TMVA New Features

Sergei V. Gleyzer

Omar Zapata Mesa

Lorenzo Moneta

Data Science at the LHC Workshop

Nov. 9, 2015

Outline

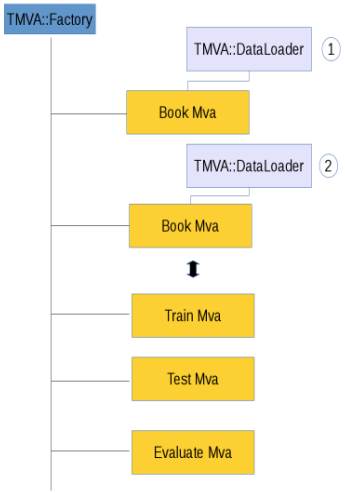
- New TMVA Features
- Data Loader
- Feature Selection
- RMVA
- PyMVA

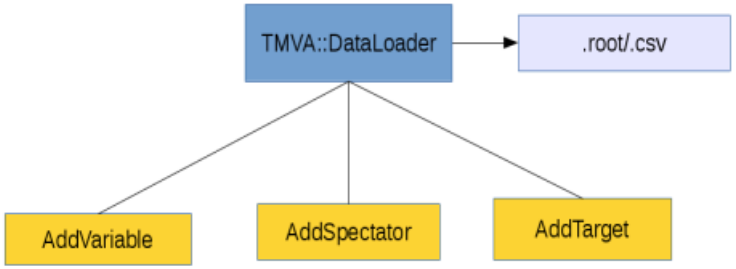
New Features

- Improved design that allows
  - Greater flexibility (modularization)
  - Feature Selection, Cross-Validation, new techniques for data storage and manipulation
  - Parallelization (OpenMP/MPI/cuda/TBB)
  - Integration with Python and R (PyMVA/RMVA)

Data Loader

- **TMVA::DataLoader class allows greater flexibility in dealing with data**
  - **Connection of different features to different classifier methods: Useful for optimization**





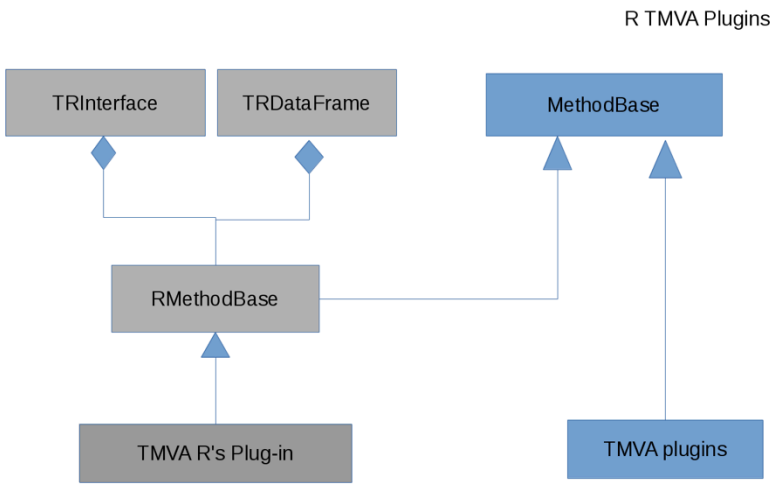

```
factory->BookMethod( DataLoader &,Types::EMVA , TString methodTitle, TString theOption);
```

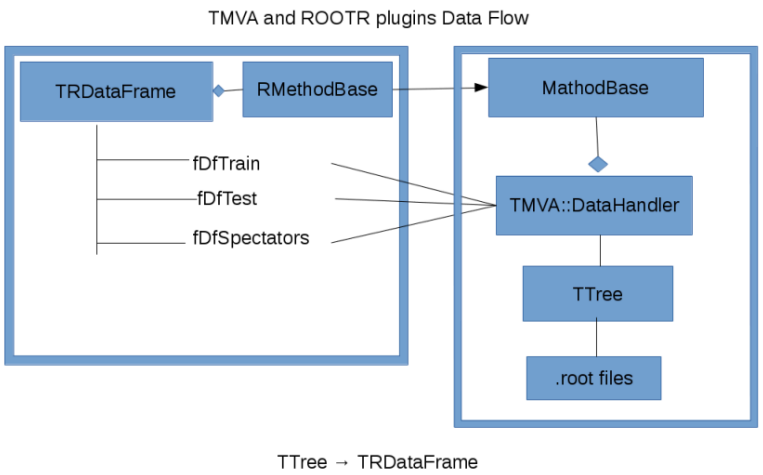
Feature Selection

- **Based on the FAST stochastic wrapper algorithm**
  - **See previous talk and today's tutorial for details**

