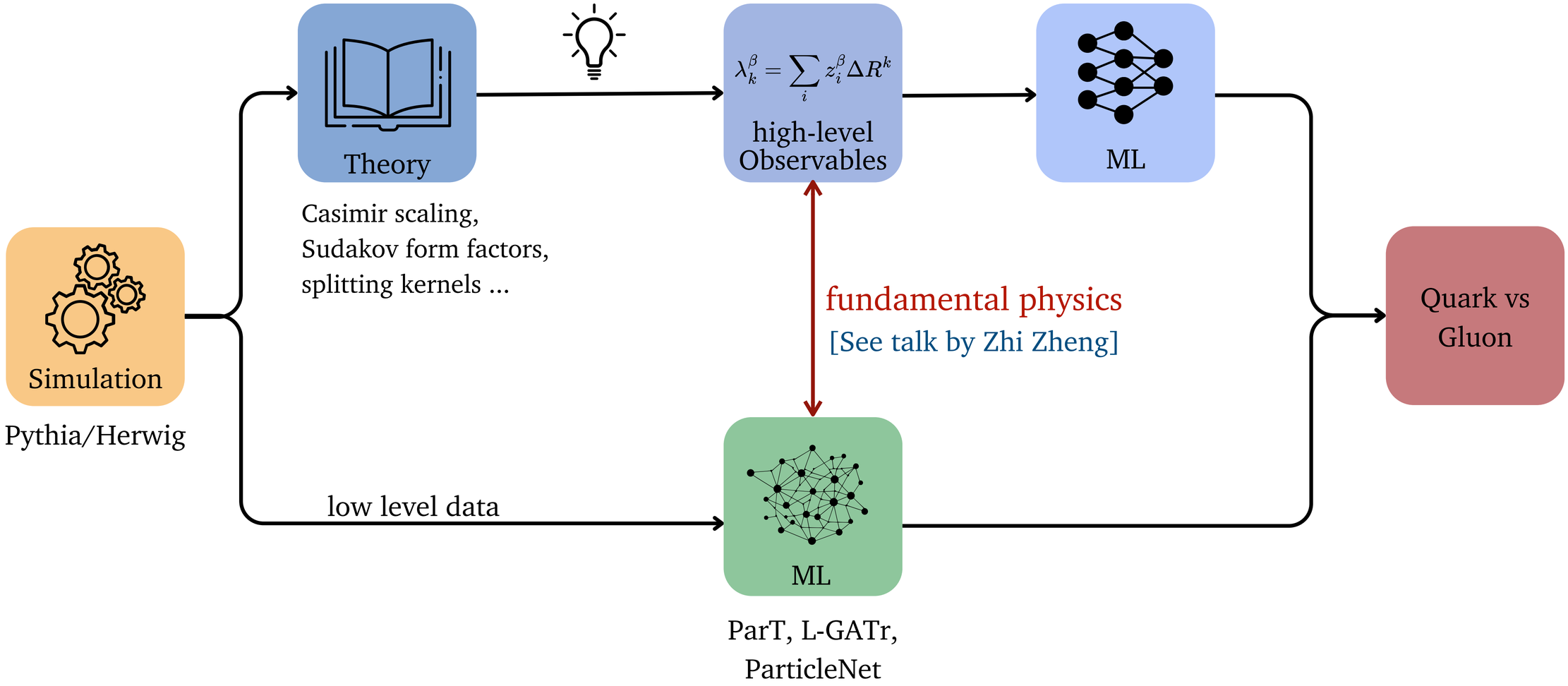
How to Deep-learn the Theory behind Quark Gluon Tagging

Boost 2025 | Brown University

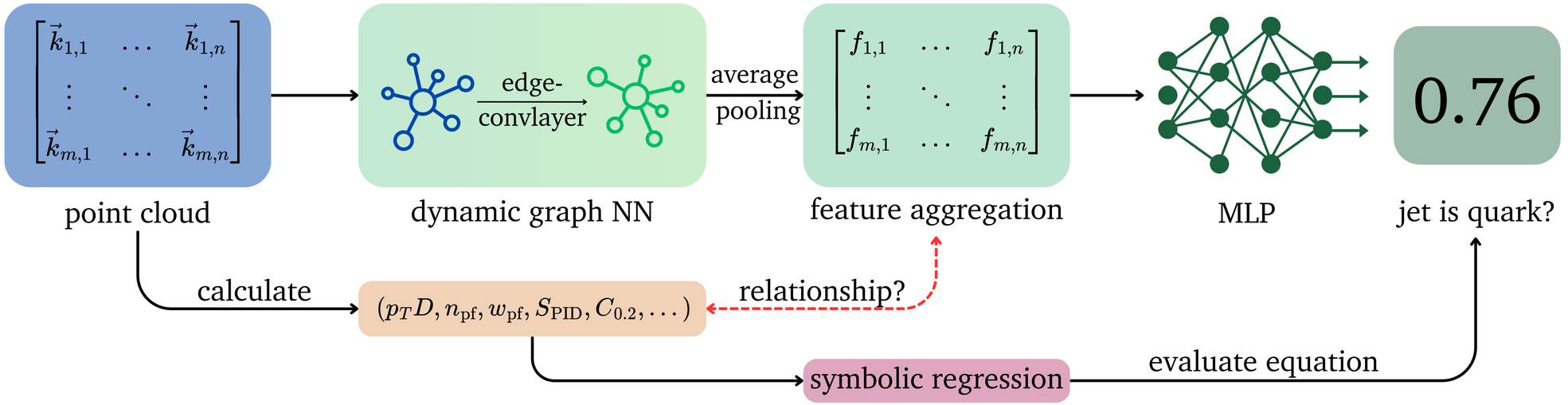
Sophia Vent | Institute for theoretical physics Heidelberg
Ramon Winterhalder & Tilman Plehn

