

Multivariate Analysis Techniques

Exercise Session: Example Results

BABAR Analysis School
SLAC National Laboratory
25th-30th October 2009

Adrian Bevan (a.j.bevan@qmul.ac.uk)

Comparing classifiers

- Note these plots use the default variable combination as set up in `tmva_example.cc`. If you modified the variables/methods to train, then your results may differ.
 - All methods are trained, and the weight files are written to a sub-directory `weights` that should be created in your current directory.
 - In addition to this you should have a ROOT file `TMVA.root` that contains the necessary information required to inspect the classifier performance.
 - You can plot the correlation histograms in `TMVA.root` by running the `plotCorrelations.cc` macro.
 - You can compare the separation between background and signal by running the `plotEfficiencies.cc` macro.

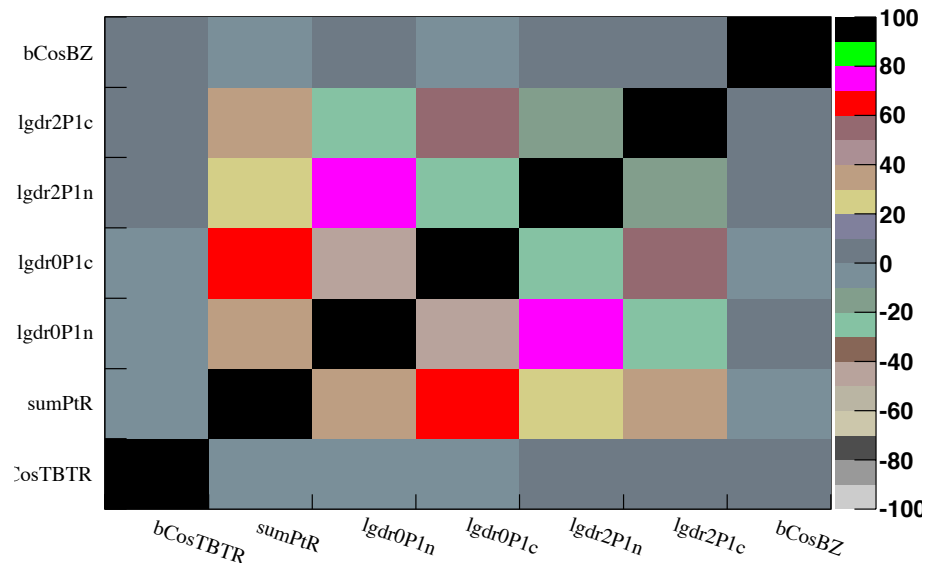
Correlations between variables

- Pearson correlation coefficients are plotted below

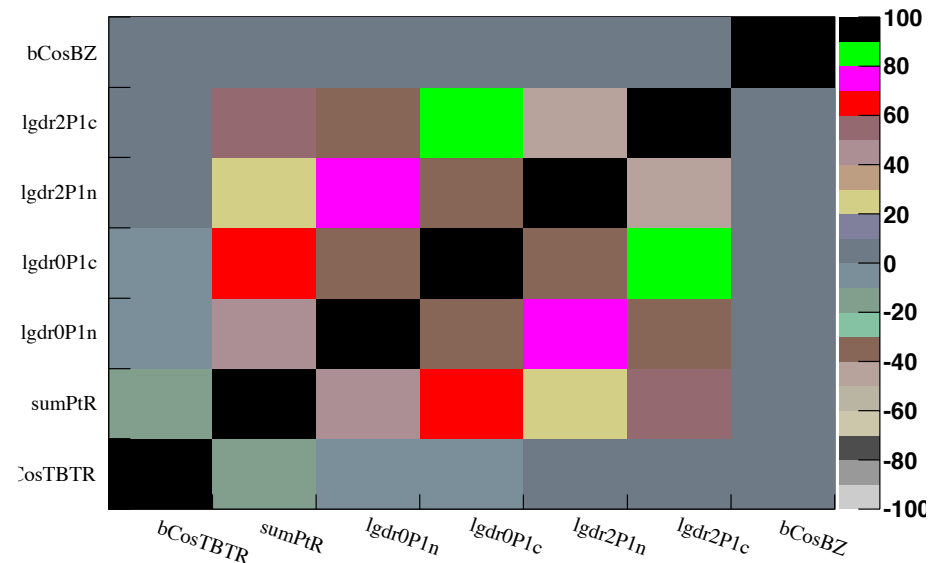
Signal

Background

Correlation Matrix (signal)



Correlation Matrix (background)

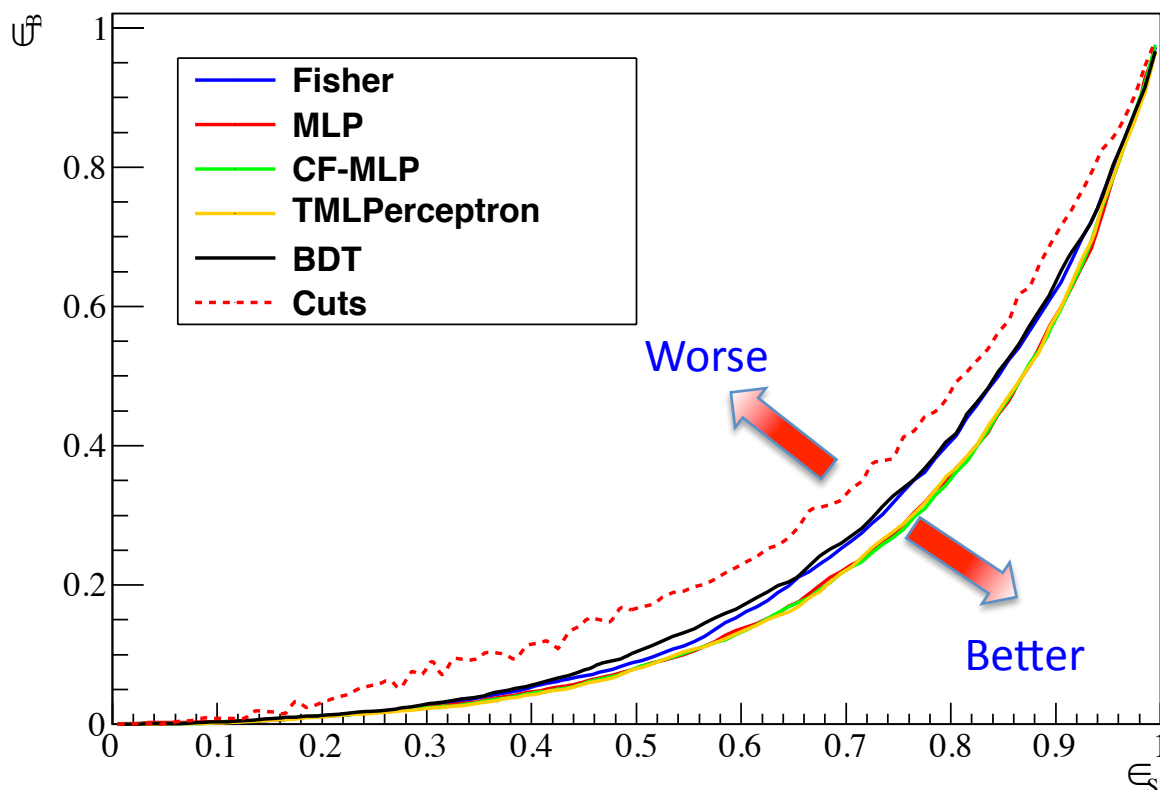


- Some variables are highly correlated...
 - Q) will they all be useful in the MVA?

Separation of signal and background

- The plot produced by `plotEfficiencies.cc` is:

MVA Performance Comparison



There are 3 levels of separation:

Worst is the cuts based method.

