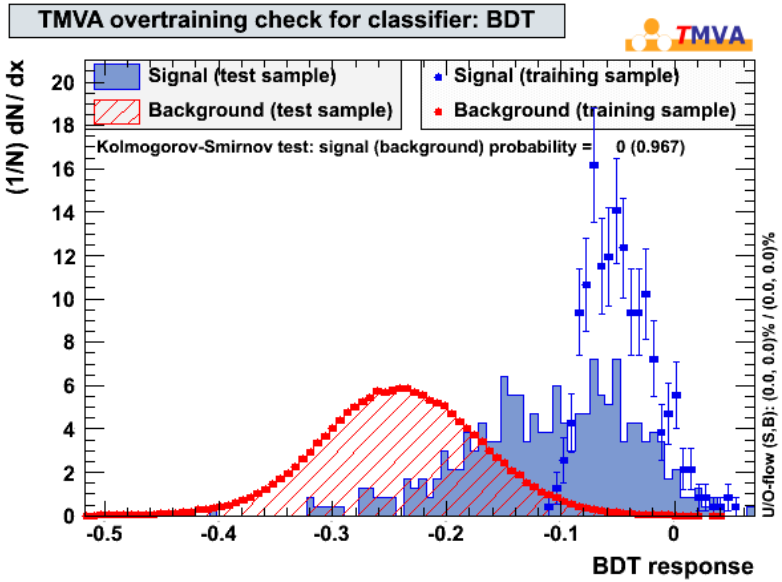
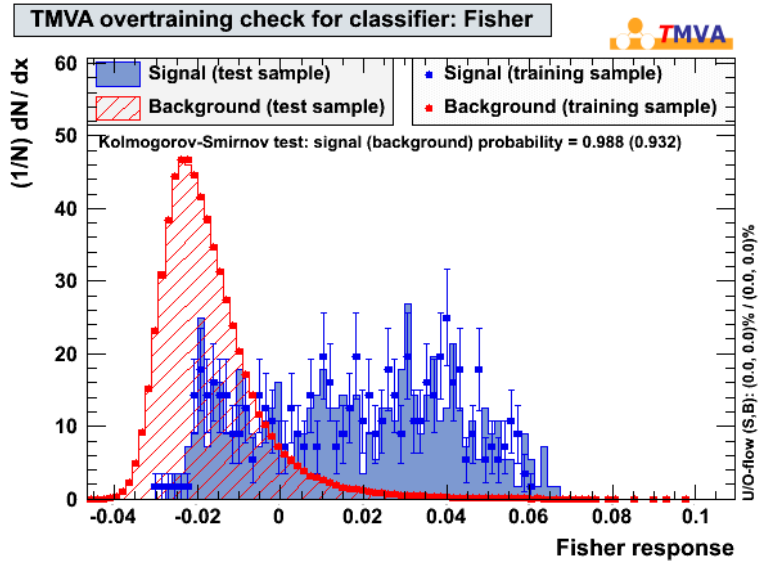
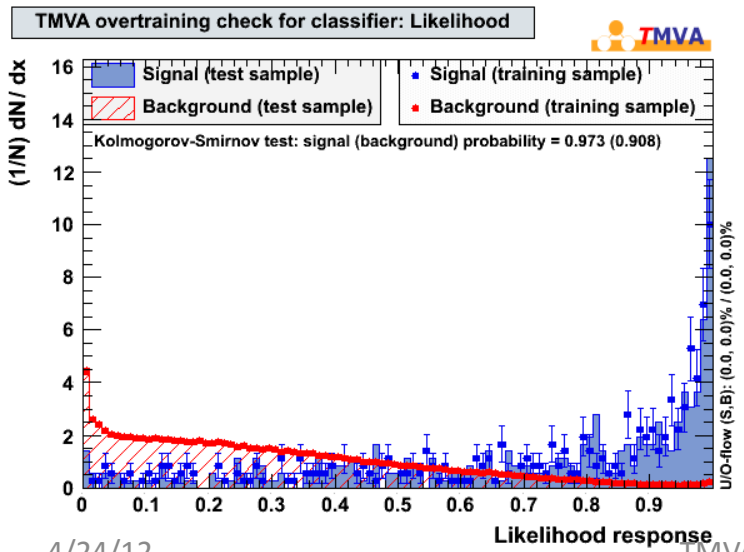
TMVA : embedding sample

- Goal : check that a cut over a given value of a classifier really selects signal, signal+back or back according to the training distributions.
- Signal sample : D0 embeding data (~700 pairs) correlated with the minimc files (directory 700_102/)
- Background : corresponding ++,-- pairs (700_102)
- Classifiers : BDT, Likelihood, Fisher

