山东大学 SHANDONG UNIVERSITY

Department of Physics, Shandong University

Compressed EWK study(ISRC1N2)

Chengxin Liao
liaocx@ihep.ac.cn

Wed, Jan 15, 2024

Outline

1. Grid Search for hyperparameter of BDTG

2. Problems in TMVA

Task-list

- ~~Fundamentals of Machine Learning & BDT introduction(DONE)~~
  ref: Bishop, C. M. & Bishop, H. Cham, S., ed. (2023), Deep Learning - Foundations and Concepts.
  ref: https://indico.cern.ch/event/472305/contributions/1982360/attachments/1224979/1792797/ESIPAP_MVA160208BDT.pdf

- ~~Learning TMVA based on it's user guide and tutorial (DONE)~~

- Machine learning for HH channel
  - check more Variable and select significance var for ML(In Progress)
  - BDTG hyperparameters optimization/ Setup a Grid Search framework (In Progress)

- Preliminary study on multibody quantum mechanics [高量] (In Progress)
  QFT Lecture (Peskin part I)

- BSc thesis: *https://www.overleaf.com/project/674e7119837a2580151a0868*

- CS61A (python): *https://cs61a.vercel.app/index.html*

Grid Search for hyperparameter of BDTG

Input(HH-Channel):

Sample:

Sig: ISRC1N2(mass_C1 = 100GeV, mass_N2 = 70GeV)->12180 entries
Bkg: 513850 entries

All input data(C1N2_100_70 and Bkg) already passed pre-selection

Variables:

pt_lep, pt_tau, METsig, Mll, mt_tau, MT2_50, dPhit1x

Startegy:

method: BDTG/AdaBDT
hyper parameter: NTrees=300,learning rate=0.05,max depth=10, MinNodeSize=1%(default)
Separate sig(bkg) into five folders, one for test, the other three for train, and last one for validation set, then traverse all possibilities.

```
Number of training and testing events
-------------------------------------------------------
Signal     -- training events            : 7311
Signal     -- testing events             : 2436
Signal     -- training and testing events: 9747
Background -- training events            : 308329
Background -- testing events             : 102770
Background -- training and testing events: 411099
```

Pre-Selection

$$had\text{-}had\ \ channel: nTaus \geq 2, nLeps = 0$$
$$pass\ MET\ trigger;\ MET \geq 200$$
$$1 \leq nBaseJet \leq 8$$
$$b - Veto$$
$$OS$$

Grid Search for hyperparameter of BDTG

Grid Search:

a simple test for grid search

Hyperparameter Grid Selection:

Ntrees: 200, 300
Max Depth: 6, 8
MinNodeSize: 1
Learning Rate: 0.01

Expanding Grid Search →

Binned significance: $Z = \sqrt{2((s_i + b_i)\log\left(1 + \frac{s_i}{b_i}\right) - s_i)}$

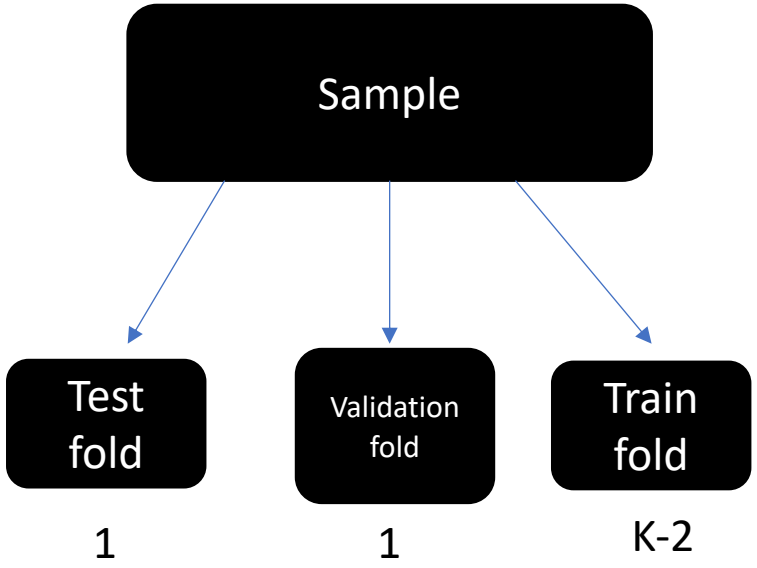
Model	Binned Significance	Zn(Syst=30%)
200_6_1_0.01	11.7823	2.84505
200_8_1_0.01	11.8305	2.90997
300_6_1_0.01	12.0497	2.95435
300_8_1_0.01	12.067	2.90483