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Motivation


Theory
$\lambda_k^\beta = \sum_i z_i^\beta \Delta R^k$
high-level Observables
ML
Casimir scaling,
Sudakov form factors,
splitting kernels ...
Simulation
Pythia/Herwig
fundamental physics
[See talk by Zhi Zheng]
Quark vs Gluon
low level data
ML
ParT, L-GATr,
ParticleNet

ParticleNet [1902.08570]



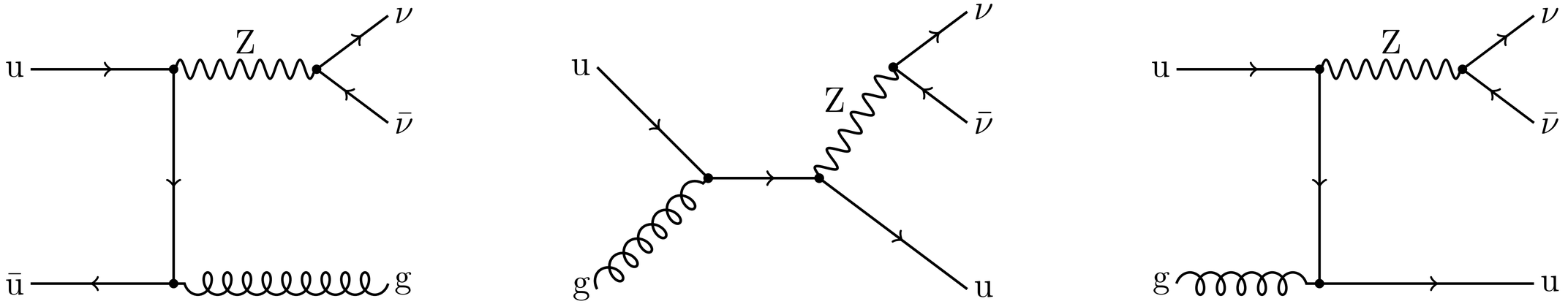
1) decorrelated latent features to physics observables

2) feature importance using SHAP
   [1705.07874]
3) turn neural network into formula using symbolic regression
   [2109.10414]

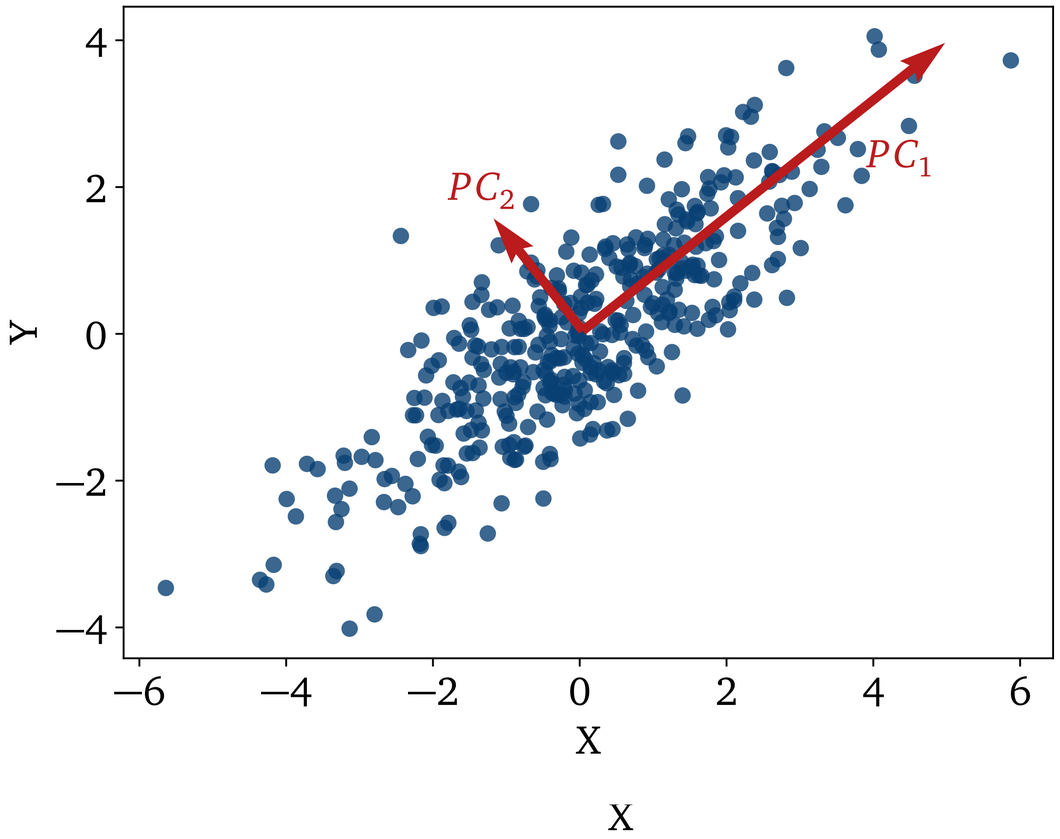
Dataset [1612.01551][1605.04692]

