Analysis Tools: Lepton ID filter

Mariana Tenorio Pita
CLAS Collaboration
June 2024

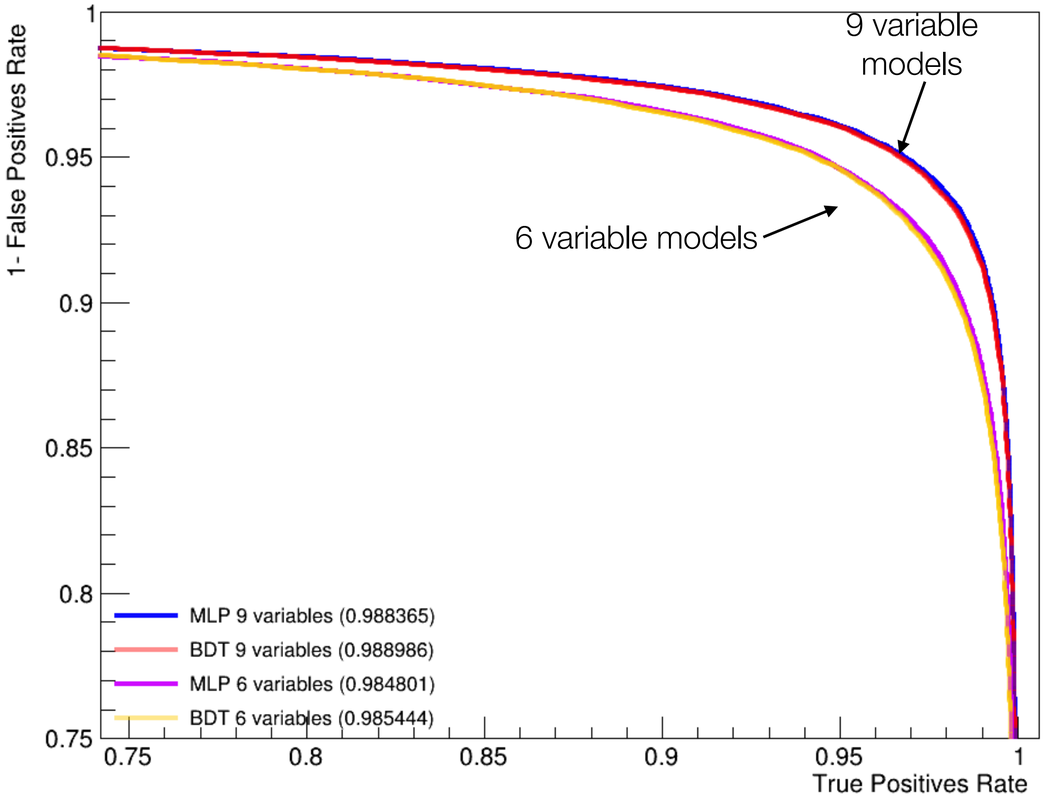
OLD DOMINION
UNIVERSITY

Lepton ID

- Enhance Lepton identification to prevent pion contamination.
- Uses machine learning models, trained on simulation.
- Validation on data showed that by retaining around 90% of leptons in the samples, the background can be reduced by a factor of 10.
- After the training we obtain for each model a weight file that containing the results of the training.