Training distribution : Classifiers output

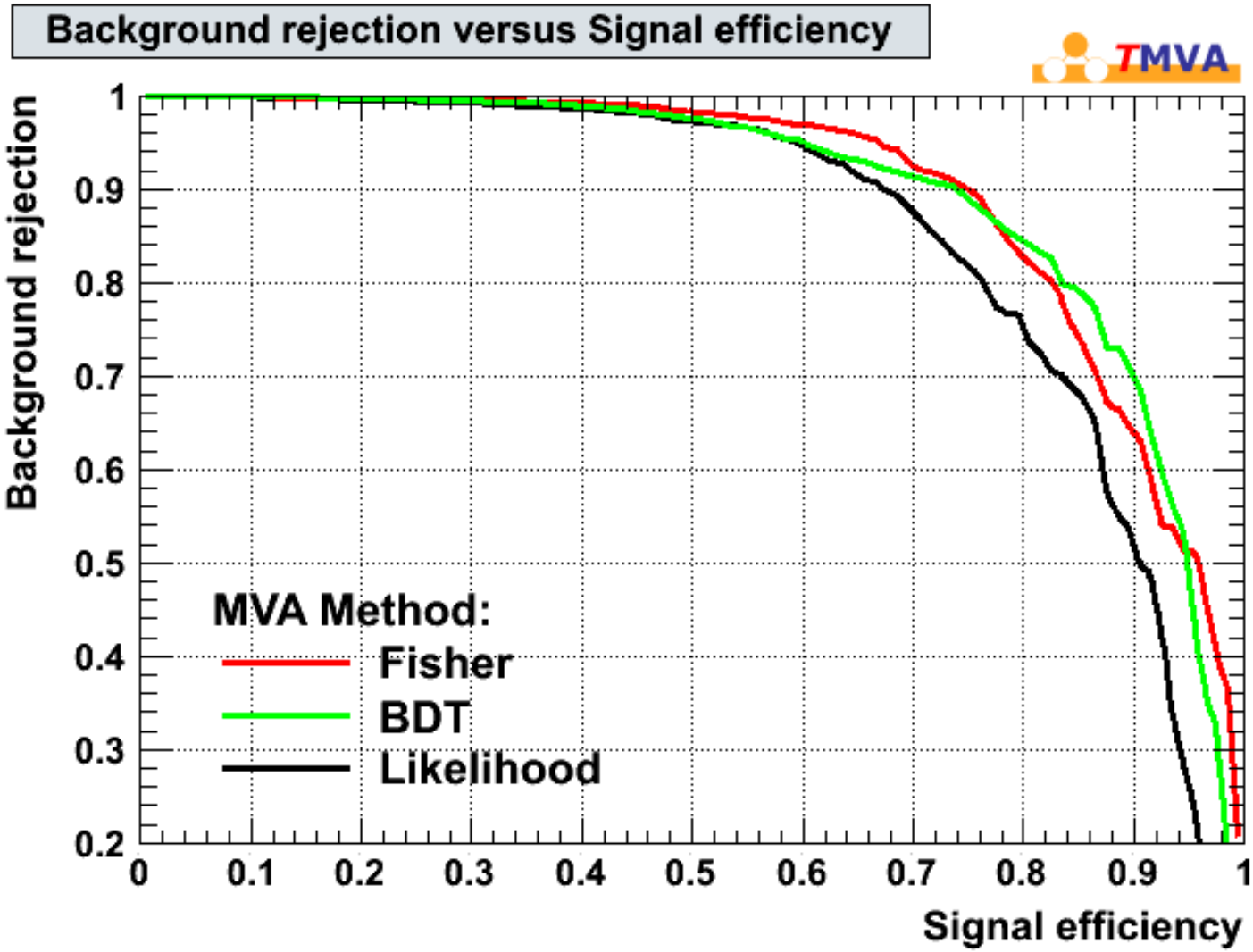






➔Overtraining due to low stats (signal) for the BDT classifier

Classifiers eff.