Fisher and BDT gave a similar response.

The three MLP algorithms gave a similar response, and have the best signal/background separation.

Q) Are all variables required?

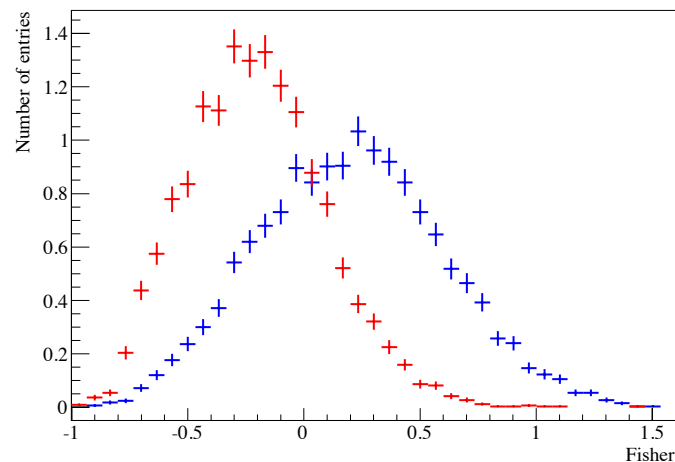
Q) Are other classifications or configurations better?

Q) What do these distributions look like?

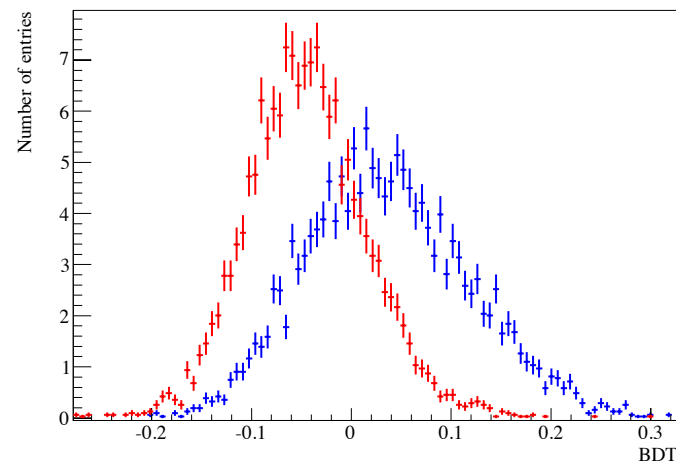
What do the distributions look like?

- The macro `plotMVADistributions.cc` can be used to plot the distributions of the MVA that you've obtained.

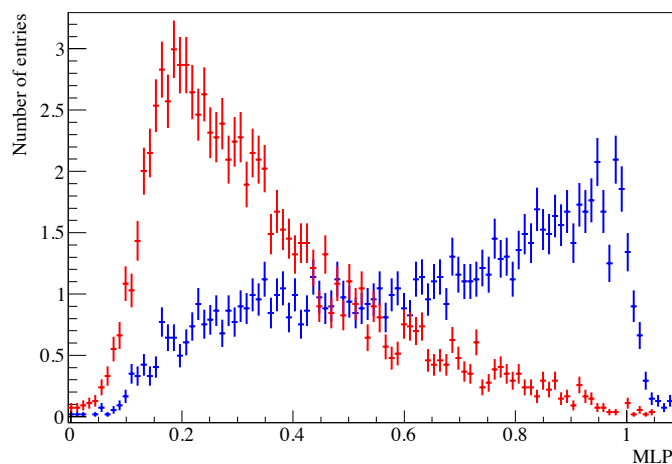
MVA distributions for the Fisher classifier



