Learned Uncertainties in LHC Physics

Tilman Plehn

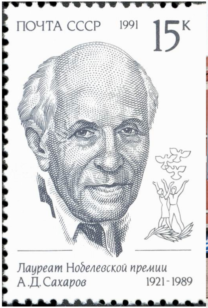
Universität Heidelberg

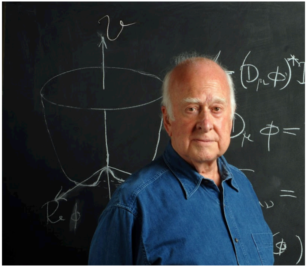
DeepChip, November 2024

Machine learning in modern LHC physics

Classic motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?





Machine learning in modern LHC physics

Classic motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

LHC physics

- fundamental questions
- huge data set
- first-principle, precision simulations
- complete uncertainty control

Successful past

- measurements of total rates
- analyses inspired by simulation
- model-driven Higgs discovery

Machine learning in modern LHC physics

Classic motivation

- dark matter?
- matter vs antimatter?
- origin of Higgs boson?

LHC physics

- fundamental questions
- huge data set
- first-principle, precision simulations
- complete uncertainty control

Successful past

- measurements of total rates
- analyses inspired by simulation
- model-driven Higgs discovery

First-principle, precision simulations

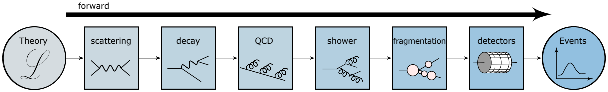
- start with Lagrangian
- calculate scattering using QFT
- simulate collisions
- simulate detectors

→ LHC collisions in virtual worlds

BSM searches

- compare simulations and data
- infer underlying theory [SM or BSM]
- publish useable results

→ understand LHC data systematically



Learned uncertainties

LHC applications

- accuracy & reliability & control
- regression: amplitudes, calibration,...
  classification: object identification, tagging, event searches...
  generation: integration, simulation, unfolding...

$\rightarrow$ Uncertainties as test and analysis input

Learned uncertainties

LHC applications

- accuracy & reliability & control
- regression: amplitudes, calibration,...
  classification: object identification, tagging, event searches...
  generation: integration, simulation, unfolding...

→ Uncertainties as test and analysis input

Remember a fit

- learn underlying function $f_\theta(x) \approx f(x)$
- maximize parameter probability given $(f, \sigma)_j$ [Gaussian]

$$p(\theta|x) = \frac{p(x|\theta)\; p(\theta)}{p(x)}$$

$$p(x|\theta) = \prod_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{|f_j - f_\theta(x_j)|^2}{2\sigma_j^2}\right)$$

$$\Rightarrow \qquad \mathcal{L} \equiv -\log p(x|\theta) = \sum_j \frac{|f_j - f_\theta(x_j)|^2}{2\sigma_j^2} + \text{const}(\theta)$$

→ parameterized function from Gaussian likelihood maximization

Learned uncertainties

LHC applications

- accuracy & reliability & control
- regression: amplitudes, calibration,...
  classification: object identification, tagging, event searches...
  generation: integration, simulation, unfolding...

$\to$ Uncertainties as test and analysis input

Many parameters θ, unknown σ

- minimize Gaussian log-likelihood loss

$$\mathcal{L}_{\text{het}} = \frac{|f(x) - f_\theta(x)|^2}{2\sigma_\theta(x)^2} + \log \sigma_\theta(x) + \cdots$$

- double network
  learning $f_\theta(x)$ and $\sigma_\theta(x)$
- generalization: $\sigma_\theta(x)$ from GMM
- LHC: phase space interpretable

$\to$ Local uncertainty learned

Regression with uncertainties

Detector calibration, transition amplitude, jet energy...

