DNP 2025
Fall Meeting of the APS
Division of Nuclear Physics
Chicago, IL | Oct. 17-20, 2025

Omnifold uncertainties

How statistical uncertainties propagate through unbinned unfolding

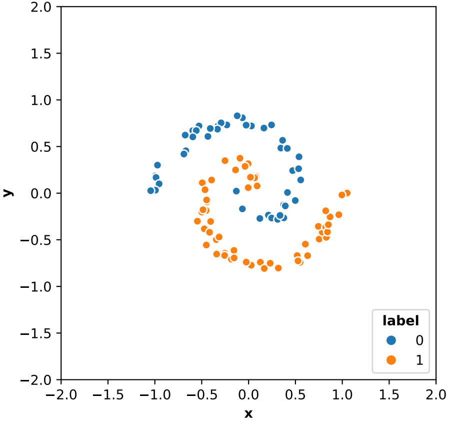
Tanmay Pani

Rutgers University

2025-10-14

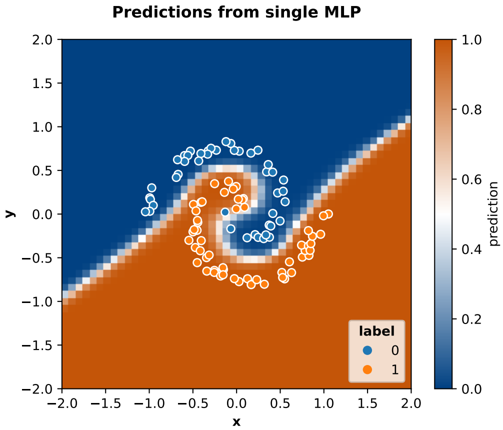
U.S. DEPARTMENT OF ENERGY
Office of Science

Multi-Layer Perceptron (MLP) - Toy Example

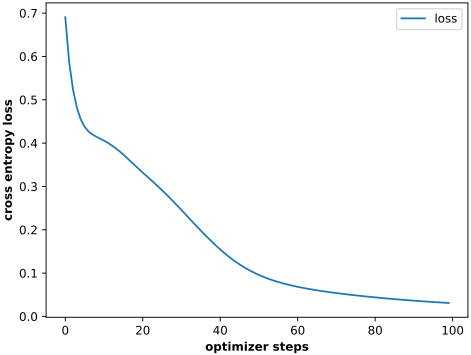


Dataset: 100 samples from 2 gaussian-smeared spirals, “0” (upper spiral / blue points) and “1” (lower spiral / orange points)
Task: classify the blue and orange points

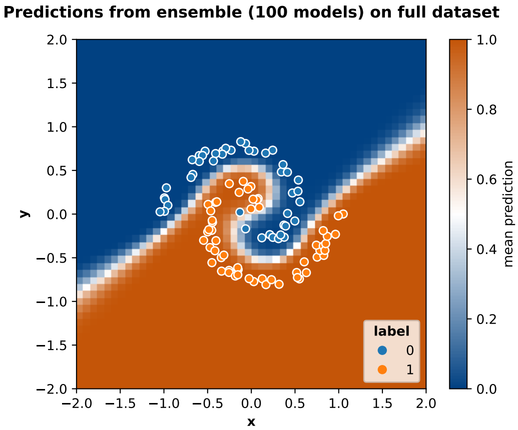
Multi-Layer Perceptron (MLP) - Toy Example



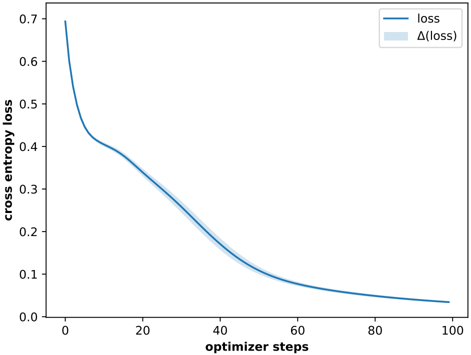
Attempt 1: A single Multi-Layer Perceptron (MLP)