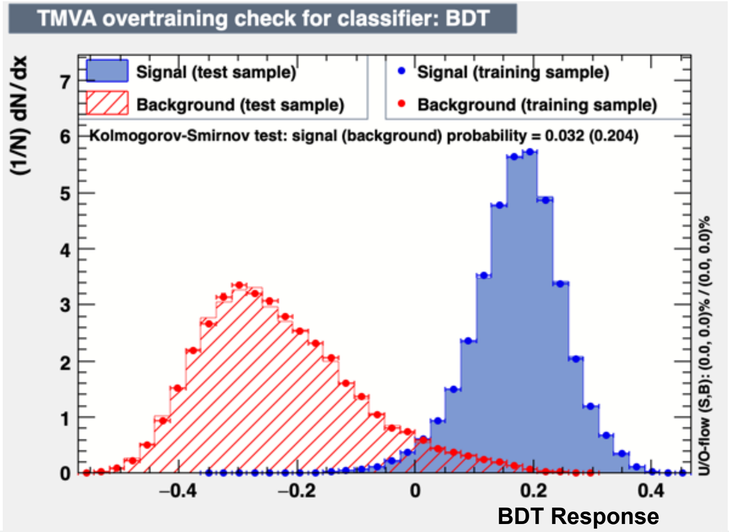
Lepton ID

- Classifiers: e^+ and e^- identification on each Pass2 RGA configuration:
  - Fall 18 Inbending (10.6 Gev)
  - Fall 18 Outbending (10.6 GeV)
  - Spring 19 Inbending (10.2 GeV)
- Methods tested:Multilayer Perceptron (MLP) and Boosted Decision Trees (BDT, default)
- Variables used: P, θ, ϕ, SF and m2 of PCAL, ECIN and ECOUT.
- 6 and 9 (default) Variable models.



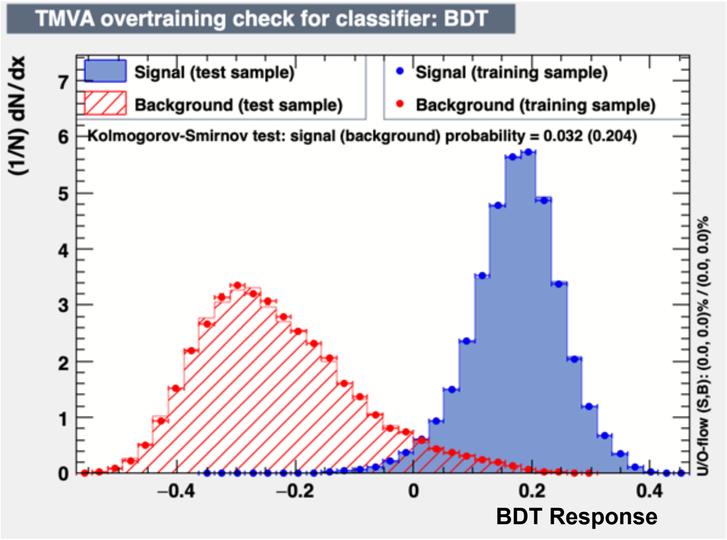
Lepton ID

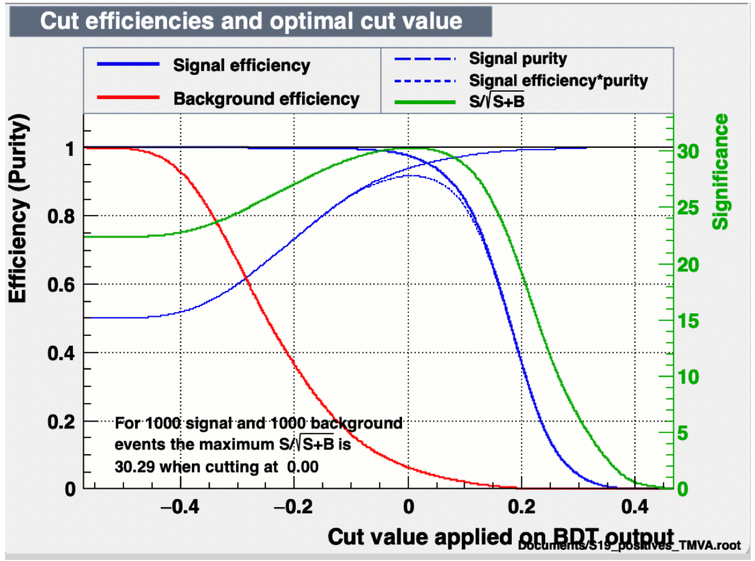
- The algorithm reads REC::Particle and REC::Calorimeter data banks for particles with PID=11/-11
- For each particle, it saves the necessary variables and then calls the TMVA reader.
- The output is a "score" value.
- We can apply a cut to discard any particles under certain value. Default cut is 0.0