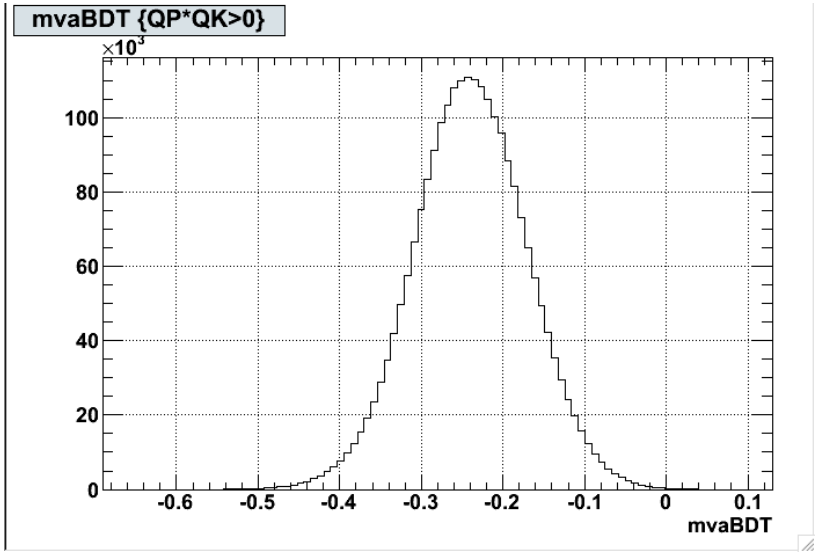


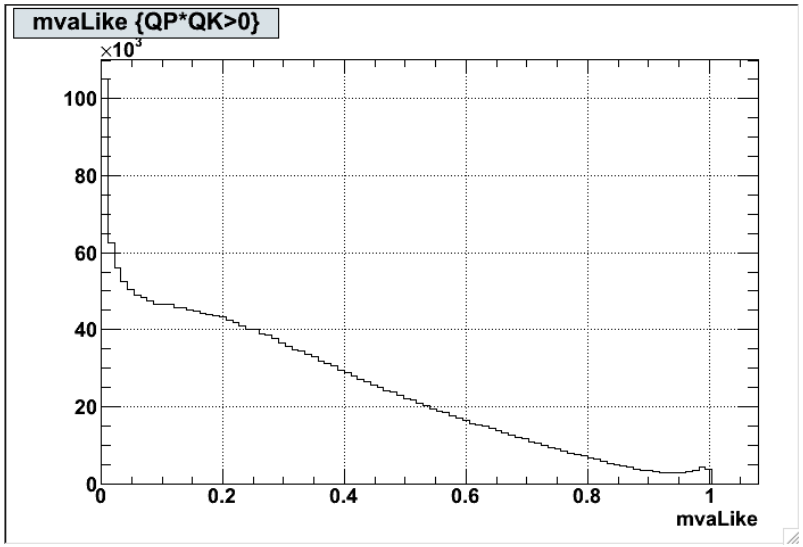
Results from training

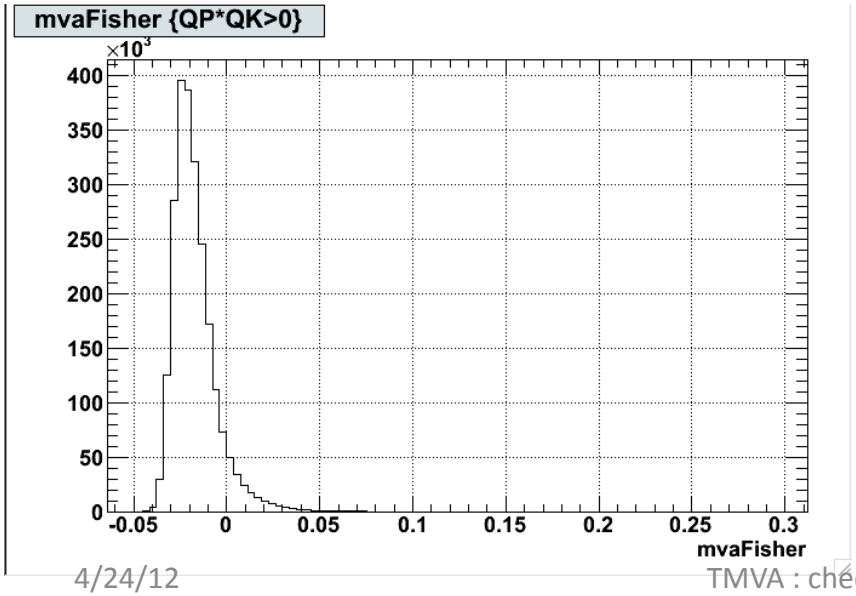
- TMVA says :

```
--- Factory          : An event is signal-like, if its MVA output exceeds the following value:
--- Factory          : -----------------------
--- Factory          :    Method:  Cut value:
--- Factory          : -----------------------
--- Factory          : Likelihood:    +0.469
--- Factory          :    Fisher:      -0.009
--- Factory          :       BDT:      -0.176
--- Factory          : -----------------------
--- Factory          : which correspond to the working point: eff(signal) = 1 - eff(background)
```