2M quark and gluon from Pythia/Herwig

$q\bar{q} \to Z\,(\to \nu\bar{\nu}) + g$ and $qg \to Z\,(\to \nu\bar{\nu}) + (uds)$



Principal component analysis (PCA)



Basis transformation

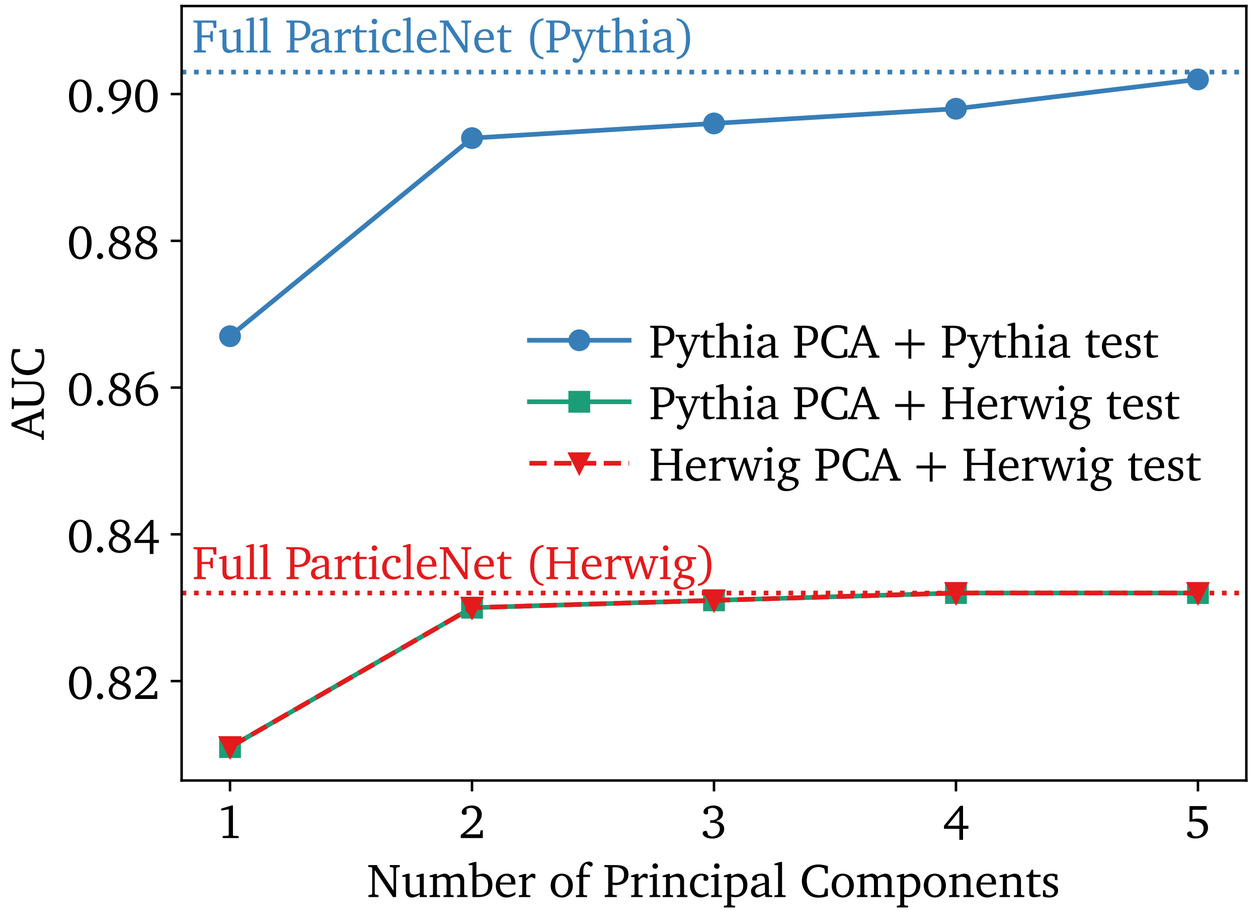
- find new axes that capture the variance
- reduce dimensionality