- energy measurement for cluster/jet j

$$\langle E \rangle = \int dE \; E \; p(E)$$

- weighted by reproducing training data $p(\theta|T)$

$$p(E) = \int d\theta \; p(E|\theta) \; p(\theta|T)$$

$\to$ θ-distributions defining Bayesian NN

Regression with uncertainties

Detector calibration, transition amplitude, jet energy...

- energy measurement for cluster/jet j

$$\langle E \rangle = \int dE \; E \; p(E)$$

- weighted by reproducing training data $p(\theta|T)$

$$p(E) = \int d\theta \; p(E|\theta) \; p(\theta|T)$$

$\rightarrow$ θ-distributions defining Bayesian NN

Variational approximation

- definition of training [think $q(\theta)$ as Gaussian with mean and width]

$$p(E) = \int d\theta \; p(E|\theta) \; p(\theta|T) \approx \int d\theta \; p(E|\theta) \; q(\theta)$$

- similarity through minimal KL-divergence [Bayes' theorem to remove unknown posterior]

$$\begin{aligned} D_{\text{KL}}[q(\theta), p(\theta|T)] &= \int d\theta \; q(\theta) \; \log \frac{q(\theta)}{p(\theta|T)} \\ &= \int d\theta \; q(\theta) \; \log \frac{q(\theta) p(T)}{p(T|\theta) p(\theta)} \\ &= D_{\text{KL}}[q(\theta), p(\theta)] - \int d\theta \; q(\theta) \; \log p(T|\theta) + \log p(T) \int d\theta \; q(\theta) \end{aligned}$$

Regression with uncertainties

Detector calibration, transition amplitude, jet energy...

- energy measurement for cluster/jet j

$$\langle E \rangle = \int dE \; E \; p(E)$$

- weighted by reproducing training data $p(\theta|T)$

$$p(E) = \int d\theta \; p(E|\theta) \; p(\theta|T)$$

$\rightarrow$ θ-distributions defining Bayesian NN

Variational approximation

- definition of training [think $q(\theta)$ as Gaussian with mean and width]

$$p(E) = \int d\theta \; p(E|\theta) \; p(\theta|T) \approx \int d\theta \; p(E|\theta) \; q(\theta)$$

- similarity through minimal KL-divergence [Bayes' theorem to remove unknown posterior]

$$\begin{aligned} D_\text{KL}[q(\theta), p(\theta|T)] &= \int d\theta \; q(\theta) \; \log \frac{q(\theta)}{p(\theta|T)} \\ &= \int d\theta \; q(\theta) \; \log \frac{q(\theta) p(T)}{p(T|\theta) p(\theta)} \\ &\approx D_\text{KL}[q(\theta), p(\theta)] - \int d\theta \; q(\theta) \; \log p(T|\theta) \equiv \mathcal{L} \end{aligned}$$

$\rightarrow$ Two-term loss: likelihood + prior

Relation to deterministic networks

Regularization

- BNN loss

$$\mathcal{L} = -\int d\theta \; q(\theta) \; \log p(T|\theta) + D_\text{KL}[q(\theta), p(\theta)]$$
$$= -\int d\theta \; q(\theta) \; \log p(T|\theta) + \frac{\sigma_q^2 - \sigma_p^2 + (\mu_q - \mu_p)^2}{2\sigma_p^2} + \dots$$

- deterministic network

$$q(\theta) = \delta(\theta - \theta_0) \quad \Rightarrow \quad \mathcal{L} \approx -\log p(T|\theta_0) + \frac{(\theta_0 - \mu_p)^2}{2\sigma_p^2}$$

$\rightarrow$ Likelihood with L2-regularization

Relation to deterministic networks

Regularization

- BNN loss

$$\mathcal{L} = - \int d\theta \; q(\theta) \; \log p(T|\theta) + D_\text{KL}[q(\theta), p(\theta)]$$
$$= - \int d\theta \; q(\theta) \; \log p(T|\theta) + \frac{\sigma_q^2 - \sigma_p^2 + (\mu_q - \mu_p)^2}{2\sigma_p^2} + \dots$$

