Status of 4Top analysis

Huiling Hua[1], Fabio Iemmi[1], Ducan Leggat[2], Hongbo Liao[1], Hideki Okawa[2], Yu Zhang[2]

1,IHEP, 2,Fudan University

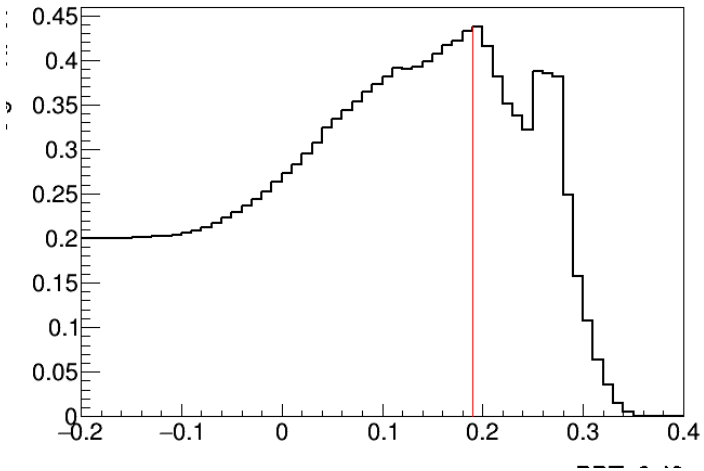
April 27, 2022

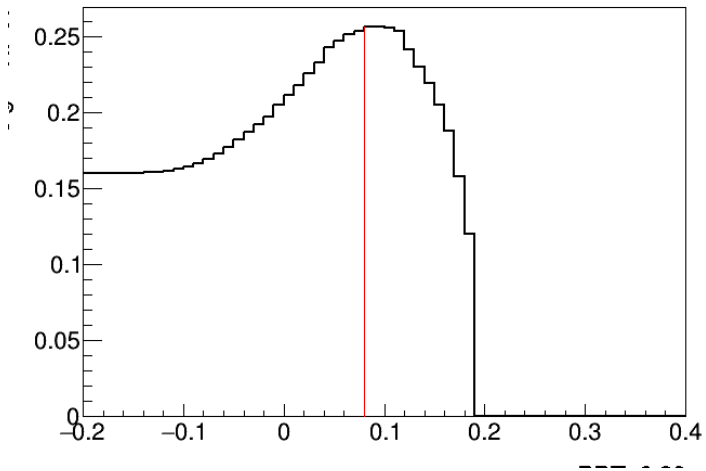
Variable bin width

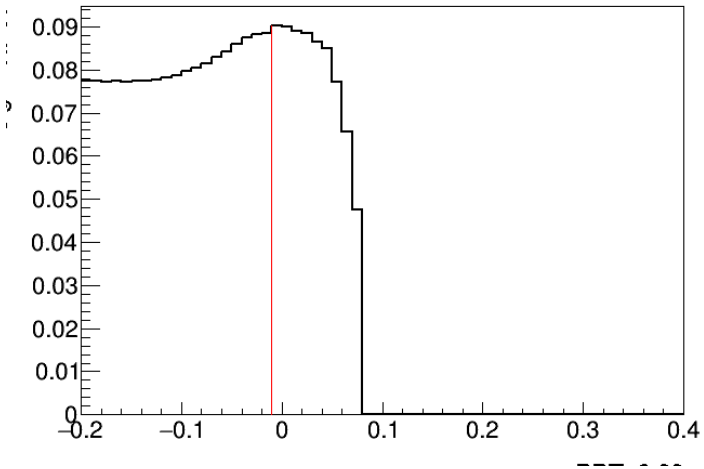
- Fit range of BDT score : [-0.2, 0.4]
- Scan the bin edge x_0 and calculate the significance in [x_0,0.4]
- Find the best threshold th_0 and determine the first bin [th_0,0.4]
- Repeat
  - Scan x_i and calculate the significance [x_i,th_{i-1}]
  - Find the best threshold th_i and determine the bin [th_i,th_{i-1}]
- Find the optimal binning of all the BDT scores.