Dimensionality reduction using PCA

dimensionality $\mathbb{R}^{64} \rightarrow \mathbb{R}^{5}$

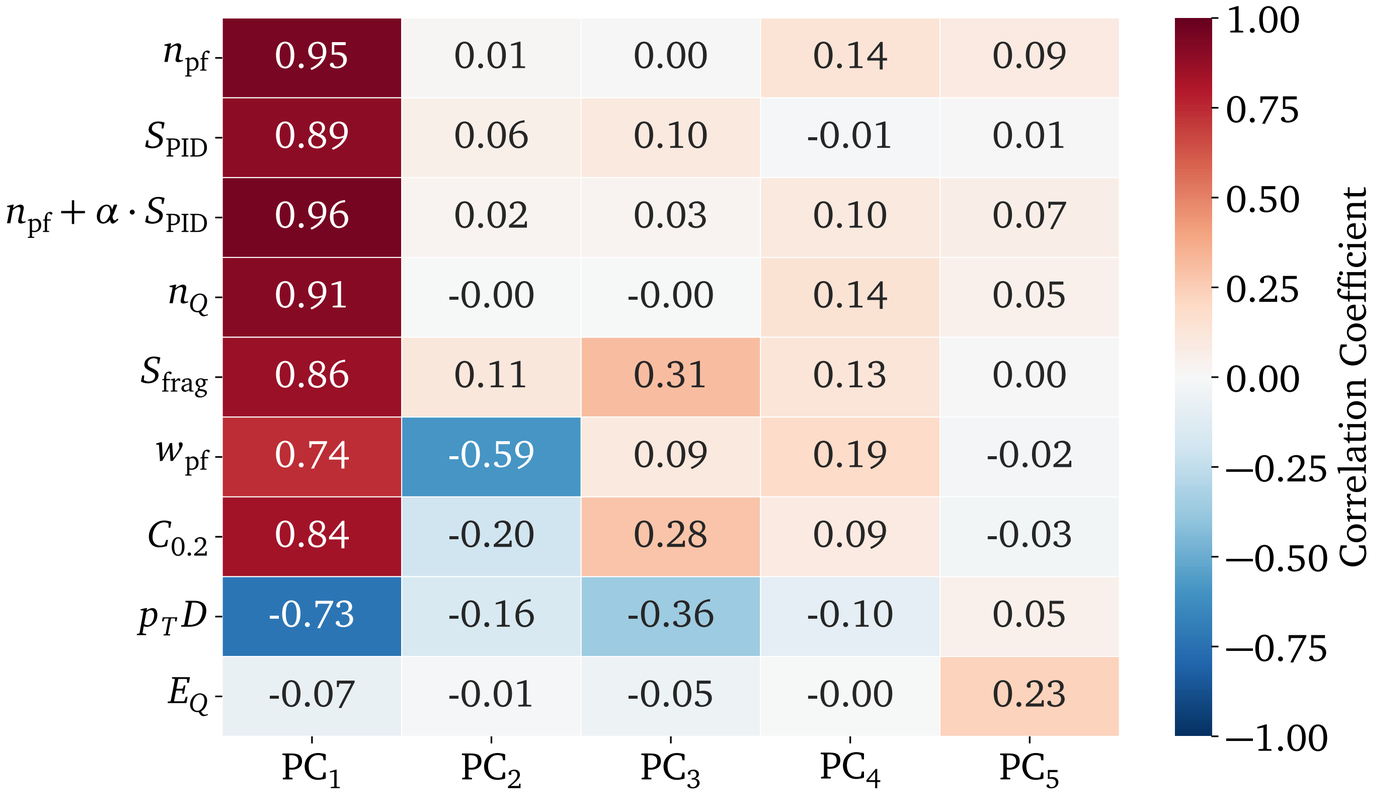
keeping discriminative power

across generators



Latent features vs observables

Latent features vs observables



PC1: multiplicity and diversity

$$S_{\text{PID}} = \sum p_i \log(p_i)$$

particle ID entropy

PC2: jet shape

PC3: fragmentation

PC5: charge

Decorrelated observables

Find observables aligned with PCA

Linear decorrelation

$$O^{\perp}_{\mathrm{PC}k} = O - \sum_{i<k} c_k \cdot \mathrm{PC}i$$

$$w^{\perp}_{\mathrm{pf}} = \alpha \cdot n_{\mathrm{pf}} - w_{\mathrm{pf}}$$

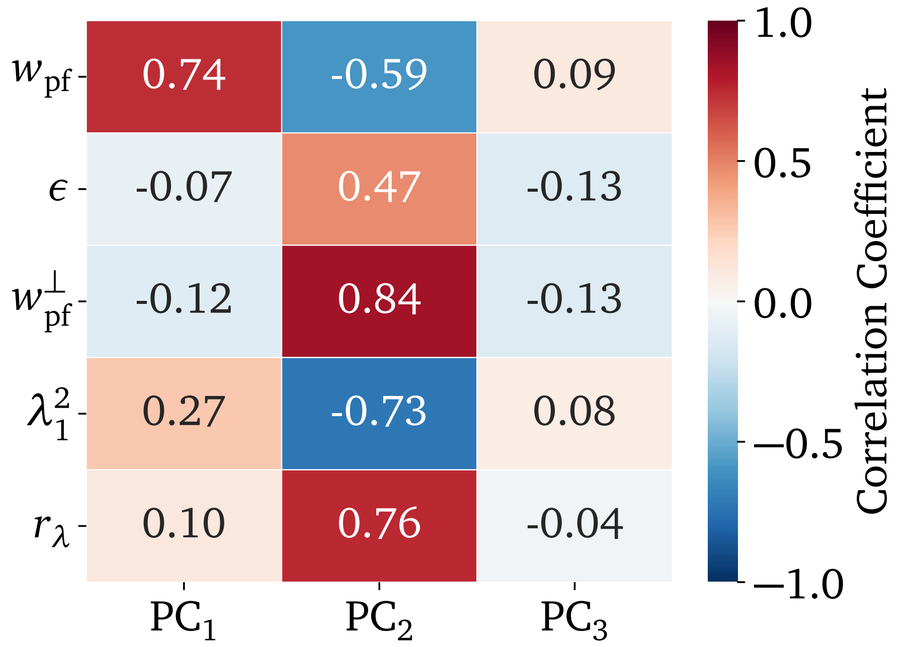
Generalized angularities:

$$\lambda^{\beta}_{k} = \sum_{i} z^{\beta}_{i} \Delta R^{k} \qquad z_i = \frac{p_{T,i}}{p_{T,\mathrm{jet}}}$$

