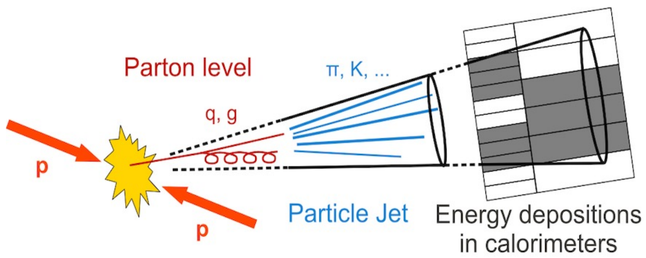
Semi-Visible Jet Classification with Boosted Decision Tree

Hebu Patil

BACKGROUND

Particle colliders record large amounts of data on jets produced from proton-proton collisions. Some jets are described by the Standard Model, but some jets contain Dark Matter that cannot be described.



Goal: Achieve higher ratio of signal to background by discriminating Semi-Visible Jets from Standard Model jets using the Boosted Decision Tree.

Model Overview

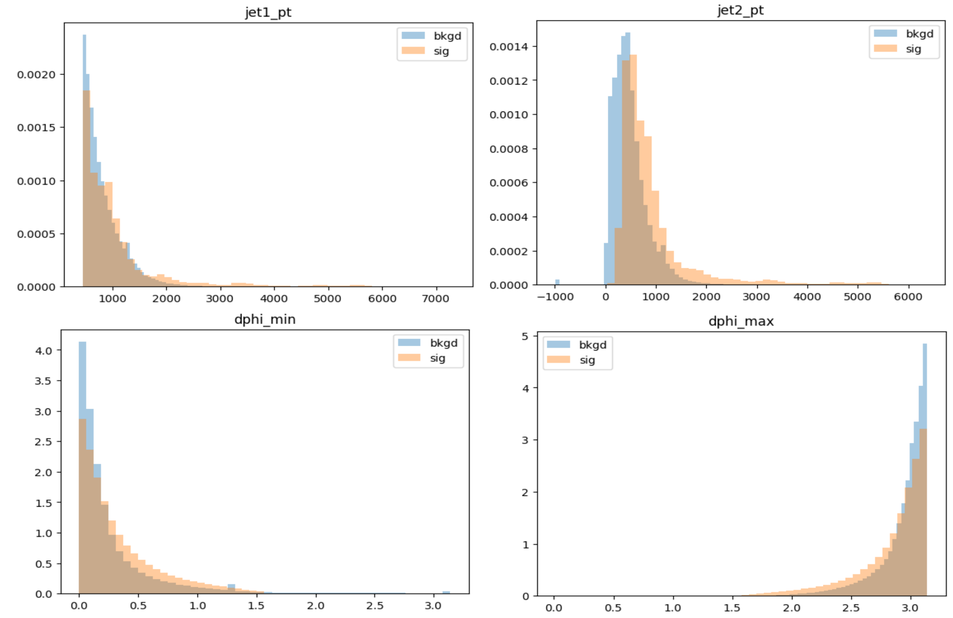
- **All features are high-level features of jets produced in simulation**
- **Around 2:1 ratio of signal to background jets**
- **Unweighted features, weighting will be implemented in the future**
- **Similar data to that used for the SVJ project conducted by Ki Park at Columbia University**

DATASET

Signal: 2294295 jets
Background: 1000000 jets

Features (19):

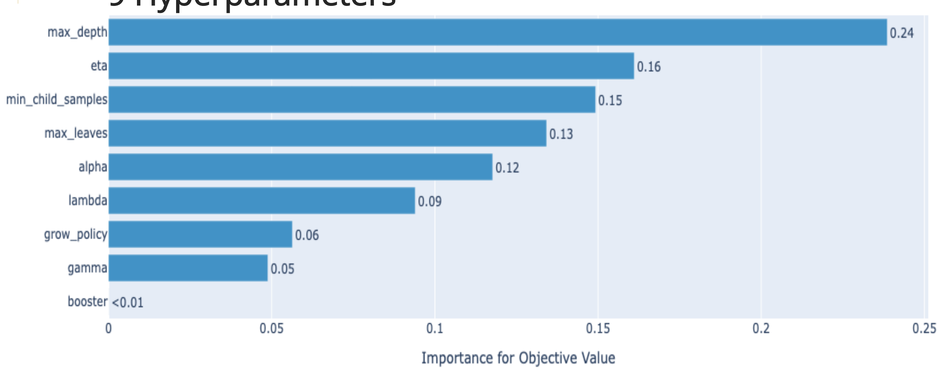
jet_pt	deta_12	mjj_12
jet1_phi	deltaY_12	mT_jj
jet2_pt	hT	dR_12
jet2_phi	rT	sphericity_T
dphi_min	aplanarity	met_met
dphi_max	sphericity	
met_phi	pt_balance_12	



HYPERPARAMETER OPTIMIZATION

Optuna used to optimize hyperparameters by setting ranges to test different values of hyperparameters

10 Trials
9 Hyperparameters



RESULTS

Base Model (without Optuna): **82.34%**

Tuned Model: **84.62%**

```
Best Accuracy: 0.84617
Best params:
        booster: gbtree
        lambda: 4.691548032259625e-11
        alpha: 0.011855309247426364
        max_depth: 40
        min_child_samples: 53
        eta: 0.26306837563252955
        gamma: 1.4900138899841883e-08
        grow_policy: depthwise
        max_leaves: 884
```

- Previous model used TMVA, new uses XGBoost
- XGBoost outperforms LGBM and Sci-kit Learn in preliminary tests
- Optuna used for hyperparameter tuning