Exercise 9: Multivariate analysis

N. Berger (`nberger@physi.uni-heidelberg.de`)

10.12.2012

Please send your solutions to `nberger@physi.uni-heidelberg.de` until 17. 12. 2012, 12:00. Put your answers in an email (subject line *SMIPP:Exercise09*).

1. **TMVA** Root comes with an extensive package for multivariate analysis called TMVA (`tmva.sourceforge.net`. On the CIP-Pool machines, TMVA is installed. To test and explore it, copy the directory `/opt/root-5.28/tmva/test` into a location that is writable by you. Change into the directory and run
   `> root -l TMVAClassification.C\(\"Fisher\"\)`.
   You should get a window with lots and lots of buttons, each of which either produces plots or another set of buttons producing even more plots for the test sample that comes with TMVA. Have a look at them and try to find out what they mean (there is extensive documentation on the TMVA website).

2. **TMVA II** Run `> root -l TMVAClassification.C`. This will train and run all classifiers available in TMVA and take about 15 minutes or so. At 90% efficiency, which classifier has the best purity?

3. **Signal and background again** On the course website, there is a pure background and a pure signal file for a resonance search in a three body decay. The signal shows a nice resonance (on which we do not want to cut). In addition, there are three measured variables for each decay particle, namely the energy loss per path length in a drift chamber, a relative time-of-flight measurement and the energy fraction deposited in a calorimeter. Explore the trees in interactive root. Modify `TMVAClassification.C` to use these samples (use the mass as a spectator variable and all others as variables to cut on) and train a few classifiers. How easy do you find it to change little bits and pieces in a large package written by someone else? Now use the classifiers to classify the data sample from the website (you can do this fairly easily by using the data tree as the test sample - classifier values will be saved to the `TMVA.root` output file. Here of course outputs like ROC curves become meaningless.
   (Attach the modified `TMVAClassification.C` file and a few representative plots - especially a mass spectrum.)