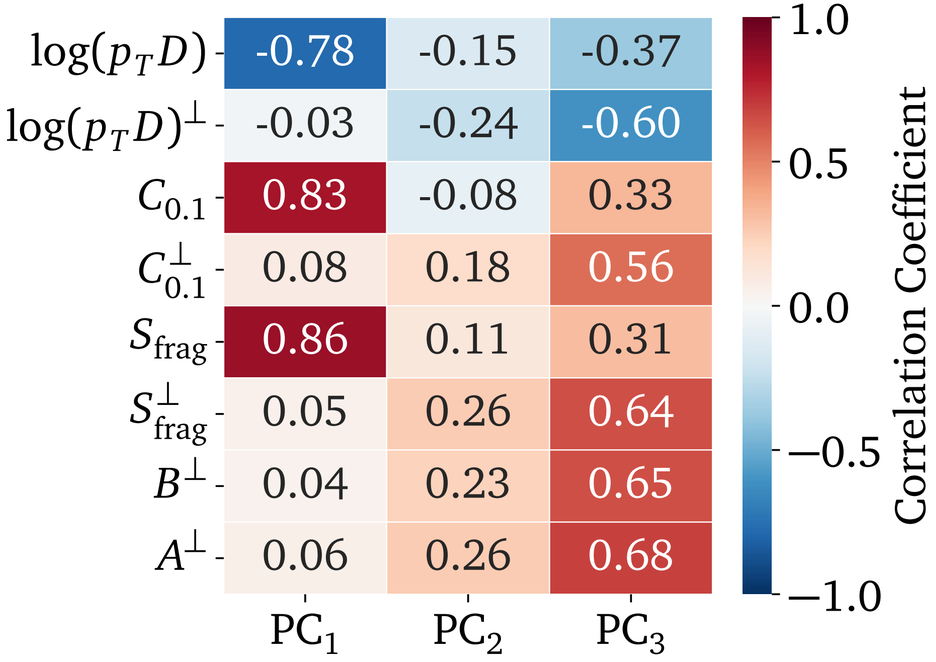
New observable:

$$r_{\lambda} = \frac{\lambda^{1}_{0.5}}{\lambda^{2}_{1}}$$

Feature engineering



PC2: jet shape



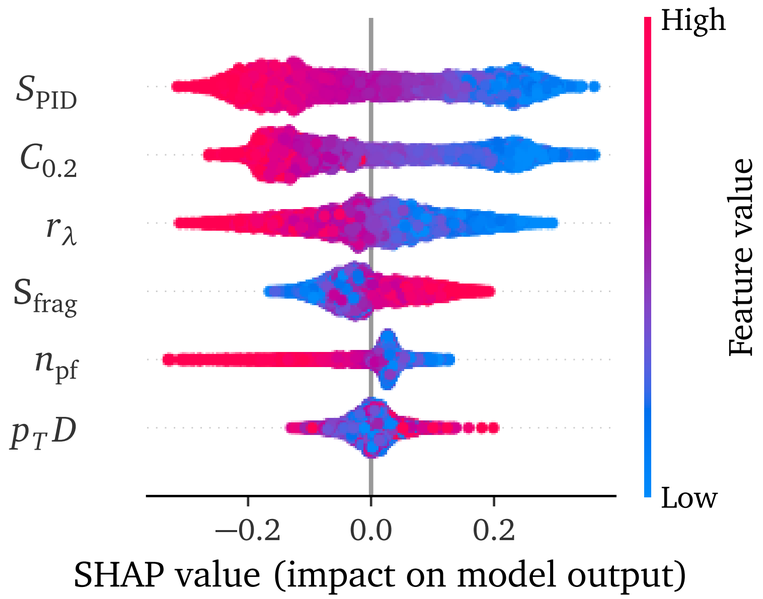
PC3: fragmentation

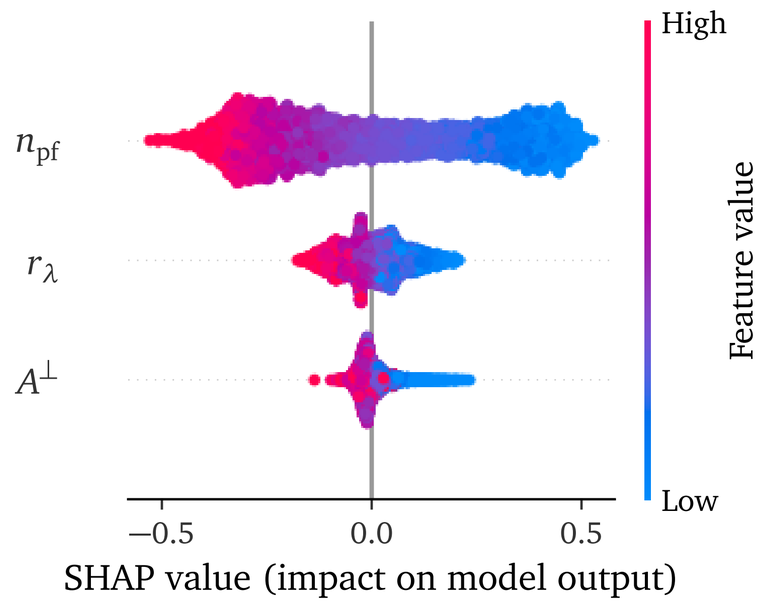
Best performing feature set $(n_{\text{pf}}, S_{\text{PID}}, r_\lambda, C_{0.2}, p_T D, S_{\text{frag}}, E_Q)$

