Symbolic Regression for Precision LHC Physics

M. Morales Alvarado*, J. Bendavid, D. Conde, M. Ubiali, V. Sanz

*University of Cambridge

1 Introduction

- In high energy physics (HEP), fundamental interactions can be described by the **Standard Model** (SM) of particle physics to a good extent.
- The SM is tested at the world's most powerful microscope: the **Large Hadron Collider** (LHC). Theoretical predictions and observables are compared.

@CERN

- **Formulas** are the actual language of theoretical predictions.
- **The question:** Is there a way to find simple, accurate, and closed analytical formulas from noisy, potentially high dimensional datasets?

2 Methodology

- We use symbolic regression (SR) with the PySR [1] library.
- Formulas are described as expression trees, as shown in Fig. 1. An evolutionary algorithm implements mutations/crossover between trees, as shown in Fig. 2, and new trees arise.

$3.1y\,(x^2+1) \leftrightarrow$

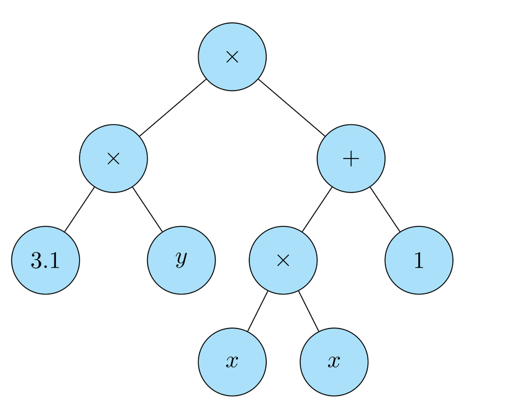


Fig. 1: Expression tree

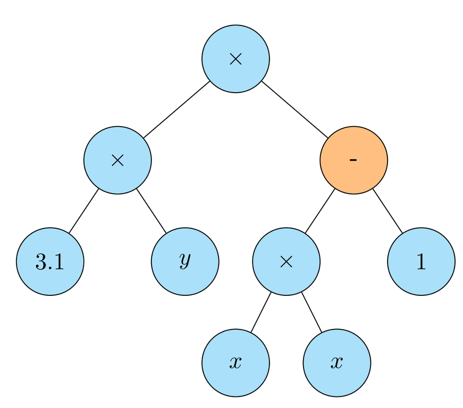


Fig. 2: Mutation

- Each tree has a complexity c, related to its number of nodes.
- Fittest trees optimise a certain metric