analysis

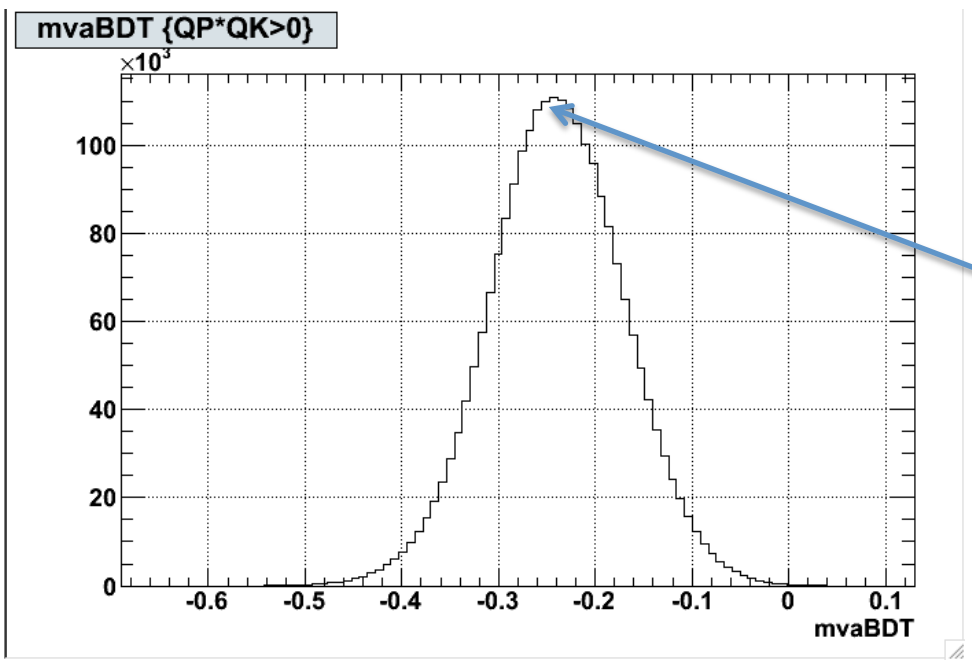


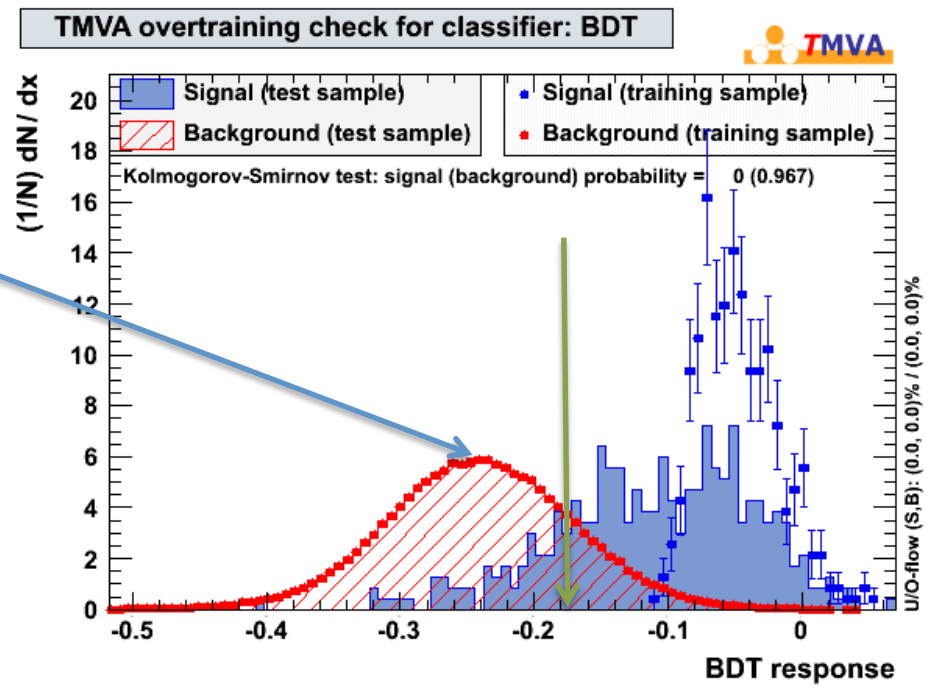




- rerunning over the embedding sample and selecting like-sign pairs **should** give back the background training sample (= red curve)
- This is mandatory *to trust* a cut on TMVA

