Exercises in Multivariate Methods for IDPASC School

For these exercises you will do a simple multivariate analysis with the TMVA package together with ROOT routines. First download the code in the subdirectories `generate`, `train`, `analyze` and `inc` from

`ttp://www.pp.r`ul.ac.uk/ cowan/stat/root/tmva/

Alternatively download the tarball `tmvaExamples.tar` to your work directory and type `tar -xvf tmvaExamples.tar`. To build the programs, type `gmake`. The ROOT libraries need to be installed; if this does not work then please ask for help.

First, use the program `generateData` to generate two data files, `trainingData.root` and `testData.root`. Each file contains 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 determine the coefficients of a Fisher discriminant. When you run the program, TMVA uses the data from the file `trainingData.root` to determine the parameters of a multivariate classifier, and the values are written into files stored into a subdirectory `weights` as xml files. Take a look at these files and identify 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::kMLP, "MLP", "H:!V:HiddenLayers=3");
```

This will store the coefficients of the multilayer perceptron in a file in the `weights` subdirectory. You can add further classifiers in a similar way; see the TMVA manual for more details. For example, a boosted decision tree and support vector machine can be created with:

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

factory->BookMethod(TMVA::Types::kSVM, "SVM", "H:C=1.0");
```

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 `MLP`). Then book and fill two more histograms for the MLP 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. Select signal events by requiring $t_{\rm MLP} > 0.5$. What are the signal and background efficiencies? What is the signal purity assuming equal prior probabilities for the two event types?

G. Cowan
RHUL Physics
May, 2013