iSTEP 2014, Beijing
TMVA Exercise

For this exercise you will do a simple multivariate analysis with the TMVA package together with ROOT routines. For instructions go to

`www.pp.rhul.ac.uk/~cowan/stat/root/tmva`

and read the file `readme.txt` to find how to build the programs in the subdirectories `generate`, `train` and `analyze`.

First, use the program `generateData` to generate two n-tuples of data whose values follow a certain three-dimensional distribution for the signal hypothesis and another for the background hypothesis. (The n-tuples are created and stored using the ROOT class `TTree`.) Using the macro `plot.C`, take a look at some of the distributions (run root and type `.X plot.C`).

Then use the program `tmvaTrain` to train a Fisher discriminant and a neural network (multilayer perceptron). When you run the program, the coefficients of the discriminating functions are written into a subdirectory `weights` as text files. You can take a look at these files and see the relevant coefficients.

Finally use the program `analyzeData` to analyze the generated data. Suppose you want to select signal events, and that the prior probabilities of signal and background are equal. Suppose you select signal events by requiring $t_{\rm Fisher} > 0$. What are the signal and background efficiencies? What is the signal purity? (Insert code into `analyzeData.cc` to count the number of signal and background events that are selected.)

Make histograms of $t_{\rm Fisher}$ for both signal and background events. (You can superimpose two histograms on the same plot by using `h1->Draw(); h2->Draw("same");`).

Modify the programs `tmvaTrain.cc` and `analyzeData.cc` to include a multilayer perceptron with one hidden layer containing 3 nodes. To book the multilayer perceptron you need a line of the form:

```
factory->BookMethod(TMVA::Types::kBDT, "BDT", "NTrees=200:BoostType=AdaBoost");
```

See the TMVA manual for more details. This will store the coefficients of the classifier in a file in the `weights` subdirectory.

Next to analyze the data using the multilayer perceptron, you will need to add a call to `reader->BookMVA` using the corresponding names (replace `Fisher` with `BDT`). Then book and fill two more histograms for the BDT statistic under both the signal and background hyothesis (do this in analogy with the histograms for the Fisher discriminant). Make plots of the resulting histograms.

Finally, select signal events by requiring $t_{\rm BDT} > 0$. What are the signal and background efficiencies? What is the signal purity assuming equal prior probabilities for the two event types?

G. Cowan