mvaBDT

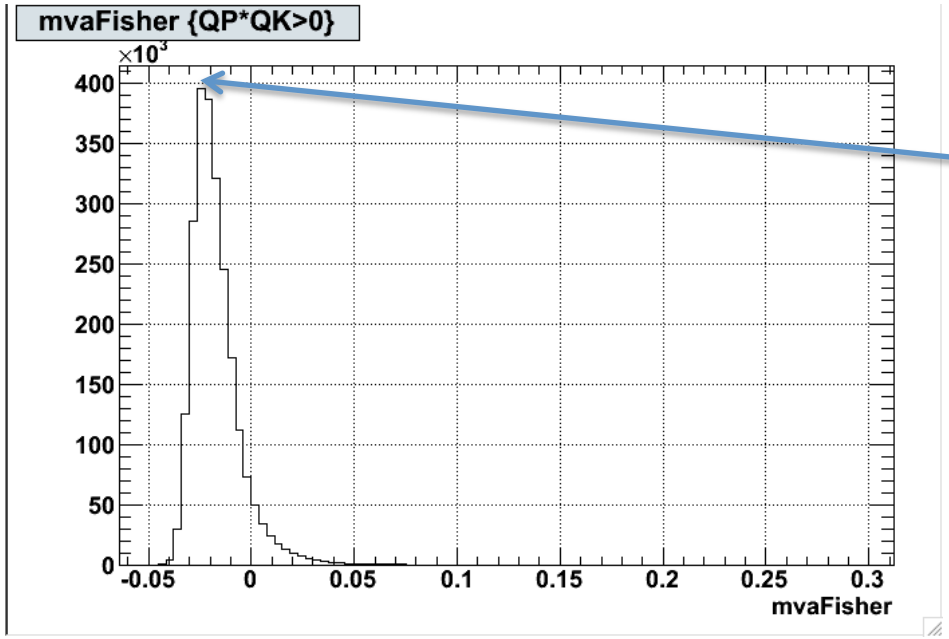


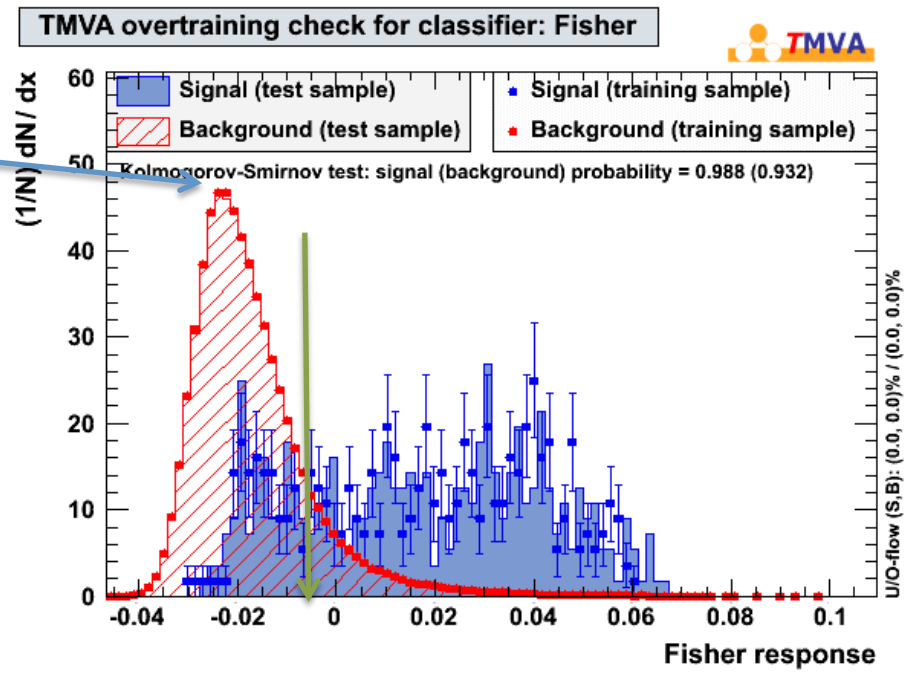


✓maxima and ranges match

- For BDT, the cut that optimizes the S/B is mvaBDT >-0.176
- But we see that it still has some background contribution
- A “background-free” cut would be ~0