just a test of Grid Search and have no practical reference value

More details and BDT Score distribution will show after expanding grid search

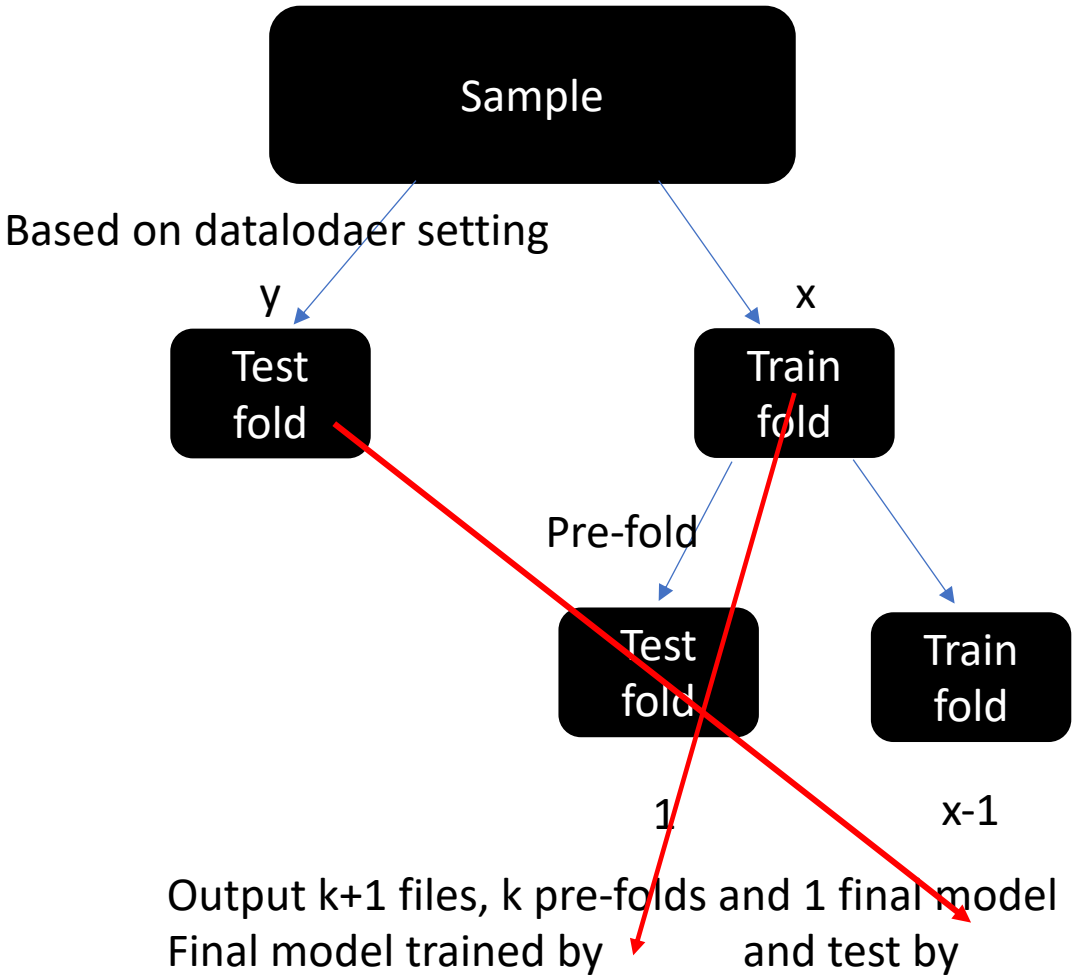
Problems in TMVA

Our k-fold method to split



Output k files

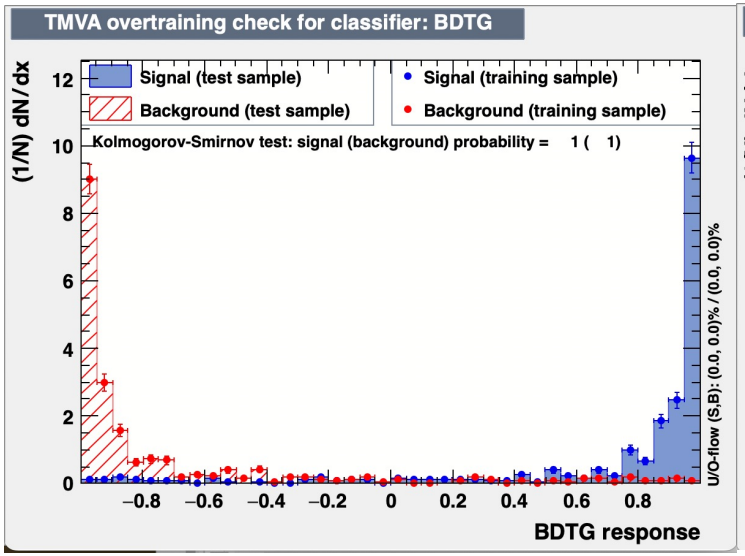
TMVA method to split

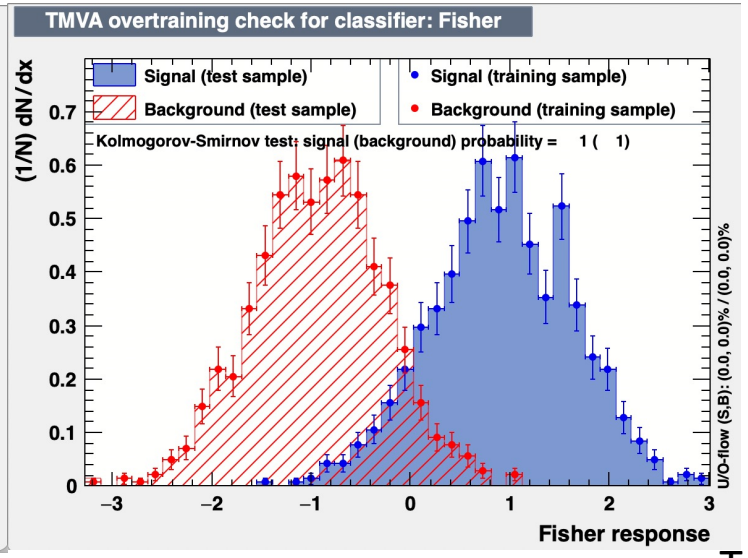


Problems in TMVA

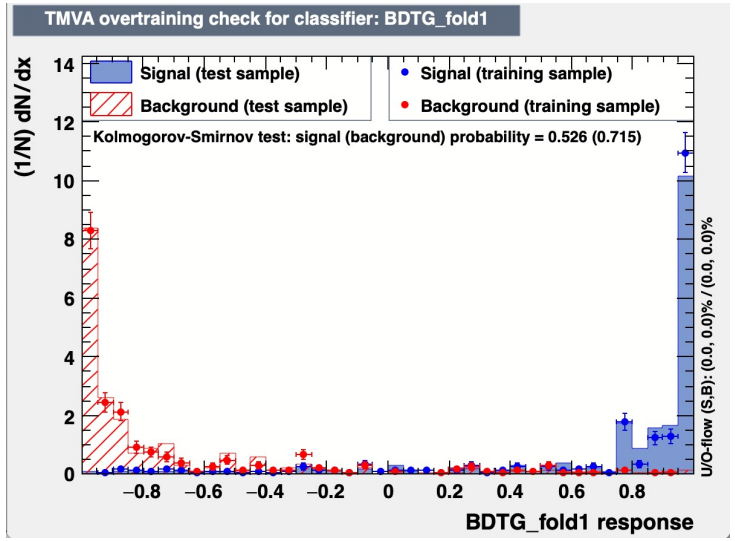
$ROOTSYS/tutorials/tmva/TMVACrossValidation.C

All KS test = 1, which is impossible





Pre-fold looks good

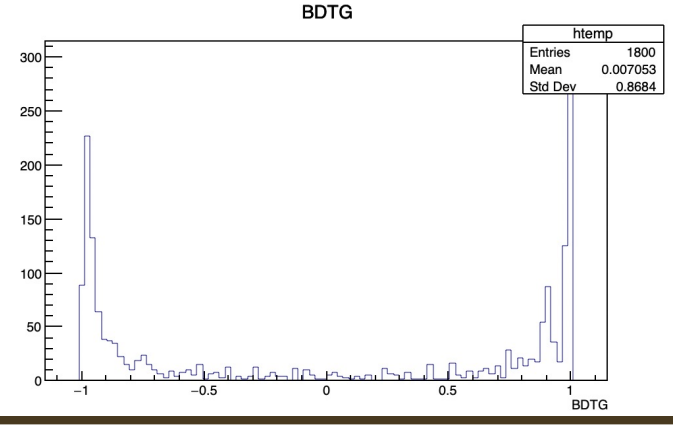


Final model