- deterministic network

$$q(\theta) = \delta(\theta - \theta_0) \quad \Rightarrow \quad \mathcal{L} \approx - \log p(T|\theta_0) + \frac{(\theta_0 - \mu_p)^2}{2\sigma_p^2}$$

$\to$ Likelihood with L2-regularization

Dropout

- Bernoulli weights

$$q(\theta) \to q(x) = \rho^x (1 - \rho)^{1-x} \Big|_{x=0,1} \quad \text{with} \quad \theta = x\theta_0$$

- Regularized likelihood with dropout

$\to$ Easy to train to high precision

Statistics vs systematics

Network evaluation

- expectation value using trained network $q(\theta)$

$$\langle E \rangle = \int dE d\theta \; E \; p(E|\theta) \; q(\theta)$$
$$\equiv \int d\theta \; q(\theta) \overline{E}(\theta) \qquad \text{with} \qquad \overline{E}(\theta) = \int dE \; E \; p(E|\theta)$$

- corresponding variance

$$\sigma_{\text{tot}}^2 = \int dE d\theta \; \left(E - \langle E \rangle\right)^2 \; p(E|\theta) \; q(\theta)$$
$$= \int d\theta \; q(\theta) \left[\overline{E^2}(\theta) - 2\langle E \rangle \overline{E}(\theta) + \langle E \rangle^2 \right]$$
$$= \int d\theta \; q(\theta) \left[\overline{E^2}(\theta) - \overline{E}(\theta)^2 + \left(\overline{E}(\theta) - \langle E \rangle \right)^2 \right] \equiv \sigma_{\text{syst}}^2 + \sigma_{\text{stat}}^2$$

Two uncertainty classes [LHC version of aleatoric/epistemic]

- systematic — vanishing for perfect network and data: $p(E|\theta) \to \delta(E - E_0)$

$$\sigma_{\text{syst}}^2 = \int d\theta \; q(\theta) \left[\overline{E^2}(\theta) - \overline{E}(\theta)^2 \right] \equiv \sigma_{\text{het}}^2$$

- statistical — vanishing for perfect training: $q(\theta) \to \delta(\theta - \theta_0)$ [see repulsive ensembles]

$\to$ LHC: systematics the problem, always!

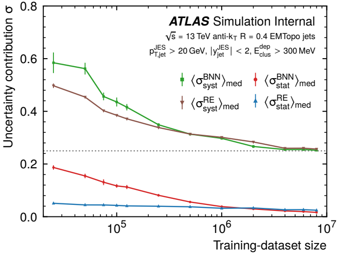
LHC Experiment

Calibration with uncertainties [Heidelberg-ML + ATLAS]

- interpretable topo-cluster phase space x
- learned calibration

$$\mathcal{R}^{\text{BNN}}(x) \approx \mathcal{R}(x) = \frac{E^{\text{EM}}(x)}{E^{\text{dep}}(x)} \qquad \text{with} \qquad \Delta\mathcal{R}^{\text{BNN}}(x)$$

- trained on simulations, statistics neglibigle
- systematics: noise in data
  network expressivity
  data representation ...



LHC Experiment

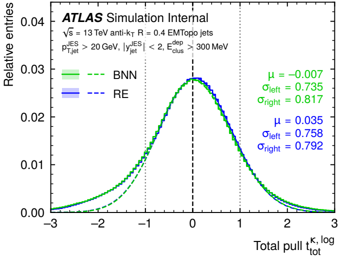
Calibration with uncertainties [Heidelberg-ML + ATLAS]

- interpretable topo-cluster phase space x
- learned calibration

$$\mathcal{R}^{\text{BNN}}(x) \approx \mathcal{R}(x) = \frac{E^{\text{EM}}(x)}{E^{\text{dep}}(x)} \qquad \text{with} \qquad \Delta\mathcal{R}^{\text{BNN}}(x)$$

- trained on simulations, statistics neglibigle
- systematics: noise in data
  network expressivity
  data representation ...