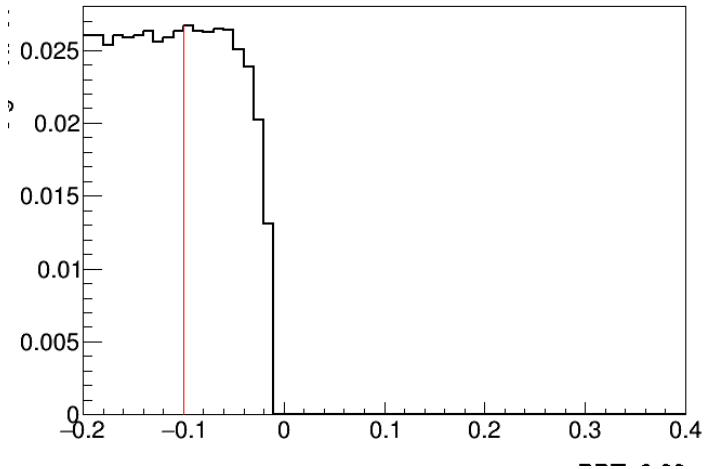
number of variables	combined σ	1st bin	2nd bin	3rd bin	4th bin
9Var	0.516±0.073	0.438±0.085	0.257±0.023	0.090±0.006	0.027±0.002
BDT	-	[0.19,0.40]	[0.09,0.19]	[-0.01,0.09]	[-0.10,-0.01]
10Var	0.475±0.057	0.424±0.063	0.201±0.016	0.069±0.005	0.026±0.002
BDT	-	[0.17,0.40]	[0.07,0.17]	[0.00,0.07]	[-0.08,0.00]