SHapely Additive exPlanations

Feature importance

Quantifying feature contributions to prediction





counterintuitive/wrong

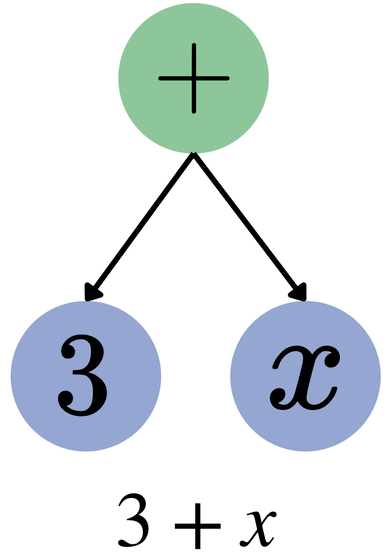
decorrelated features

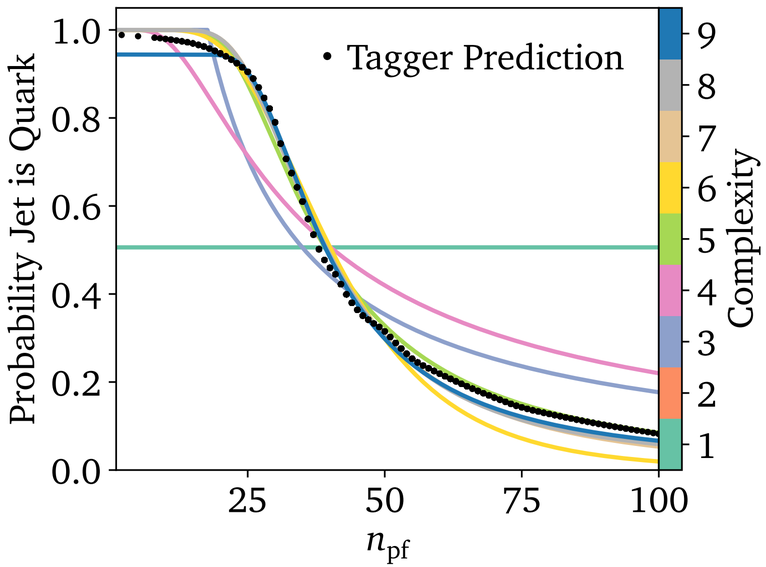
Turning networks into formulas

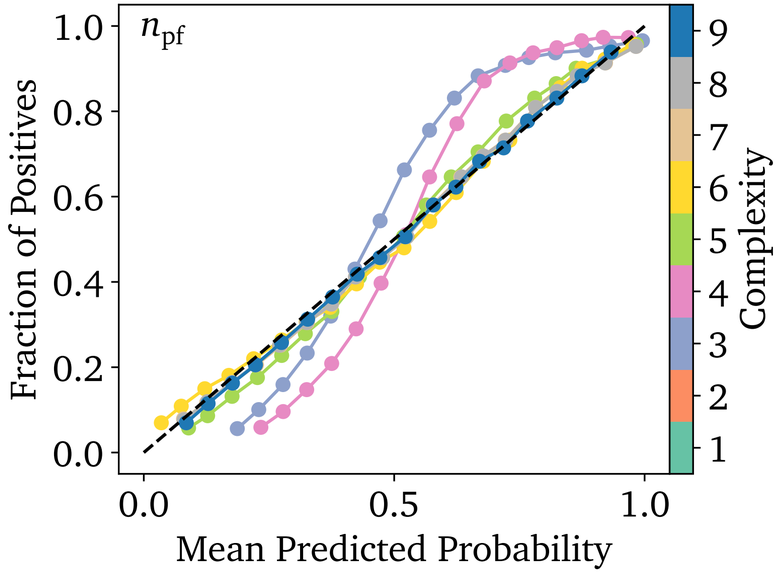
Symbolic Regression using PySR [2305.01582]

PySR fits a formula in formula space

A network is a **complex function approximation**
→ can be described by a formula







One dimensional formula

Observable	Formula
n_{pf}	$0.94 \cdot \tanh\left(21036 \cdot \left(0.005 + \frac{1}{n_{\text{pf}}}\right)^3\right)$
S_{PID}	$\tanh^2\left(\frac{1.14}{S_{\text{PID}}^3}\right)$
S_{frag}	$\tanh^2\left(\frac{18.08}{S_{\text{frag}}^3}\right)$