$\rightarrow$ calibration from pull



LHC Experiment

Calibration with uncertainties [Heidelberg-ML + ATLAS]

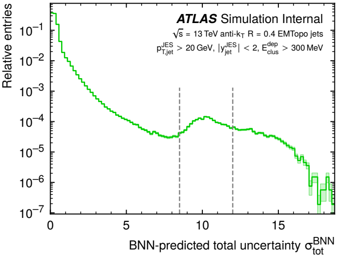
- interpretable topo-cluster phase space x
- learned calibration

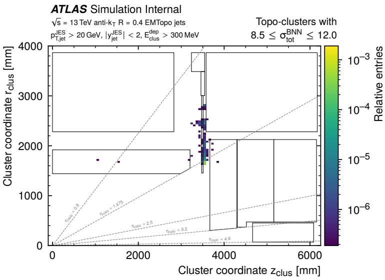
$$\mathcal{R}^{\text{BNN}}(x) \approx \mathcal{R}(x) = \frac{E^{\text{EM}}(x)}{E^{\text{dep}}(x)} \qquad \text{with} \qquad \Delta\mathcal{R}^{\text{BNN}}(x)$$

- trained on simulations, statistics neglibigle
- systematics: noise in data
  network expressivity
  data representation ...

→ calibration from pull

→ Understand data using uncertainties!



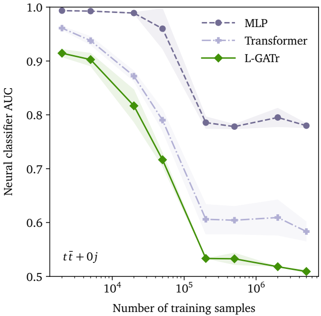


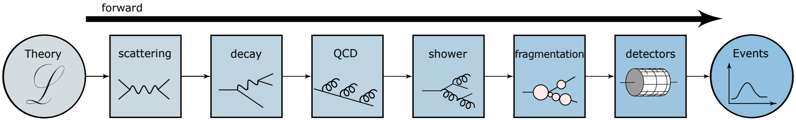
Generative AI

Forward simulations

- learn phase space density
  sample Gaussian $\to$ phase space
- Variational Autoencoder
  $\to$ low-dimensional physics
- Generative Adversarial Network
  $\to$ generator trained by classifier
- Normalizing Flow/Diffusion
  $\to$ (bijective) mapping [INN]
- JetGPT, ViT
  $\to$ non-local structures
- Equivariant L-GATr
  $\to$ guarantee Lorentz symmetry

$\to$ Combinations: equivariant transformer CFM...





Generative Uncertainties

Unsupervised Bayesian networks

- data: event sample [points in 2D space]

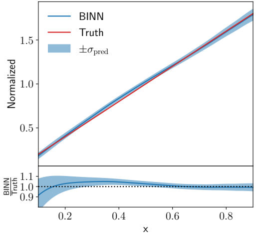
  learn phase space density
  sample from latent space
- Bayesian version

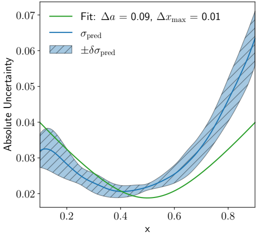
  allow weight distributions
  learn uncertainty map
- 2D wedge ramp

$$p(x) = ax + b = ax + \frac{1 - \frac{a}{2}(x_{\text{max}}^2 - x_{\text{min}}^2)}{x_{\text{max}} - x_{\text{min}}}$$

$$(\Delta p)^2 = \left(x - \frac{1}{2}\right)^2 (\Delta a)^2 + \left(1 + \frac{a}{2}\right)^2 (\Delta x_{\text{max}})^2 + \left(1 - \frac{a}{2}\right)^2 (\Delta x_{\text{min}})^2$$

explaining minimum in $\sigma(x)$





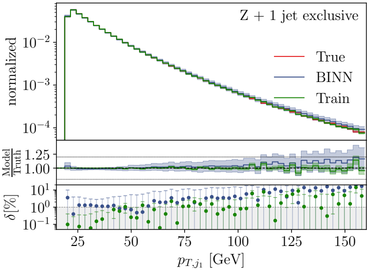
→ INNs just fit, CFM less stiff, transformer patch-wise...

