Rediscover the Higgs Boson with ML Techniques

Fatemeh Farhangian, Zahra Sadat Shobeiri
Shahid Beheshti University, Tehran, Iran

Abstract
In this report we observe the effects of different cuts on the performance of ML algorithms with the goal to find the ratio of signal to background for proton-proton collision CERN Open Data

1 Introduction

There are many ways to improve the performance of ML algorithms. One of them is to reduce the data meaningfully. Here we use cuts on processes that we know are not of much importance. Cuts on 'lepton charge' and 'lepton type' turned out to be the best resulting.

2 Choosing Cuts

Because we are limited to 4-lepton decay in the data,

$$H \rightarrow ZZ^* \rightarrow 4l$$

candidates can be as below:

- lepton transverse momentum > 6 GeV
- sum of lepton charges = 0
- sum of lepton types = 44 ($eeee$) and 48 ($e\mu e\mu$) and 52 ($\mu\mu\mu\mu$)

There were some complications due to low amount of data, so these were the best resulting cuts.

3 Results with Different Algorithms

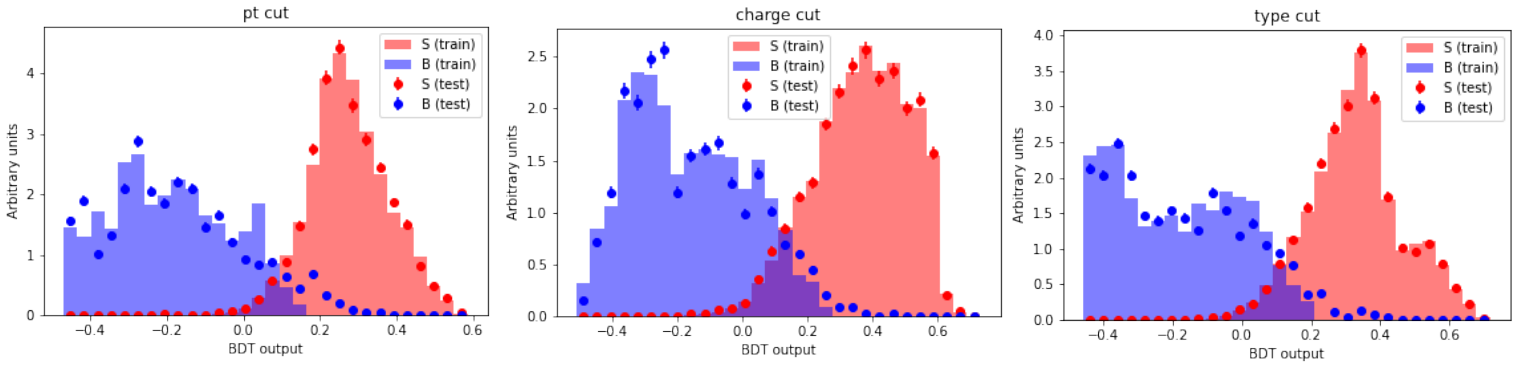
3.1 Boosted Decision Tree

Pt cut generally lowered the performance which we think is due to low amount of data.

The following are the performance on the test data.

BDT Results

Metrics	Pt cut	Charge cut	Type cut
Accuracy	0.98	0.97	0.98
F1-Score	0.93	0.91	0.91
AUC	0.98	0.98	0.98



Overfit checks for different cuts

3.2 Neural Network Model

NN Results

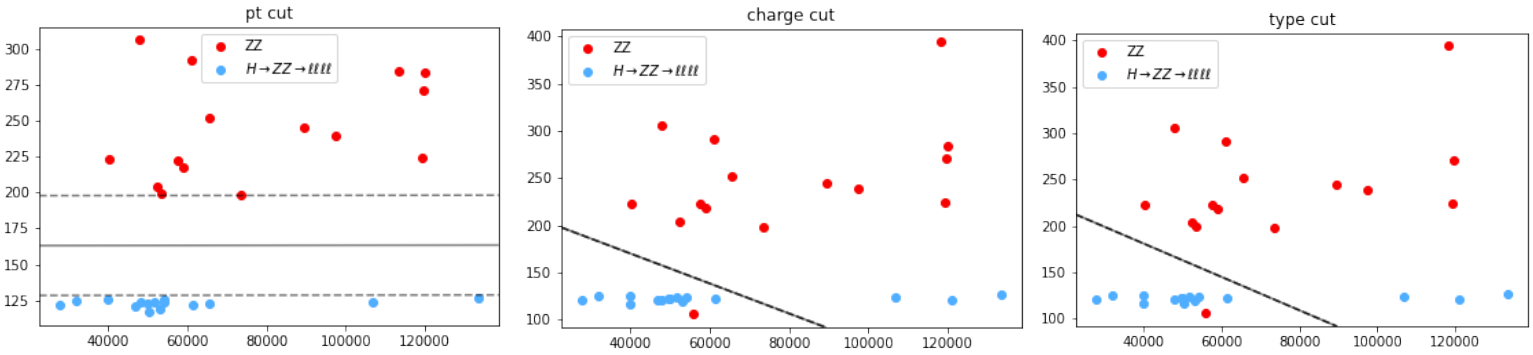
Metrics	Pt cut	Charge cut	Type cut
Accuracy	0.98	0.96	0.96
F1-Score	0.91	0.82	0.83
AUC	0.87	0.73	0.76

3.3 Support Vector Machine (SVM)

This is the result for only 15 events in the dataset.

SVM Results

Metrics	Pt cut	Charge cut	Type cut
Accuracy	1.00	0.87	0.87
F1-Score	1.00	0.87	0.87
AUC	1.00	0.90	0.88



Decision function for different cuts

4 Conclusion

We can see that the cut on transverse momentum was able to remove background more effectively. That is transverse momentum is of the most importance to remove background.