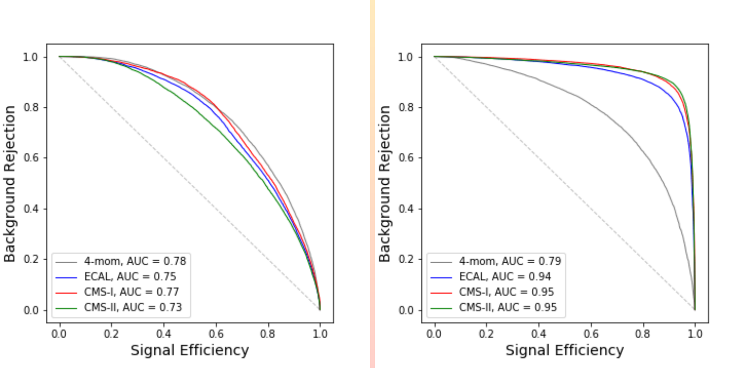


CMS-PAS-HIG-17-004, part of CMS-HIG-17-018: evidence for ttH production in multilepton final states

End-to-end event classification

- Low-level data representation
  - Tracker, ECAL, HCAL
  - Various geometries possible
- Mass decorrelation to avoid mass sculpting
  - Transform $E_{\gamma\gamma}$ to unit $M_{\gamma\gamma}$ in S and B
  - Extension of pivoting technique
- 3-class ResNet training:
  ($\mathrm{H} \to \gamma\gamma$, $\gamma\gamma$, γ+jet)
- Technique is statistically limited





From arXiv:1807.11916

What if you don't know your signal?

Gaussian processes)

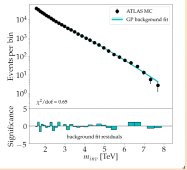
- MVA Gaussian associated to a set of random variables ($N_{dim} = N_{random\ variables}$)
  - *Kernel* as measure of similarity between bin centers (counts) and a *mean function*

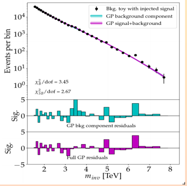
$$\mu(x) = 0\,, \tag{9}$$

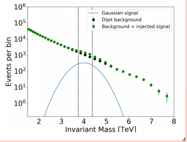
$$\Sigma_B(x,x') = A\ \exp\left(\frac{d-(x+x')}{2a}\right)\sqrt{\frac{2l(x)l(x')}{l(x)^2+l(x')^2}}\exp\left(\frac{-(x-x')^2}{l(x)^2+l(x')^2}\right), \tag{10}$$

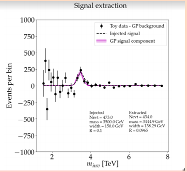
$$\Sigma_S(x,x') = C\ \exp\left(-\frac{1}{2}(x-x')^2/k^2\right)\exp\left(-\frac{1}{2}\left((x-m)^2+(x'-m)^2\right)/t^2\right), \tag{11}$$

-
  - Signal not parameterized
  - Hyperparameters fixed in B-only fit
- S: residual from B subtraction







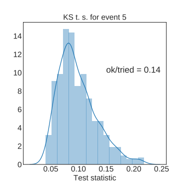


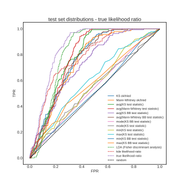
AMVA4NewPhysics deliverable 2.5 public report

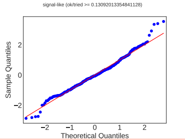
Inverse Bagging

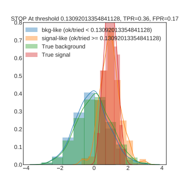
UCLouvain
Institut de recherche
en mathématique et physique

- Data: mixture model with small S
- Classification based on sample properties
  - Compare bootstrapped samples w/ pure-B reference
  - Use theorem by Metodiev to translate inference into signal fraction
- Benchmark with LR and LDA
  - Promising results