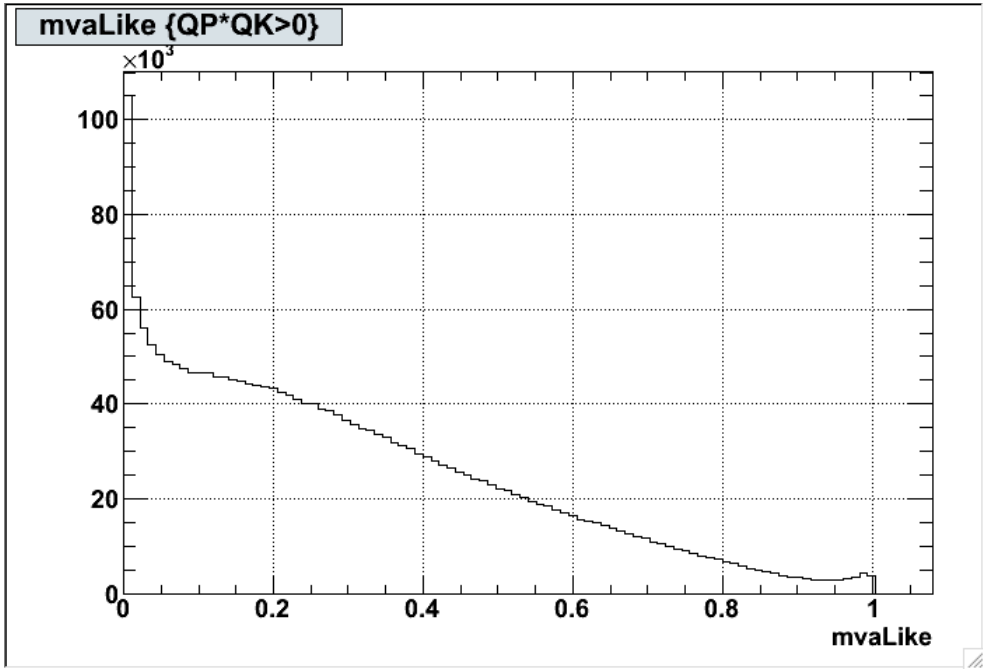
mvaFisher

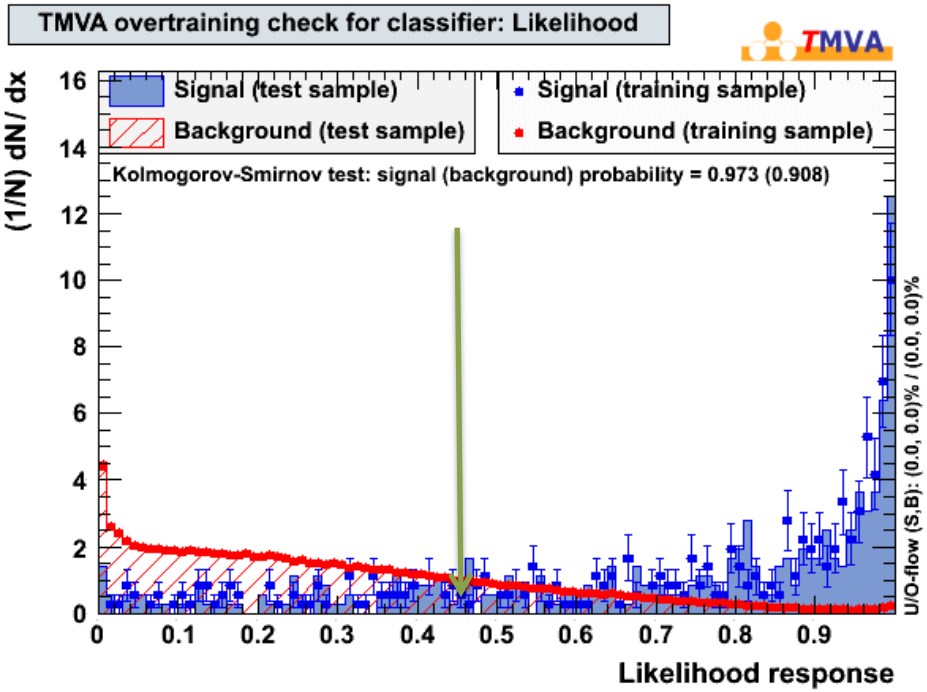




- ✓Looks agree
- ✓Should check in log scale
- ✓The optimum cut (mvaFisher> -0.009) again selects some background
- ✓A background-free cut would be ~0.02