Scanning the edge - Significance vs bin edge

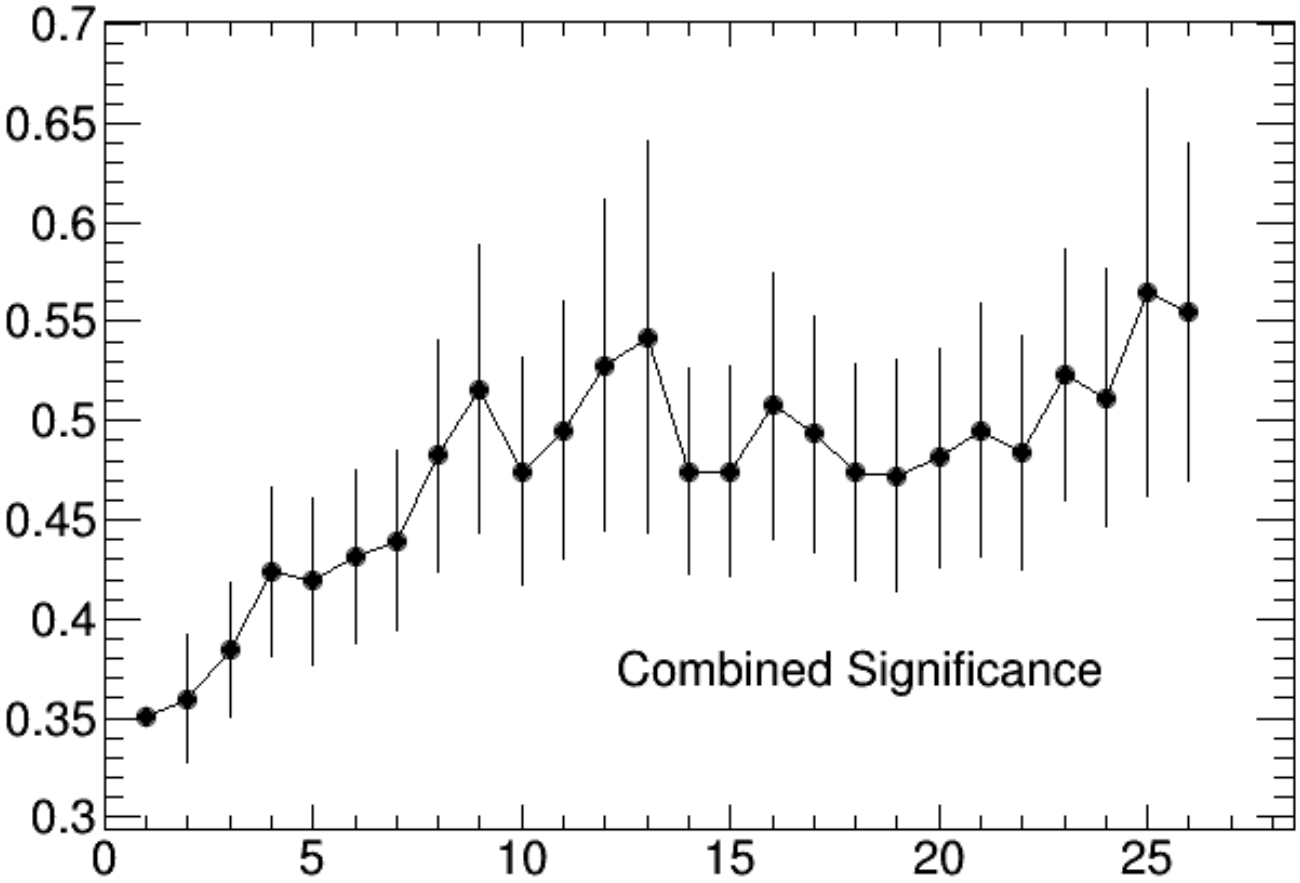






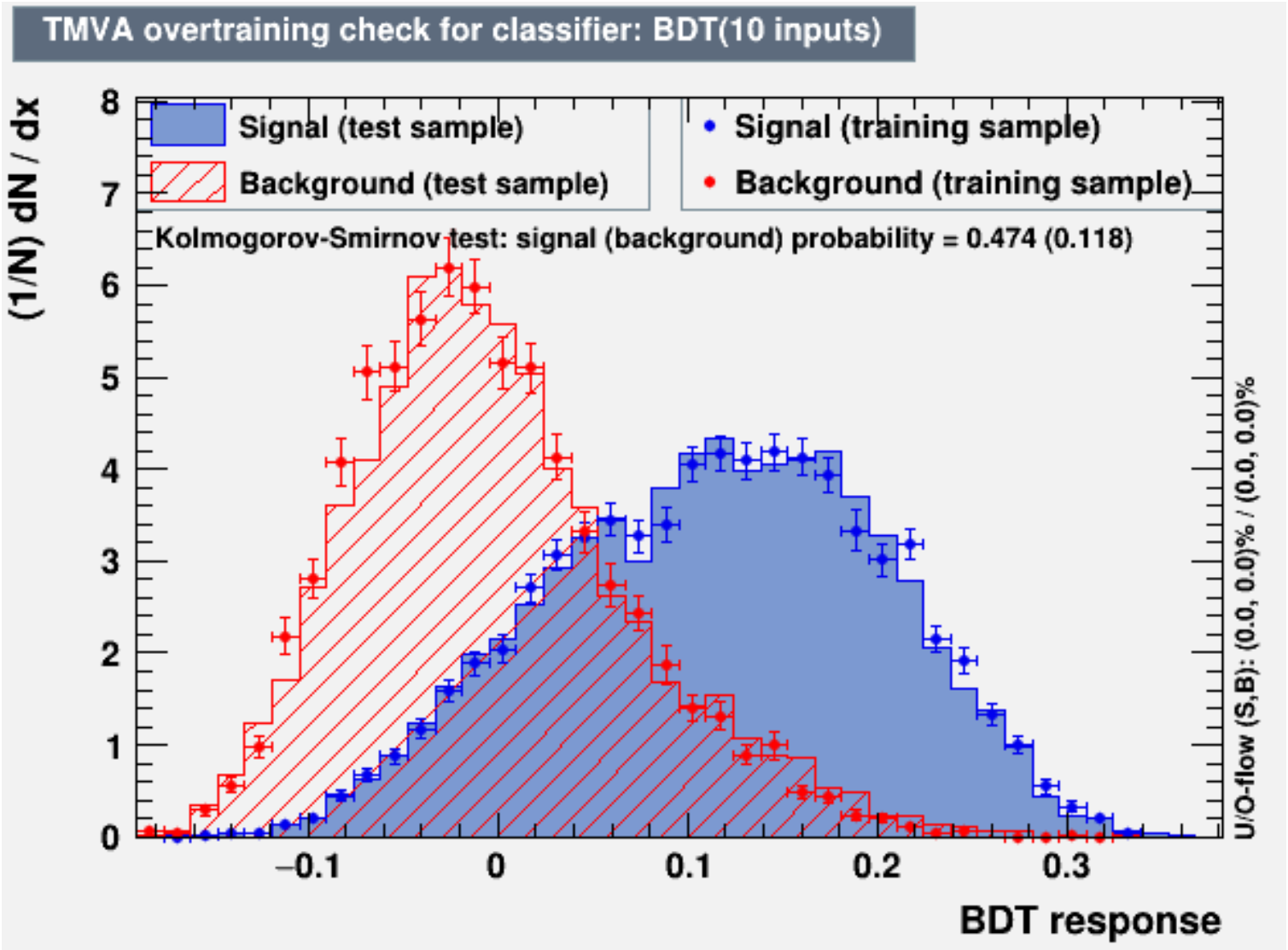


Significance as a function of nVar



- The point of 9 variables could be a candidate working point.
- The binning could be 5 bins, [-0.2,-0.1,0.0,0.1,0.2,0.4]

Backup


TMVA overtraining check for classifier: BDT(10 inputs)
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.474 (0.118)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDT response
0
1
2
3
4
5
6
7
8
−0.1
0
0.1
0.2
0.3