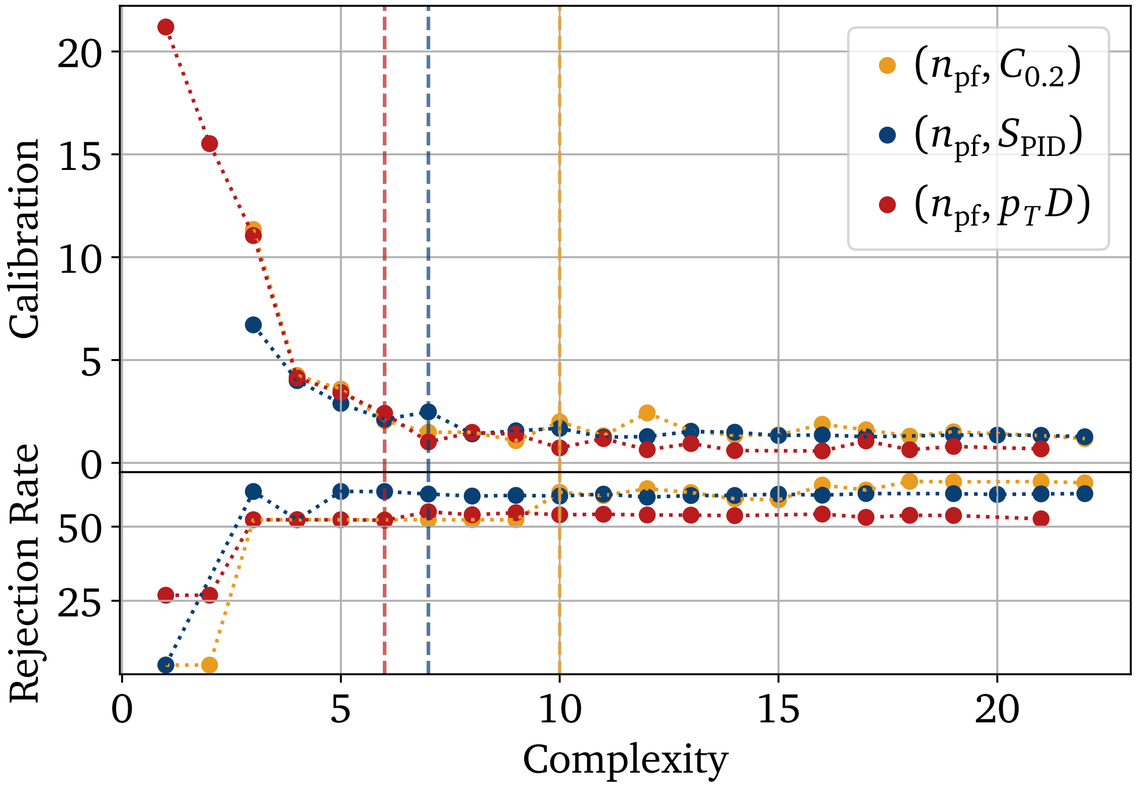
Formulas align with theoretical intuition

For high n_{pf} the term

$$\frac{1}{n_{\text{pf}}} \to 0 \text{ and } \lim_{x \to 0} \tanh(x) \to 0$$

$\to$ Gluon label

Two dimensional formula



- Equation based on only one observable up to a certain complexity
- complexity 15 needed

Tagging formula

$$\tanh^3\left(0.55 \cdot C_{0.2} + 2 \cdot \left(-0.02 \cdot r_\lambda \cdot \left(C_{0.2} \cdot p_T D \cdot S_{\text{PID}} \cdot S_{\text{frag}} - 0.25\right) + 1\right)^3\right)$$

- neural network AUC $= 0.872$
- SR formula AUC $= 0.871$

Summary

- Explainability of **trained** networks
- ParticleNet re-learns known observables
- New observables r_λ and S_{PID}
- SHAP condition not fulfilled
- Approximate tagger formula

[2507.21214]

Thank you!