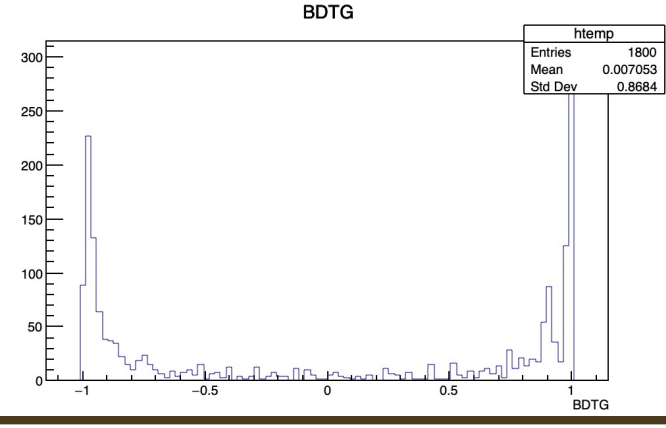
input

```
Number of training and testing events
-------------------------------------------------------
Signal     -- training events            : 900
Signal     -- testing events             : 100
Signal     -- training and testing events: 1000
Background -- training events            : 900
Background -- testing events             : 100
Background -- training and testing events: 1000
```

Test tree



Train tree



TODO

1. Expand Grid Search and add more variables
2. Build the Grid Search framework for TMVA
3. Completing the theoretical section of the my thesis