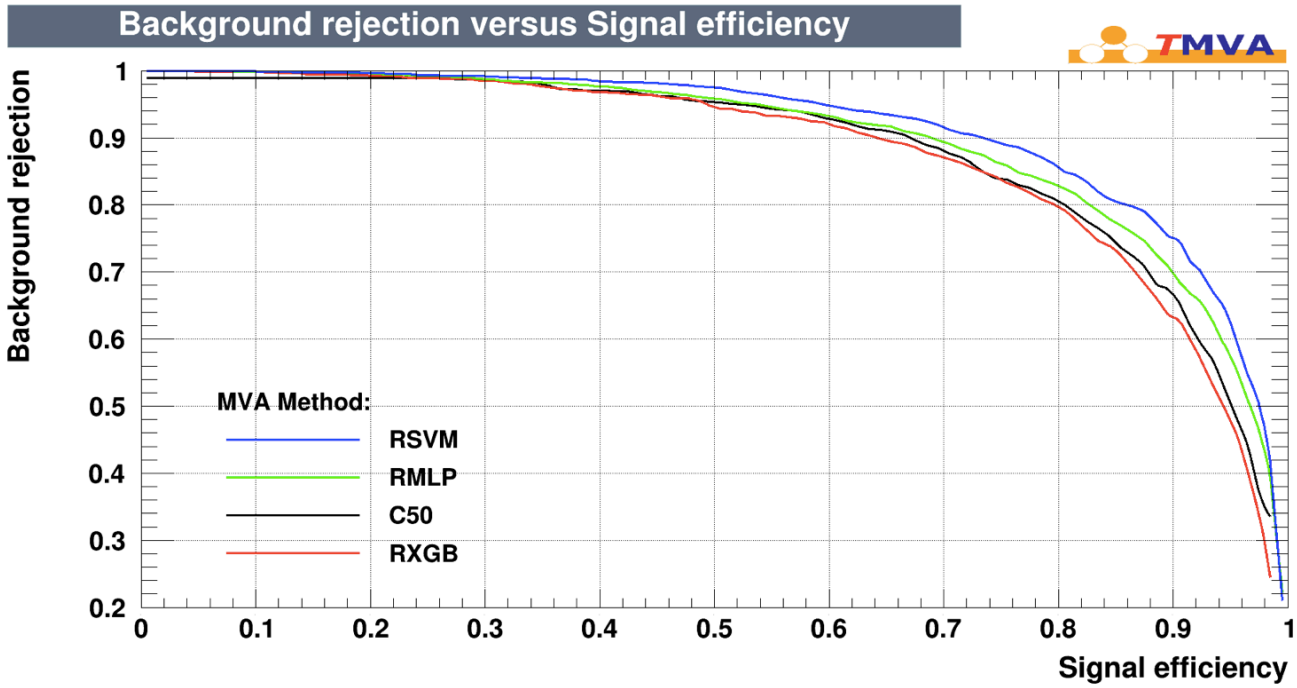
RMVA

- **RMVA is a set of plugins for TMVA that allows the use of R's classification and regression packages**