Events with uncertainties

Bayesian network generator

- network with weight distributions
- weights with uncertainty

$\Rightarrow$ Training-related error bars



Events with uncertainties

Bayesian network generator

- network with weight distributions
- weights with uncertainty

$\Rightarrow$ Training-related error bars

Control classifier

- train classifier training vs generated
- quantify accuracy
  identify failure modes
  reweight generator

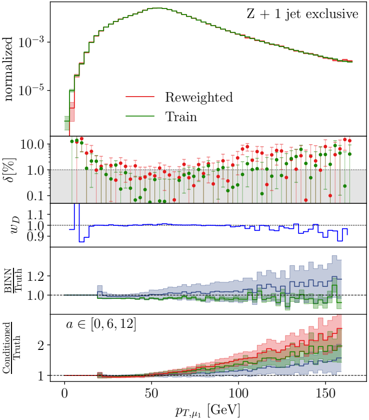
Theory uncertainties

- systematics from data augmentation

$$w = 1 + a \left(\frac{p_{T,j_1} - 15 \ \text{GeV}}{100 \ \text{GeV}} \right)^2$$

- train conditionally on a $[a = 0 \ldots 30]$
  error bar from sampling a

$\Rightarrow$ Systematic/theory error bars



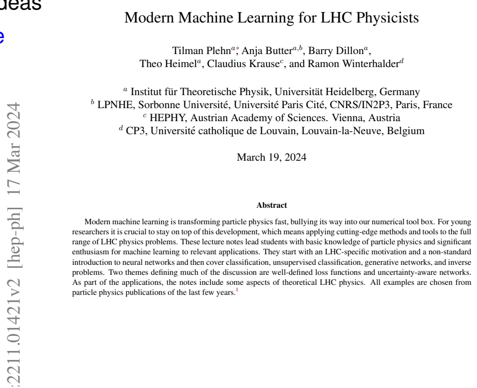
Modern ML for LHC

ML for the best science

- old people: just another numerical tool for a numerical field
- young people: transformative new common language
- uncertainties crucial for physics
- be 10000 Einsteins,

  ...improving established tools
  ...developing new tools for established tasks
  ...transforming through new ideas

→ It's the future, also of Science



:2211.01421v2 [hep-ph] 17 Mar 2024

Modern Machine Learning for LHC Physicists

Tilman Plehn[a*], Anja Butter[a,b], Barry Dillon[a],
Theo Heimel[a], Claudius Krause[c], and Ramon Winterhalder[d]

[a] Institut für Theoretische Physik, Universität Heidelberg, Germany
[b] LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France
[c] HEPHY, Austrian Academy of Sciences. Vienna, Austria
[d] CP3, Université catholique de Louvain, Louvain-la-Neuve, Belgium

March 19, 2024

Abstract

Modern machine learning is transforming particle physics fast, bullying its way into our numerical tool box. For young researchers it is crucial to stay on top of this development, which means applying cutting-edge methods and tools to the full range of LHC physics problems. These lecture notes lead students with basic knowledge of particle physics and significant enthusiasm for machine learning to relevant applications. They start with an LHC-specific motivation and a non-standard introduction to neural networks and then cover classification, unsupervised classification, generative networks, and inverse problems. Two themes defining much of the discussion are well-defined loss functions and uncertainty-aware networks. As part of the applications, the notes include some aspects of theoretical LHC physics. All examples are chosen from particle physics publications of the last few years.[1]