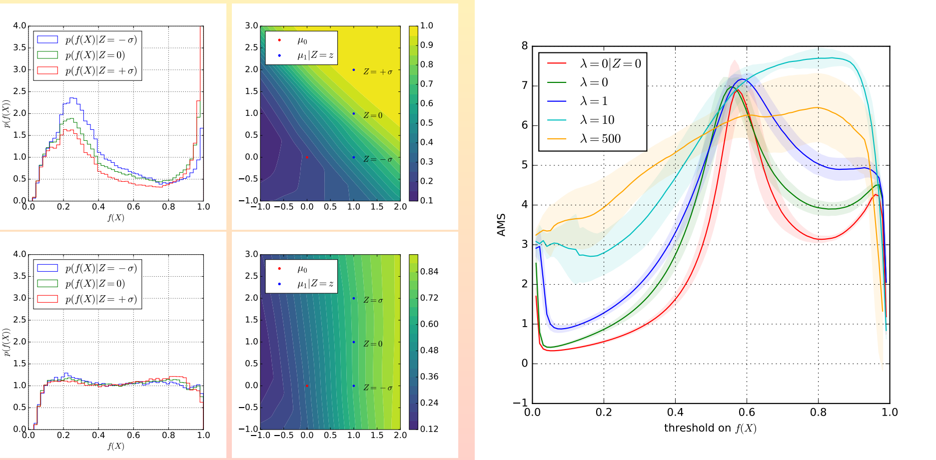


Vischia-Dorigo arXiv:1611.08256, doi:10.1051/epjconf/201713711009, and P. Vischia's talk at EMS2019

Deal with uncertainties

Can we reduce the impact of uncertainties on our result?

- Adversarial networks can be used to build pivot quantities
  - Quantities invariant in some parameter (typically nuisance parameter, e.g. pileup)
- Best Approximate Mean Significance as a tradeoff optimal/pivotal
  $E_\lambda(\theta_f,\theta_r) = \mathcal{L}_f(\theta_f) - \lambda \mathcal{L}_r(\theta_f,\theta_r)$

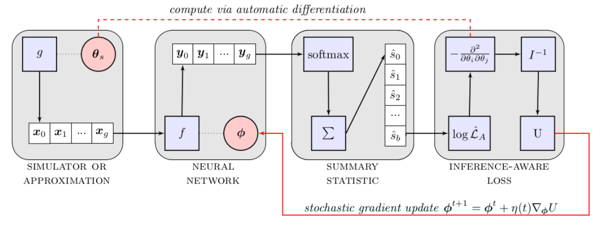


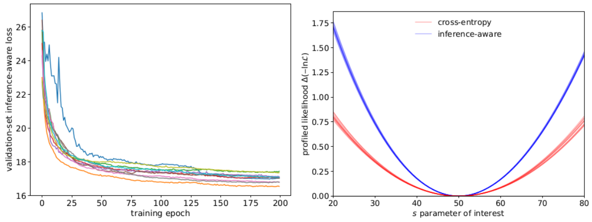
From Louppe-Kagan-Cranmer, arXiv:1611.01046

INFERNO: inference-aware optimization

UCLouvain
Institut de recherche
en mathématique et physique

- Build a non-parametric simulation-based likelihood to be used as summary statistic
- Minimize the expected variance of the parameters of interest
  - Obtain Fisher information matrix by automatic differentiation, and use it as loss function
  - For (asymptotically) unbiased estimators, Rao-Cramér-Frechet (RCF) bound $V[\hat{\theta}] \sim \frac{1}{\hat{\theta}}$ (see my lecture in this conference for details)
  - Constraints from auxiliary measurements (nuisance parameters) included in covariance matrix





(a) inference-aware training loss (b) profile-likelihood comparison

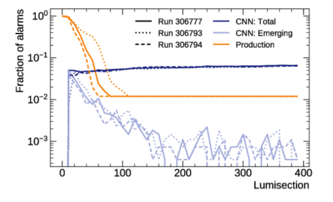
From De Castro-Dorigo, arXiv:1806.04743, and AMVA4NewPhysics deliverable 1.4 public report

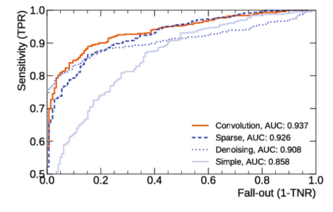
Which data should we take?

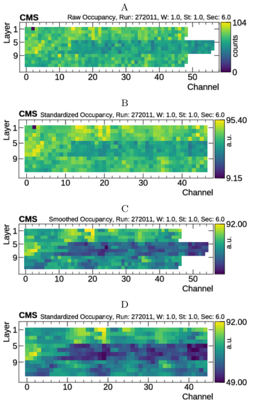
What if you don’t know which data to take?

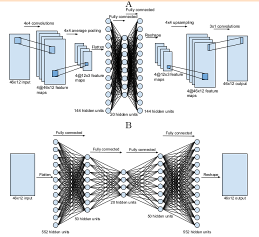
UCLouvain
Institut de recherche
en mathématique et physique

- Represent data as images organized geographically
  - Local approach: layers treated independently
  - Regional approach: layers treated independently but simultaneously (spot intra-chamber issues)
- Autoencoders (noise detection, dimensionality reduction)
  - Encode input to hidden layer
  - Decode hidden layer back to approximate representation of input