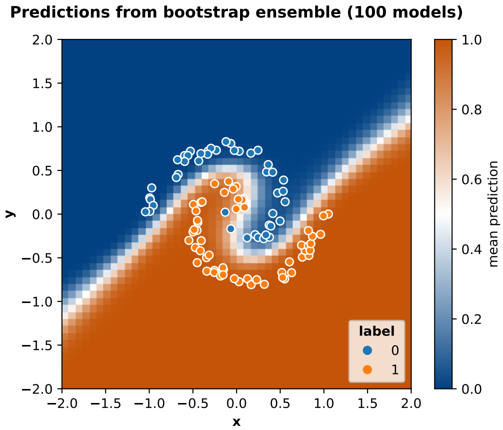
Multi-Layer Perceptron (MLP) - Toy Example



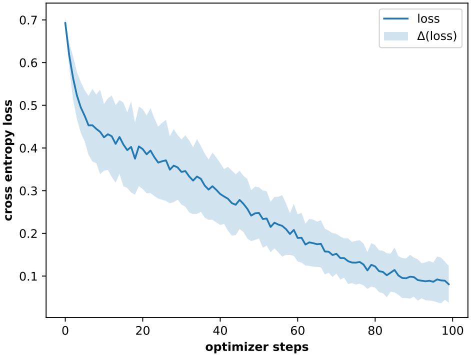
Attempt 2: 100 MLPs, each initialized differently, but sees the whole dataset



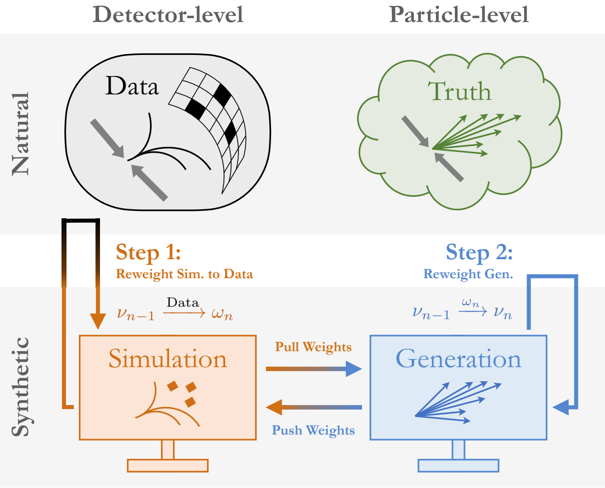
Multi-Layer Perceptron (MLP) - Toy Example