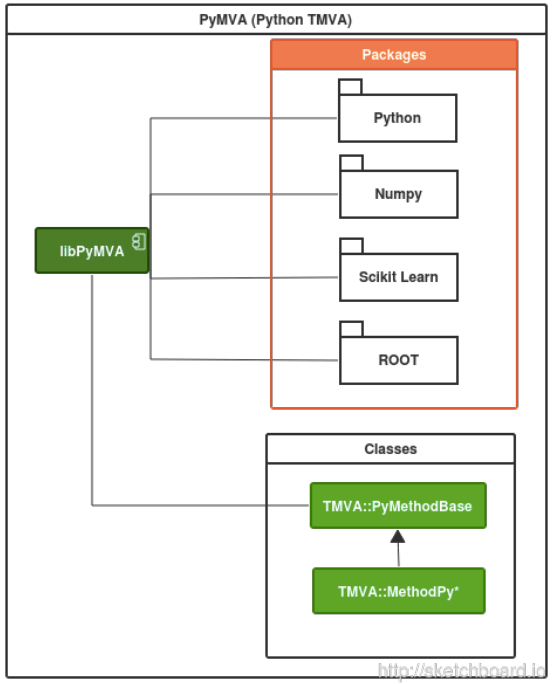
RMVA

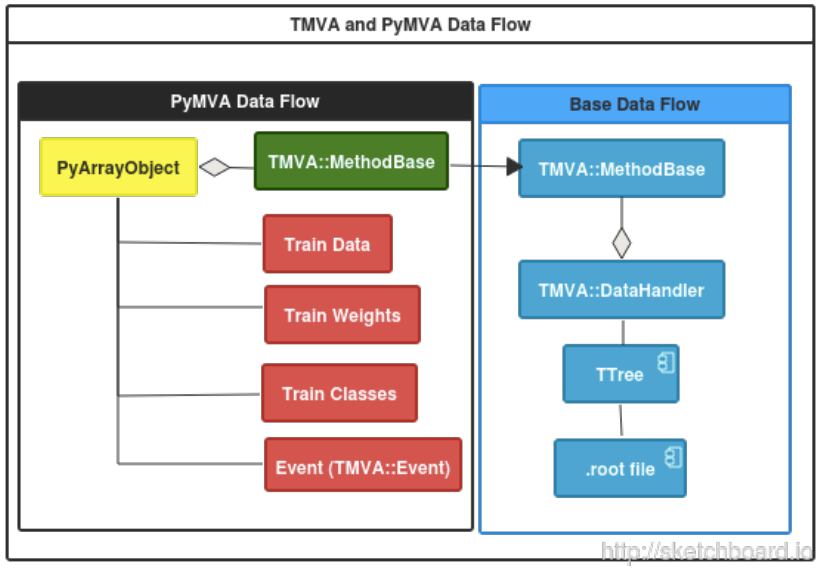



```
Evaluation results ranked by best signal efficiency and purity (area)
-------------------------------------------------------------------------------
MVA              Signal efficiency at bkg eff.(error):        | Sepa-    Signifi-
Method:          @B=0.01    @B=0.10    @B=0.30    ROC-integ. | ration:  cance:
-------------------------------------------------------------------------------
RSVM           : 0.328(08)  0.735(08)  0.924(04)    0.913    | 0.526    1.355
RMLP           : 0.286(08)  0.689(08)  0.899(05)    0.897    | 0.481    1.310
C50            : 0.000(00)  0.671(08)  0.878(05)    0.881    | 0.462    1.253
RXGB           : 0.233(07)  0.643(08)  0.867(06)    0.875    | 0.434    1.194
-------------------------------------------------------------------------------
```

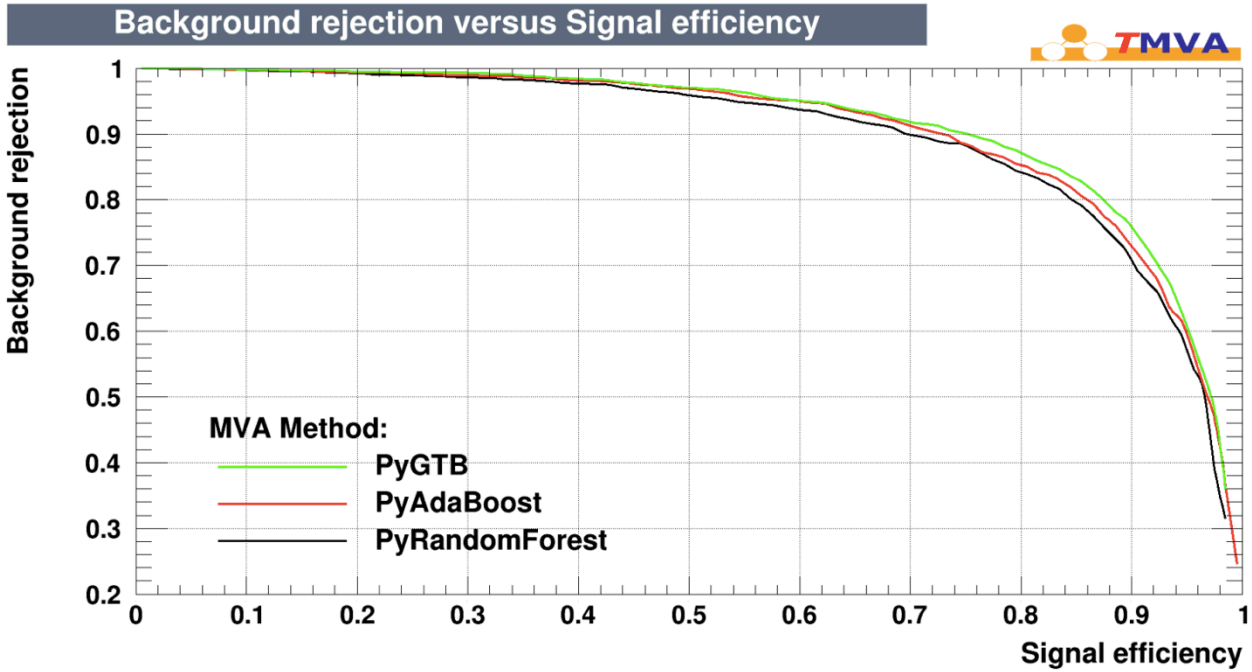
PyMVA

- **PyMVA is a set of plugins for TMVA based on python api that allows use of python based ML methods**





PyMVA




```
: Evaluation results ranked by best signal efficiency and purity (area)
: -------------------------------------------------------------------------------
: MVA              Signal efficiency at bkg eff.(error):       | Sepa-    Signifi-
: Method:          @B=0.01    @B=0.10    @B=0.30    ROC-integ. | ration:  cance:
: -------------------------------------------------------------------------------
: PyGTB          : 0.343(08)  0.751(07)  0.924(04)    0.914    | 0.539    1.514
: PyAdaBoost     : 0.331(08)  0.741(07)  0.918(05)    0.911    | 0.761    0.943
: PyRandomForest : 0.245(07)  0.702(08)  0.905(05)    0.898    | 0.497    1.375
: -------------------------------------------------------------------------------
:
```

Tutorials

Try the new features tutorial today and the interfaces tutorial on Thursday

More features coming soon.

More Information

ROOT
Data Analysis Framework

TMVA

Websites: http://root.cern.ch

http://oproject.org