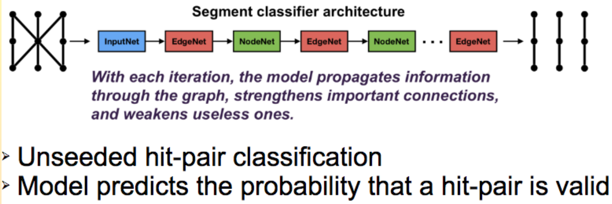


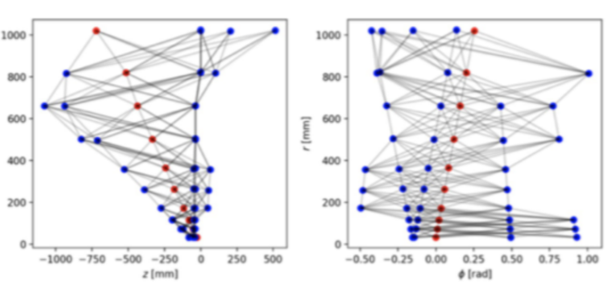


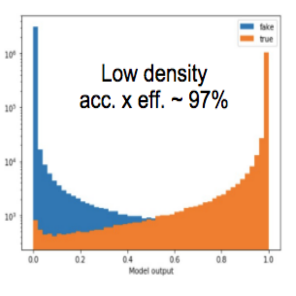
From arXiv:1808.00911

UCLouvain
Institut de recherche
en mathématique et physique

- Graph networks to literally connect the dots



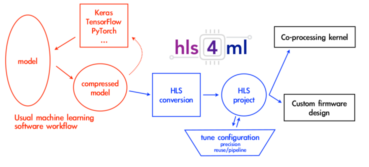


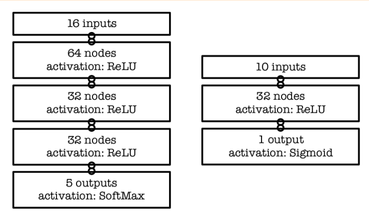


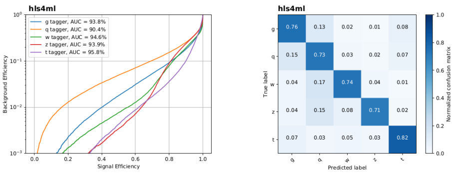
The HEP.TrkX project, S. Gleyzer's talk at 3rd IML workshop

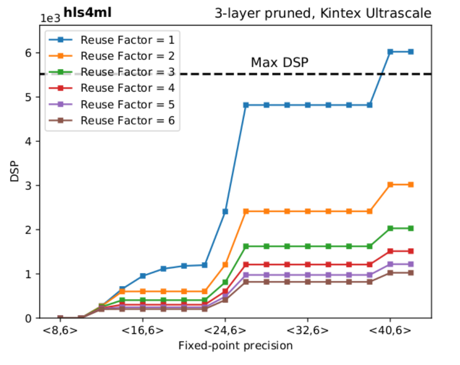
What if you need speed?

- Real-time event processing requires low-latency, low-power hardware: FPGAs
- Case study: jet substructure classification
- Compression, quantization, parallelization digital signal processing (arithmetic) blocks (DSPs),





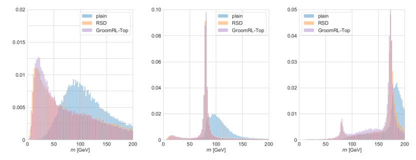




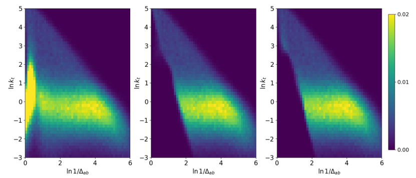
From arXiv:1804.06913

Next? Probably Deep Q learning (Reinforcement Learning)

- Boosted objects decay to collimated jets reconstructed as single fat jet
- Fat jet grooming: remove soft wide-angle radiation not associated with the underlying hard substructure



(a) QCD (b) W (c) top



(a) Plain (b) GroomRL-W (c) GroomRL-Top

Images from arXiv:1903.09644 and The Auckland Dog Coach

Summary...

- Focused on searches
- Many applications of machine learning!
  - Object identification/reconstruction
  - Event classification
  - Regression of physical observables
  - Reduce impact of uncertainties
  - End-to-end reconstruction
  - Data acquisition
  - Data quality monitoring
  - Online ML with dedicated hardware
  - Reinforcement learning: the next step?
- Many validation techniques carefully employed
  - Results match the physics we know or the physics we plug in
- Much more to come in the future!!!
- In the meantime I hope I have conviced you that...

Image from Amazon's website

Machine learning: a black box? Not really!

Image from Etsy listing: only 12 Australian dollars! (content not included)

THANK YOU FOR YOUR ATTENTION!!

Backup