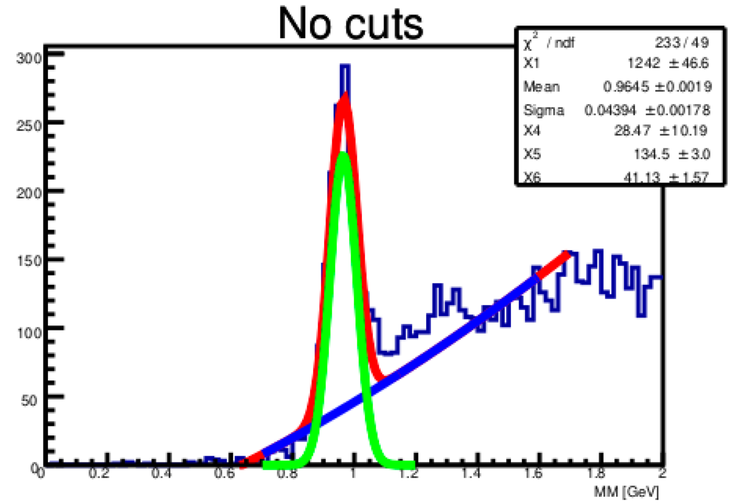


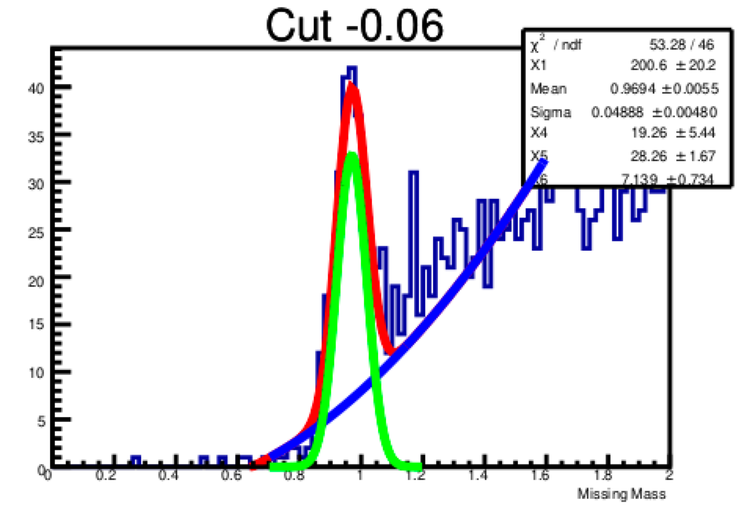

```
//Get TMVA reader
TMVA::Reader *readerTMVA = new TMVA::Reader( "!Color:!Silent" );
  // Create a set of variables and declare them to the reader
Float_t P, Theta, Phi, PCAL,ECIN,ECOUT,m2PCAL,m2ECIN,m2ECOUT;

readerTMVA->AddVariable( "P",&P );
readerTMVA->AddVariable( "Theta",&Theta);
readerTMVA->AddVariable( "Phi",&Phi);
readerTMVA->AddVariable( "SFPCAL",&PCAL);
readerTMVA->AddVariable( "SFECIN",&ECIN);
readerTMVA->AddVariable( "SFECOUT",&ECOUT );
readerTMVA->AddVariable( "m2PCAL",&m2PCAL);
readerTMVA->AddVariable( "m2ECIN",&m2ECIN);
readerTMVA->AddVariable( "m2ECOUT",&m2ECOUT);

m_log->Debug("Add variables to readerTMVA");

readerTMVA->BookMVA( "BDT", o_weightfile );

m_log->Debug("TMVA method booked");

lepton.score=readerTMVA->EvaluateMVA("BDT");
```

Lepton ID


TMVA overtraining check for classifier: BDT
Signal (test sample)
Signal (training sample)
Background (test sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.032 (0.204)
(1/N) dN/dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDT Response


Cut efficiencies and optimal cut value
Signal efficiency
Background efficiency
Signal purity
Signal efficiency*purity
S/√S+B
Efficiency (Purity)
Significance
For 1000 signal and 1000 background
events the maximum S/√S+B is
30.29 when cutting at 0.00
Cut value applied on BDT output
Documents/S19_positives_TMVA.root


No cuts
χ² / ndf 233 / 49
X1 1242 ±46.6
Mean 0.9645 ±0.0019
Sigma 0.04394 ±0.00178
X4 28.47 ±10.19
X5 134.5 ±3.0
X6 41.13 ±1.57
MM [GeV]


Cut -0.06
χ² / ndf 53.28 / 46
X1 200.6 ±20.2
Mean 0.9694 ±0.0055
Sigma 0.04888 ±0.00480
X4 19.26 ±5.44
X5 28.26 ±1.67
X6 7.139 ±0.734
Missing Mass