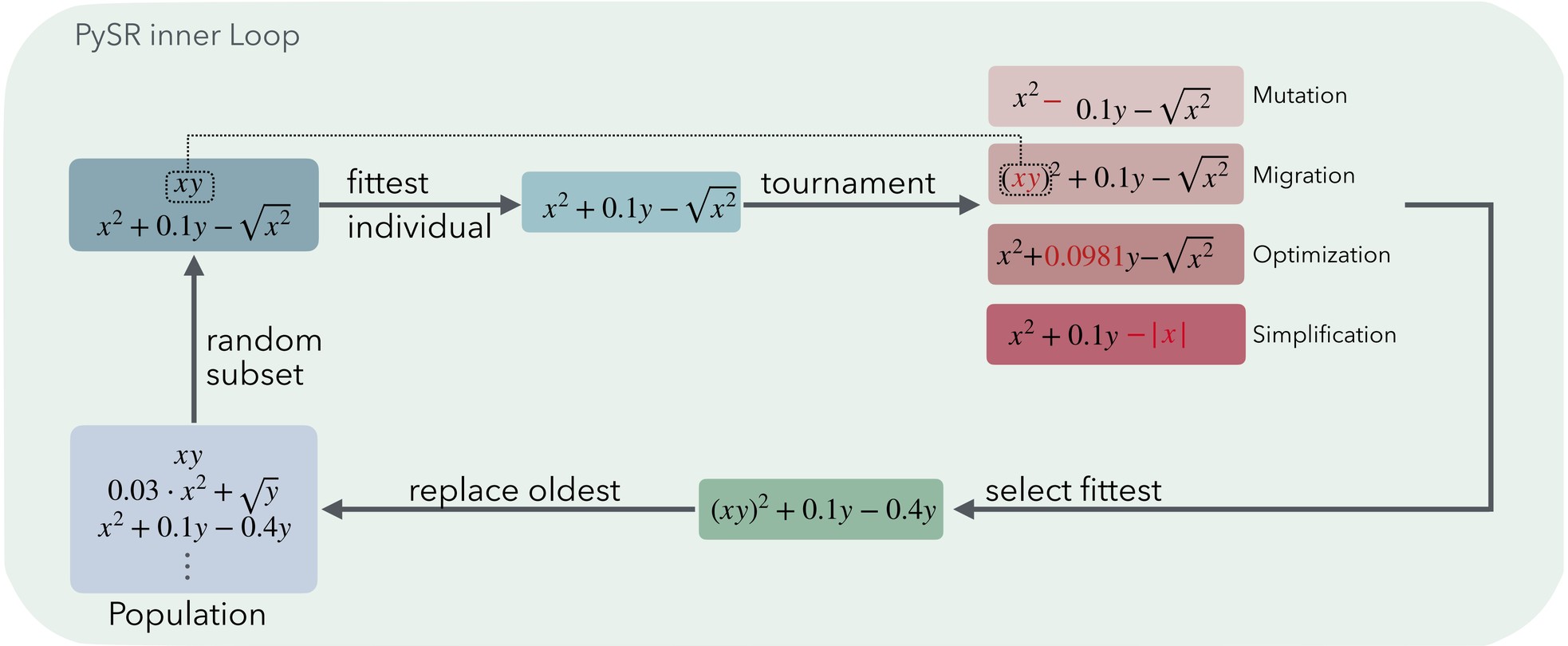
Additional slides

Definition of observables

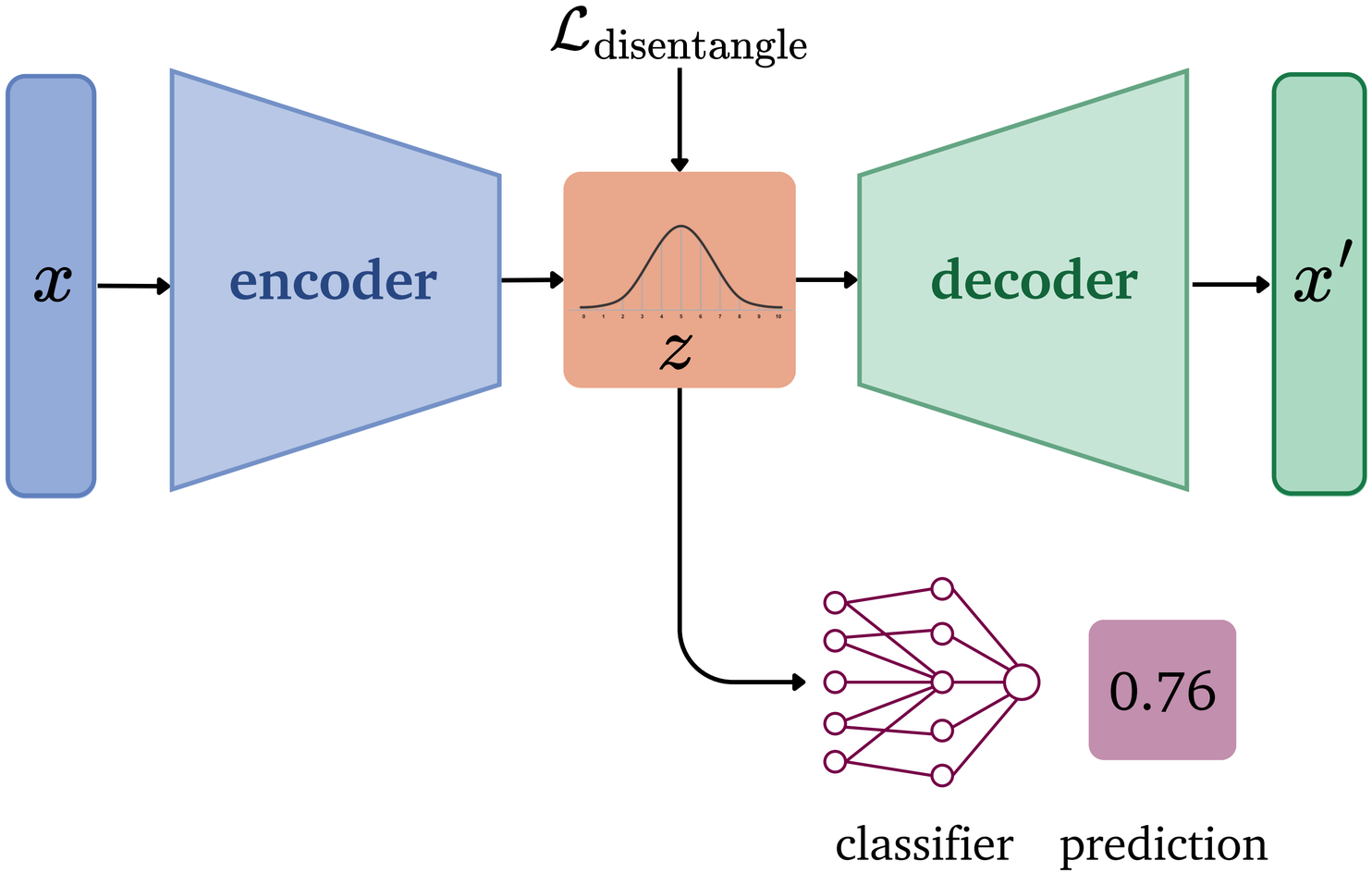
$$n_{\mathrm{pf}} = \sum_i 1 \qquad w_{\mathrm{pf}} = \frac{\sum_i p_{T,i} \Delta R_{i,\mathrm{jet}}}{p_{T,\mathrm{jet}}}$$

$$C_\beta = \frac{\sum_{i<j} p_{T,i} p_{T,j} (\Delta R_{ij})^\beta}{\left(\sum_i p_{T,i}\right)^2} \qquad p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \,.$$

Symbolic regression using PySR



Decorrelated Latent Classifier



SHAP (SHapley Additive exPlanations)

Quantify each feature's contribution to the model's prediction

Shapely values from gametheory

$$\mathscr{V}_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \Big[f(S \cup \{i\}) - f(S) \Big]$$

Uncover:

1) Feature importance ranking

2) Identifying which regions of a feature drive predictions