MVA distributions for the BDT classifier



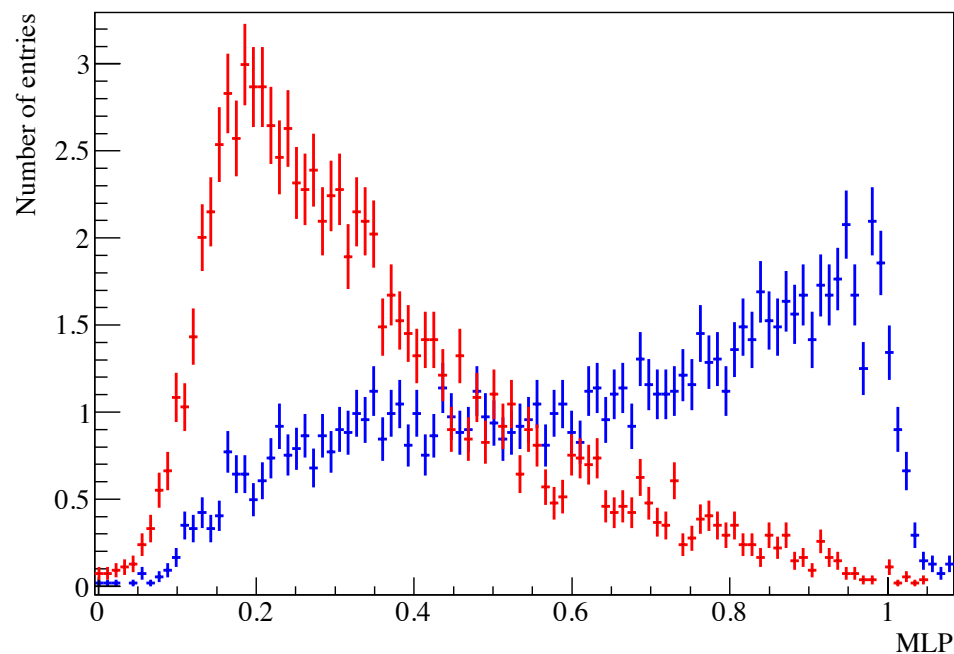
MVA distributions for the MLP classifier



What do the distributions look like?

- What if we want to use the MLP in a fit to data?

MVA distributions for the MLP classifier



Can try to parameterise this shape with a non-parametric PDF.

Can try to transform the shape by a 1:1 mapping into something a little easier to parameterise.

Can use a mapping function such as:

$$\mathcal{N} = 1 - \arccos(MLP + \xi)$$

What are the powerful discriminating variables?

- You need to look at the information printed to the standard output for this.
 - e.g. for the Fisher classifier:

```
--- TFHandler_Factory      : Ranking input variables...
--- IdTransformation      : Ranking result (top variable is best
ranked)
--- IdTransformation      : -----
--- IdTransformation      : Rank : Variable  : Separation
--- IdTransformation      : -----
--- IdTransformation      :      1 : bCosTBTR   : 1.412e-01
--- IdTransformation      :      2 : lgdr2P1n   : 7.891e-02
--- IdTransformation      :      3 : lgdr2P1c   : 6.519e-02
--- IdTransformation      :      4 : bCosTBZ    : 2.674e-02
--- IdTransformation      :      5 : lgdr0P1n   : 2.314e-02
--- IdTransformation      :      6 : bCosBZ     : 2.143e-02
--- IdTransformation      :      7 : lgdr0P1c   : 9.404e-03
--- IdTransformation      :      8 : sumPtR     : 8.101e-03
```

Summary #3

- The problem of distinguishing between signal and background classes of events is a relatively simple one.
 - Only two target classes (straightforward to extend).
 - Covers many situations we encounter in real life (as often one background dominates over all others).
 - ... and you've now done this!
- The computation of an MVA is simplified greatly by the tools of the trade now available in High Energy Physics.
- As the saying goes: '*A poor workman always blames his tools*'. Don't use an algorithm if you don't understand what it is and what the limitations of the method are.
 - This warning is equally valid for both financial institutions and particle physicists.
- Hopefully these lectures have provided you with an introduction to the subject and given you a few things to think about.