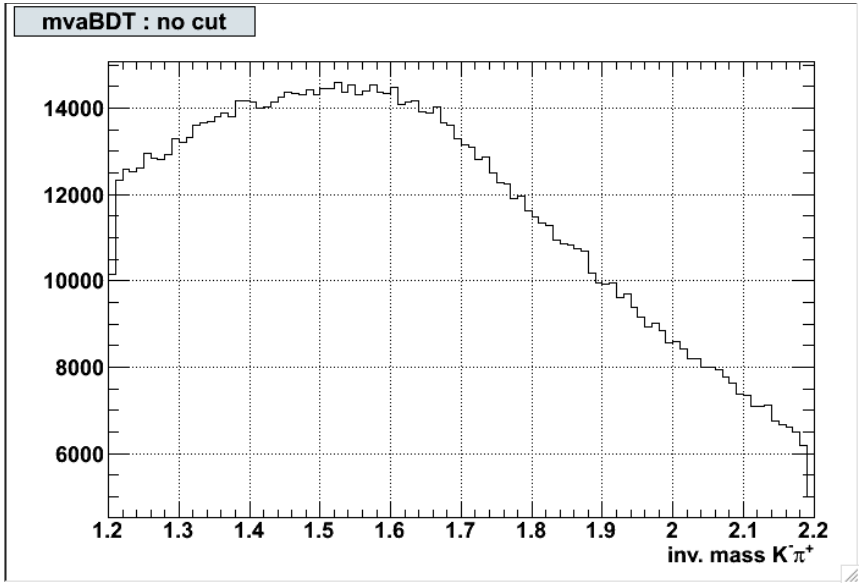
mvaLike

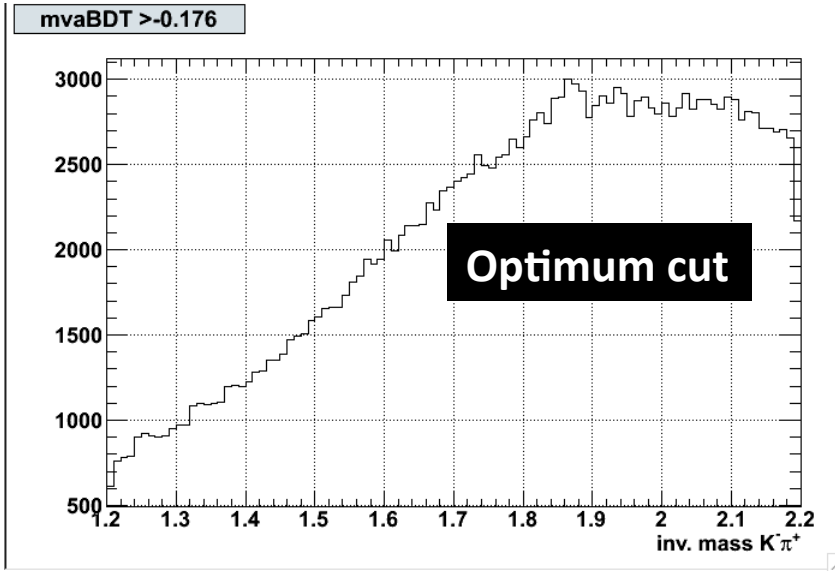




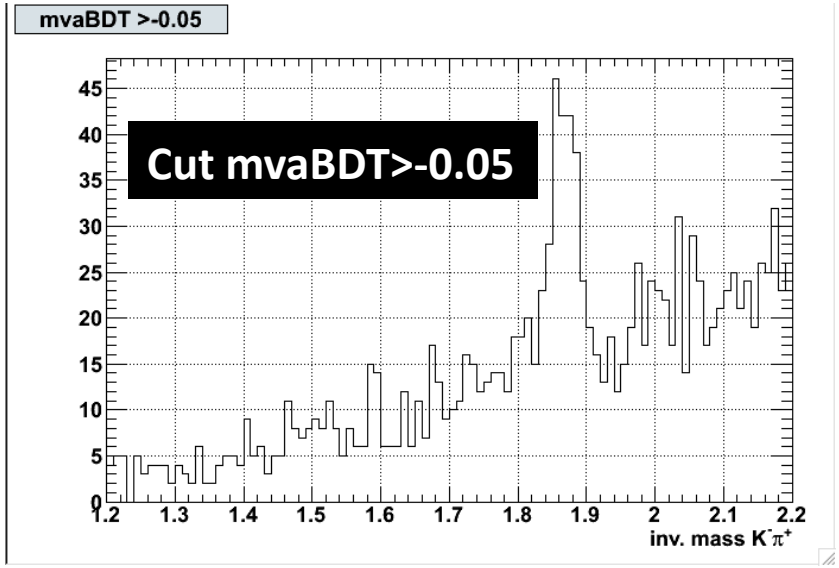
- ✓maxima and ranges match
- ✓The optimum cut (mvaLike> 0.469) again selects some background
- ✓A background-free cut would be ~0.9

For mvaBDT

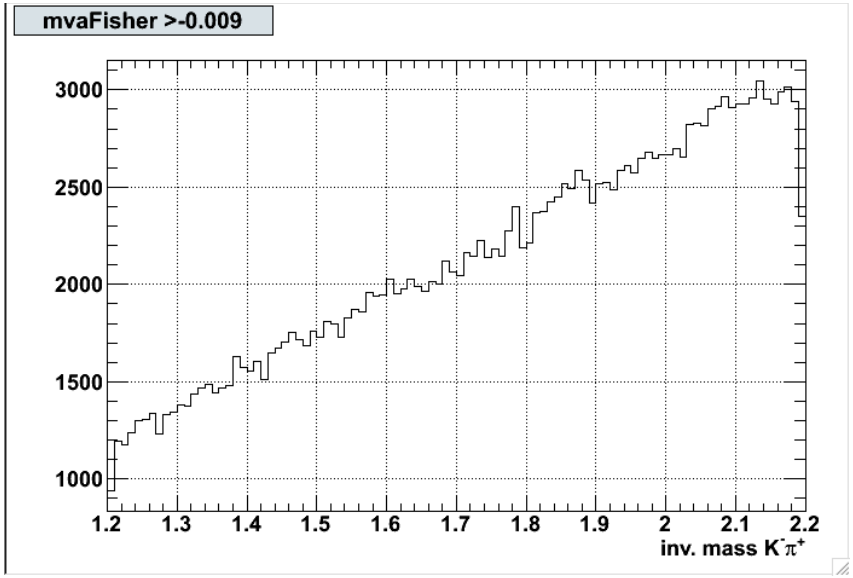

mvaBDT : no cut
14000
12000
10000
8000
6000
1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2
inv. mass K⁻π⁺


mvaBDT >-0.176
3000
2500
2000
1500
1000
500
Optimum cut
1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2
inv. mass K⁻π⁺

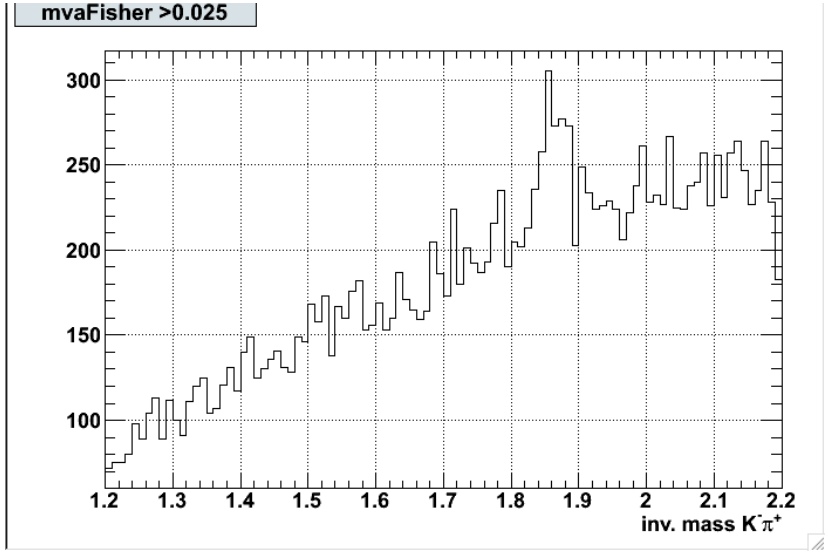
mvaBDT >-0.1
500
450
400
350
300
250
200
150
100
1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2
inv. mass K⁻π⁺


mvaBDT >-0.05
45
40
35
30
25
20
15
10
5
0
Cut mvaBDT>-0.05
1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2
inv. mass K⁻π⁺

