

69703 - ANALISI STATISTICA DEI DATI NELLA FISICA NUCLEARE E SUBNUCLEARE

Modulo 3 : Laboratorio di analisi statistica per la fisica delle alte energie

docente: G. Sirri

<http://www.unibo.it/docenti/gabriele.sirri2>

ESERCIZIO 4 : TMVA

23.04.2015

da spedire via mail entro Giovedì 30.04.2015

Per favore spedite la soluzione via mail a gabriele.sirri2@unibo.it entro la data di scadenza in modo che possiamo discutere la soluzione durante la lezione del giorno successivo.

Assicuratevi che l'oggetto della mail sia: "**ASD-2015:Esercizio04 COGNOMI**".

E' importante che tutti i vostri programmi e macro siano stati testati prima di spedirli.

Per favore commentate il codice in maniera ragionata e comprensibile.

E' raccomandato il lavoro in gruppi di due, ma **non dimenticate di dichiarare entrambi i nomi**.

Riferimenti:

- TMVA web site: <http://tmva.sourceforge.net>

- TMVA manual: TMVA - Toolkit for Multivariate Data Analysis , [arXiv:physics/0703039v5](https://arxiv.org/abs/physics/0703039v5) [[physics.data-an](https://arxiv.org/abs/physics/0703039v5)]

- TMVA workshop (<http://indico.cern.ch/conferenceDisplay.py?confId=112879>)

In particolare "My tips and tricks"

<http://indico.cern.ch/contributionDisplay.py?contribId=5&confId=112879>

- <http://www.unibo.it/docenti/gabriele.sirri2> --> Contenuti utili --> Analisi stat... Calendario e Materiale

We will use the TMVA version provided by root 5.34 which provide a set of useful root macros located in `$ROOTSYS/tmva/test/`

For each exercise, you must create a separate working folder.

You can **load** the macros directly from the default folder, in this way :

```
root -l $ROOTSYS/tmva/test/TMVAClassification.C\ ("LD,MLP,BDT\"")
root -l $ROOTSYS/tmva/test/TMVAGui.C
```

It is highly recommended to not modify the macros located in `$ROOTSYS/tmva/test/` .When you are requested to modify something (for example `TMVAClassification.C`) please copy the macro in your working folder .

You can copy all macros from `$ROOTSYS/tmva/test/` to the working folder if needed.

Don't run the macro in the default folder.

Note: for Windows is `.../test/TMVAClassification.C ("LD,MLP,BDT")` ,without the ``` before parentheses.

[0] tmva_ex0.C

- Create a working folder “tmva_ex0”
- download http://root.cern.ch/files/tmva_class_example.root to the working folder
- Run your first job using the macro `TMVAClassification.C` .
Train the classifiers LD, MLP, BDT on the test data.
You are not requested to modify the macro. You have just to run it:

```
root -l $ROOTSYS/tmva/test/TMVAClassification.C\(\"LD,MLP,BDT\")
```

- Open `TMVAClassification.C` and have a look to the code. Locate where are defined :
 - i) input variables for the training;
 - ii) spectator variables;
 - iii) signal and background trees;
 - iv) signal and background weights;
 - v) selection cuts on signal and background;
 - vi) number of training and testing events;
 - vii) booking of MVA methods;
 - viii) calls to start of training, testing and method comparison

To use the TMVA collections of macros (and see the output of your training and testing), type :

```
root -l $ROOTSYS/tmva/test/TMVAGui.C
```

(Allegare i .png dei plot che ritenete più significativi)

(SKIP THE FOLLOWING)

Next, run the reader application using `root -l TMVAClassificationApplication.C\(\"LD,BDT\")`

[1] tmva_ex1.C

DA http://hep.fi.infn.it/ciulli/Site/Analisi_Dati.htm

Lo scopo di questo esercizio è fare una semplice analisi multivariata con il pacchetto TMVA di ROOT. Scaricate il tar file con il codice per l'esercizio da:

http://hep.fi.infn.it/ciulli/Site/Analisi_Dati_files/tmvaExamples.tar

quindi in una directory date il comando `tar -xvf tmvaExamples.tar`.

Per prima cosa usate la macro `generateData.C` per generare due n-tuple di dati, i cui valori seguono una distribuzione tridimensionale per il segnale e un'altra per il fondo. La macro `plot.C` può essere usata per guardare le distribuzioni (eseguite `root` e poi dal prompt date il comando `.x plot.C`).

Usate poi la macro `tmvaTrain.C` per determinare i coefficienti del discriminante di Fisher. Questi coefficienti sono scritti in un file nella sotto-directory `weights` come file testo. Guardate il log del comando e il contenuto del file per individuare i coefficienti.

Infine usate `analyzeData.C` per analizzare i dati generati. Supponete che le probabilità a priori di segnale e fondo siano uguali. Quali sono le efficienze per segnale e fondo se richiedete $t_{\text{Fisher}} > 0$? E qual'è la purezza del segnale selezionato con questo taglio? (Modificate il codice `analyzeData.C` inserendo dei contatori per rispondere a queste domande). Scrivete una macro per visualizzare e confrontare gli istogrammi `hFishSig` e `hFishBkg`. Potete partire come esempio dalla macro `plotUniform.C` del problema 1.

Adesso modificate il programma `tmvaTrain.cc` e `analyzeData.C` per includere una rete neurale con uno strato nascosto con 3 nodi. Per creare la rete neurale dovete inserire la linea:

```
factory->BookMethod(TMVA::Types::kMLP, "MLP", "H:!V:HiddenLayers=3");
```

dove "MLP" sta per "Multi Layer Perceptron" (si veda il manuale di TMVA per maggiori dettagli). Anche i coefficienti della rete neurale sono salvati in un file nella sotto-directory `weights`.

Analizzate infine i dati usando la rete neurale. Dovrete aggiungere la chiamata

```
reader->BookMVA;
```

usando il nome corrispondente (rimpiazzate `Fisher` con `MLP`). Create e riempite altri due istogrammi per guardare la distribuzione della statistica MLP per il segnale e il fondo (analogamente agli istogrammi per il discriminante di Fisher). Quali sono le efficienze su segnale e fondo se si richiede $t_{\text{MLP}} > 0.5$? Qual'è la purezza del segnale?