Accuracy: Minimise

$$L = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

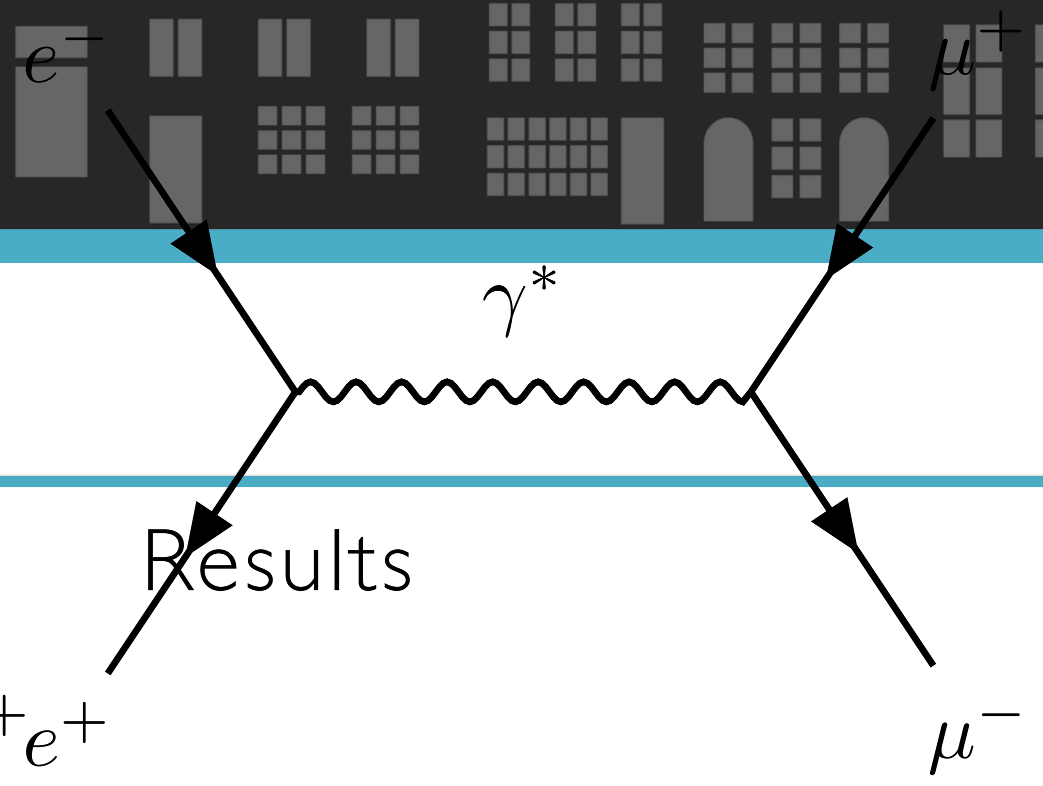
Score: Maximise

$$-\frac{\partial \log(L)}{\partial c}$$

Best: Highest score with

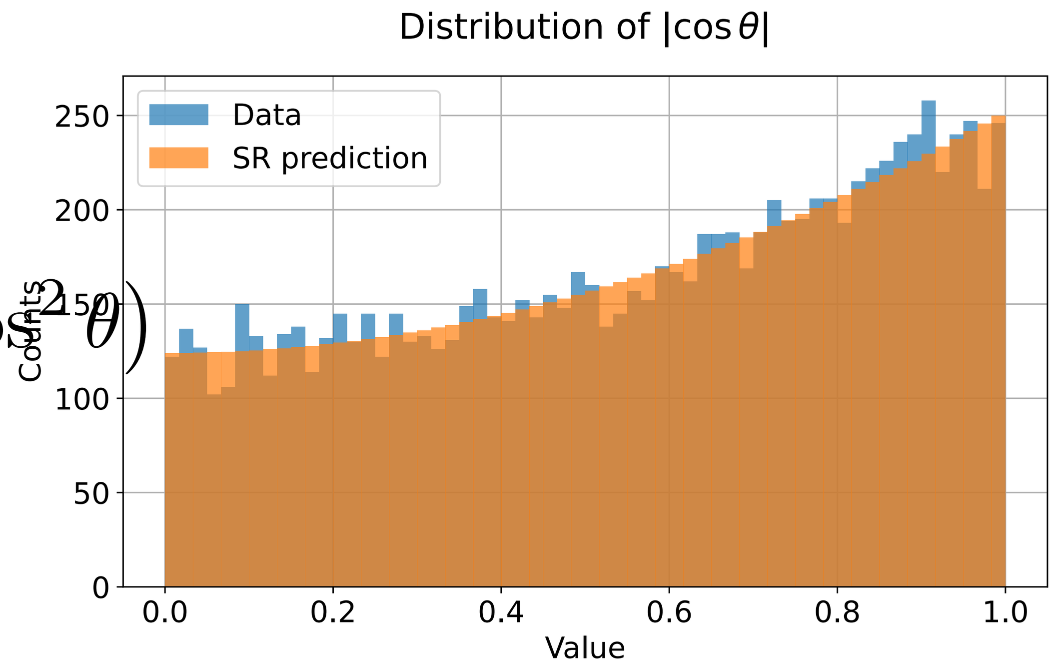
$$L \leq 1.5 \times L_{\min}$$

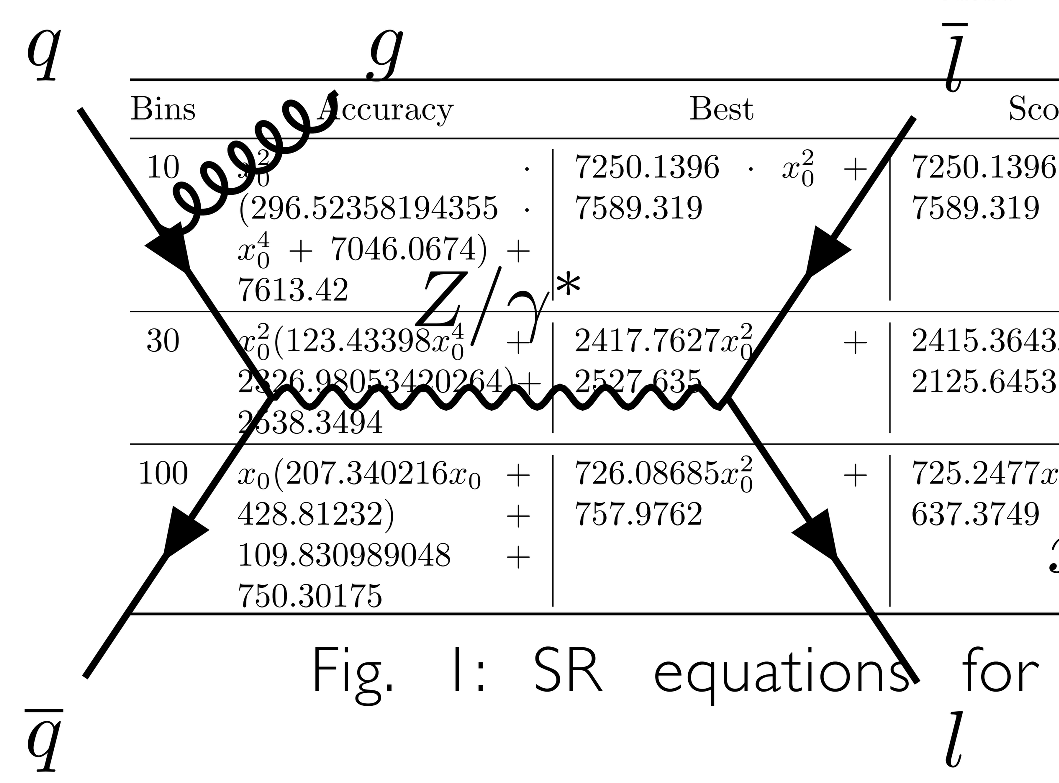
3 Results



$$e^- e^+ \to \mu^- \mu^+ e^+$$

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{4s}\left(1+\cos^2\theta\right)$$





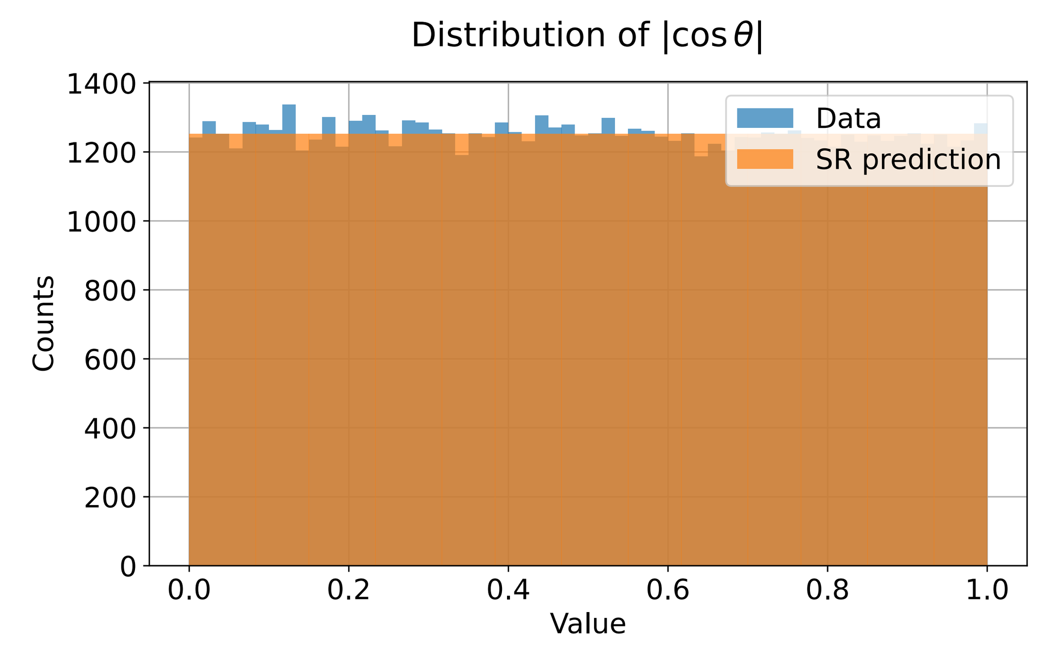
Bins	Accuracy	Best	Score
10	$x_0^2 \cdot (296.52358194355 \cdot x_0^4 + 7046.0674) + 7613.42$	$7250.1396 \cdot x_0^2 + 7589.319$	$7250.1396 \cdot x_0^2 + 7589.319$
30	$x_0^2(123.43398x_0^4 + 2326.98053420264) + 2538.3494$	$2417.7627x_0^2 + 2527.635$	$2415.3643x_0 + 2125.6453$
100	$x_0(207.340216x_0 + 428.81232) + 109.830989048 + 750.30175$	$726.08685x_0^2 + 757.9762$	$725.2477x_0 + 637.3749$

$$x_0 = \cos\theta$$

Fig. 1: SR equations for

Reweighted distribution (constant)

$$\sigma = \frac{4\pi\alpha^2}{3s}$$



Denoise	Accuracy	Best	Score
True	6833.678	6833.678	6833.678
False	$x_0^2(661.2393x_0 - 706.5968)(x_0^2 + x_0) + 7593.010$	$7614.049 - 194.01741x_0$	$7614.0547 - 194.02724x_0$

$$p\,p \to \gamma^*/Z \to l^+ l^-$$

$$\frac{d^5\sigma}{dp_T\,d\eta\,dm\,d\cos\theta\,d\phi} = \frac{3}{16\pi}\frac{d^3\sigma^{U+L}}{dp_T\,d\eta\,dm}\left[(1+\cos^2\theta) + \sum_{i=0}^{7} P_i(\theta,\phi)\,A_i\right]$$

$$A_i = A_i(p_T, \eta, m)$$

References

UNIVERSITY OF CAMBRIDGE

PBSP

erc European Research Council

morales.alvarado@damtp.cam.ac.uk