For mvaFisher



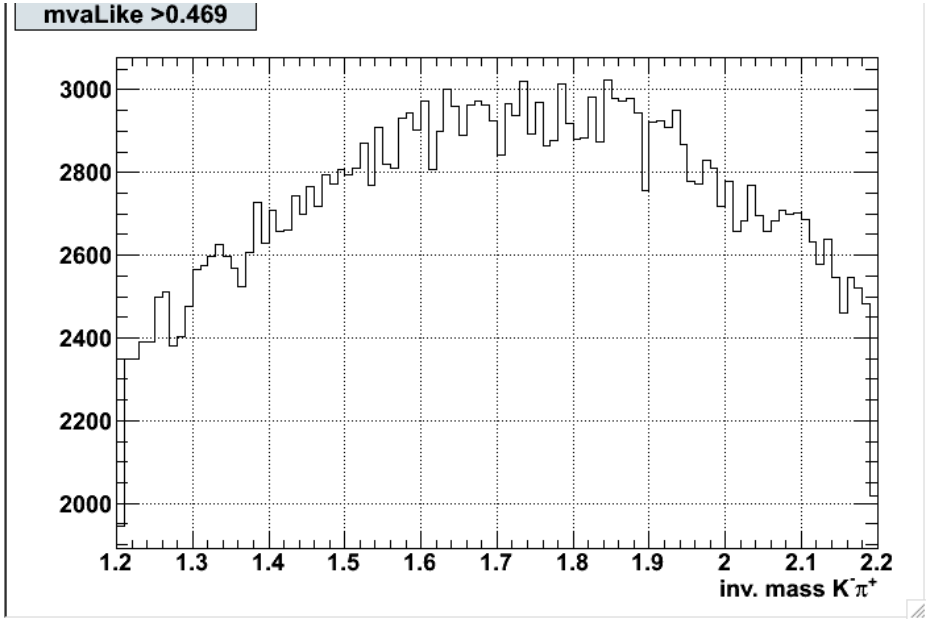
Optimum cut



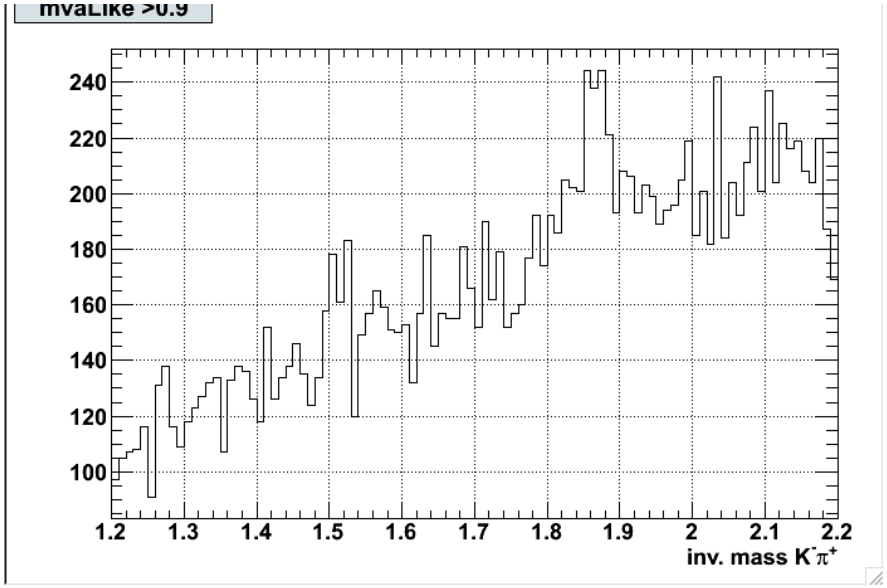
Cut mvaFisher>0.025

- The optimal cut does not show a peak.
- not as good as with mvaBDT but cutting on higher value also gives a clear peak

For mvaLike



Optimum cut



Cut mvaLike>0.9

summary

- I think these are good results :
  - even the “optimal” cut does not show clearly a peak, we can trust a cut on a classifier because the distribution after training and after rerunning with the training results give same background distribution
  - Cutting on slightly higher mva responses show clearly the D0 peak
  - It might be optimized with a different training (more variables, more stats)
  - It might be optimized by choosing another classifiers
- It might not be as this good in real life because the D0 here have flat Pt
- We also see the background reduction when applying the BDT cut (slide 9, upper left = no cut , bottom right cut >-.05)
- As a proof of principle of TMVA, it looks good.