Attempt 3: 100 MLPs, each sees different bootstrap sampling of the dataset

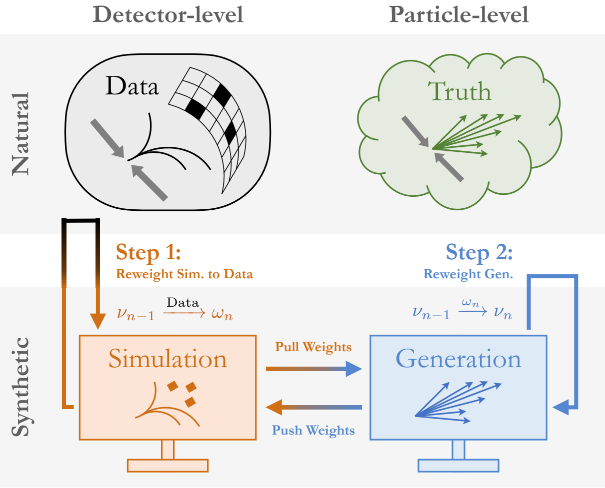


Omnifold : Basics



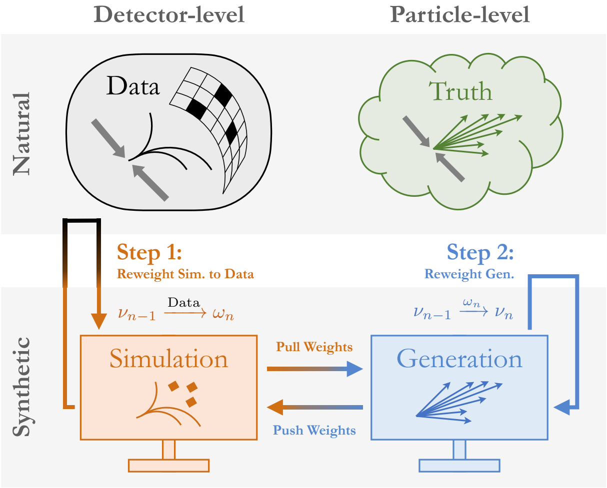
- Iterative unfolding using unbinned data

Omnifold : Basics



- Iterative unfolding using unbinned data
- Steps 1 and 2 each uses MLP classifier to calculate per-sample conditional probabilities

Omnifold : Basics



- Iterative unfolding using unbinned data
- Steps 1 and 2 each uses MLP classifier to calculate per-sample conditional probabilities
- Step 1 reweights reco to data and Step 2 propagates them to the gen

Generalized angularities

$$\lambda_\beta^\kappa = \sum_{\text{const}\in\text{jet}} \overbrace{\left(\frac{p_{\text{T,const}}}{p_{\text{T,jet}}}\right)^\kappa}^{\text{soft/hard radiation}} \times \overbrace{r(\text{const},\text{jet})^\beta}^{\text{collinearity sensitive}}$$

$$r(\text{const},\text{jet}) = \sqrt{(\eta_{\text{jet}} - \eta_{\text{const}})^2 + (\phi_{\text{jet}} - \phi_{\text{const}})^2}$$

- **LHA angularity** $\lambda_{0.5}^1 = \frac{\sum_{\text{trk}\in\text{jet}} p_{\text{T,trk}}\sqrt{\Delta R}}{p_{\text{T,jet}}}$,
- **Jet girth:** $\text{g} = \lambda_1^1 = \frac{\sum_{\text{trk}\in\text{jet}} p_{\text{T,trk}}\Delta R}{p_{\text{T,jet}}}$, measure of jet broadening
- **Thrust** $\lambda_2^1 = \frac{\sum_{\text{trk}\in\text{jet}} p_{\text{T,trk}}\Delta R^2}{p_{\text{T,jet}}}$, related to jet mass
- **Momentum dispersion :** $p_T^D = \lambda_0^2$
- Done for both inclusive jets and hard-core component of the inclusive jets
- Hard-core component calculated by vector summing constituents with $p_T > 2.0$ GeV/c

Dataset and Simulations

- **System:** p+p (2012) @ $\sqrt{s_{\mathrm{NN}}} = 200$GeV

Dataset and Simulations

- **System:** p+p (2012) @ $\sqrt{s_{\rm NN}} = 200$GeV
- **High Tower (HT) triggered** events ($\exists$ tower with $E_{\rm tower} \geqslant 4$ GeV) to enhance jet signal

Dataset and Simulations

- **System:** p+p (2012) @ $\sqrt{s_{\mathrm{NN}}} = 200$GeV
- **High Tower (HT) triggered** events ($\exists$ tower with $E_{\mathrm{tower}} \geqslant 4$ GeV) to enhance jet signal
- **Embedding simulation:**

Dataset and Simulations

- **System:** p+p (2012) @ $\sqrt{s_{\mathrm{NN}}} = 200$GeV
- **High Tower (HT) triggered** events ($\exists$ tower with $E_{\mathrm{tower}} \geqslant 4$ GeV) to enhance jet signal
- **Embedding simulation:**
- GEN: PYTHIA-6 Perugia-STAR dijet events (J. K. Adkins, PhD thesis (Kentucky U., 2015))
- RECO: PYTHIA-6 Perugia-STAR + GEANT3 + STAR p+p Run12 ZeroBias (2023 version)

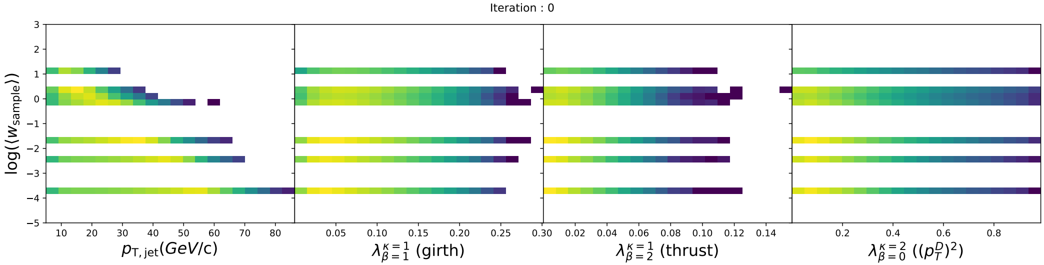
Dataset and Simulations

- **System:** p+p (2012) @ $\sqrt{s_{\mathrm{NN}}} = 200$GeV
- **High Tower (HT) triggered** events ($\exists$ tower with $E_{\mathrm{tower}} \geqslant 4$ GeV) to enhance jet signal
- **Embedding simulation:**
- GEN: PYTHIA-6 Perugia-STAR dijet events (J. K. Adkins, PhD thesis (Kentucky U., 2015))
- RECO: PYTHIA-6 Perugia-STAR + GEANT3 + STAR p+p Run12 ZeroBias (2023 version)
- Bootstraped ensembles consists of 10 parallel models and each model sees a bootstrap subsample consisting of 500000 data jets and 500000 reco-jets

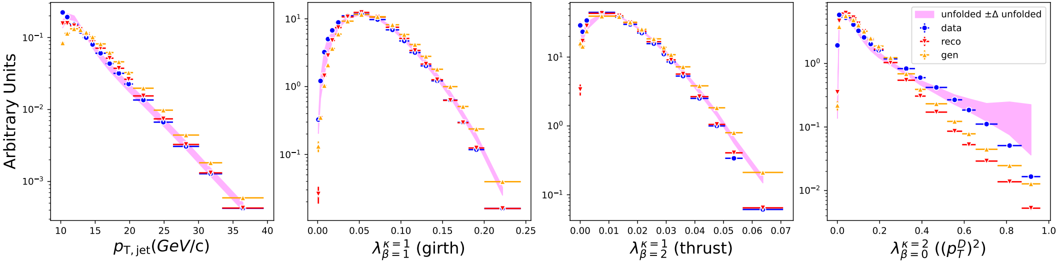
Dataset and Simulations

- **System:** $p{+}p$ (2012) @ $\sqrt{s_{\mathrm{NN}}} = 200$GeV
- **High Tower (HT) triggered** events ($\exists$ tower with $E_{\mathrm{tower}} \geqslant 4$ GeV) to enhance jet signal
- **Embedding simulation:**
- GEN: PYTHIA-6 Perugia-STAR dijet events (J. K. Adkins, PhD thesis (Kentucky U., 2015))
- RECO: PYTHIA-6 Perugia-STAR + GEANT3 + STAR $p{+}p$ Run12 ZeroBias (2023 version)
- Bootstraped ensembles consists of 10 parallel models and each model sees a bootstrap subsample consisting of 500000 data jets and 500000 reco-jets
- Each model runs for 10 unfolding iterations

Bootstrapped Omnifolding : Mean Model Prediction



Bootstrapped Omnifolding : Unfolded Histograms



Conclusions and Outlook

- Bootstraped ensembling results in better predictions which depend on phase space distribution of the features
- Omnifolded weights from bootstraped ensemble results in smoothing of sample-weights distributions
- Unfolded distributions from bootstraped omnifold show uncertainties that increase with decrease in data statistics
- Need to run bigger ensembles to get better convergence of histograms with iterations