



UF UNIVERSITY of
FLORIDA

Machine

Learning

Sergei

Gleyzer

PART

III

UERJ Mini-course in HEP Statistics

Nov. 26, 2015



Machine Learning Tutorials

Today's Outline



- TMVA New Features
 - Data Loader: Modularity
 - PyMVA
 - RootR
 - RMVA
 - Feature Selection
- Tutorial(s) and notebook demos

IML: Inter-experimental LHC Machine-Learning working group

- **Exchange of HEP-ML expertise** and experience among LHC experiments
- **ML Forum for LHC-related** development and discussions
- **ML software development and maintenance**
- **Current and future ML R&D in HEP**
- **Exchange between HEP and ML communities**
- **Education** (Tutorials)

- **Website:** <http://iml.cern.ch>
- **Latest meeting** Oct. 28
<https://indico.cern.ch/event/453344/>
- **Next meeting** Dec. 4
<https://indico.cern.ch/event/463561/>
 - Forum/Mailing-list
 - <https://groups.cern.ch/group/lhc-machinelearning-wg/default.aspx>
 - Please join if you are interested in ML topics



TMVA

New Features

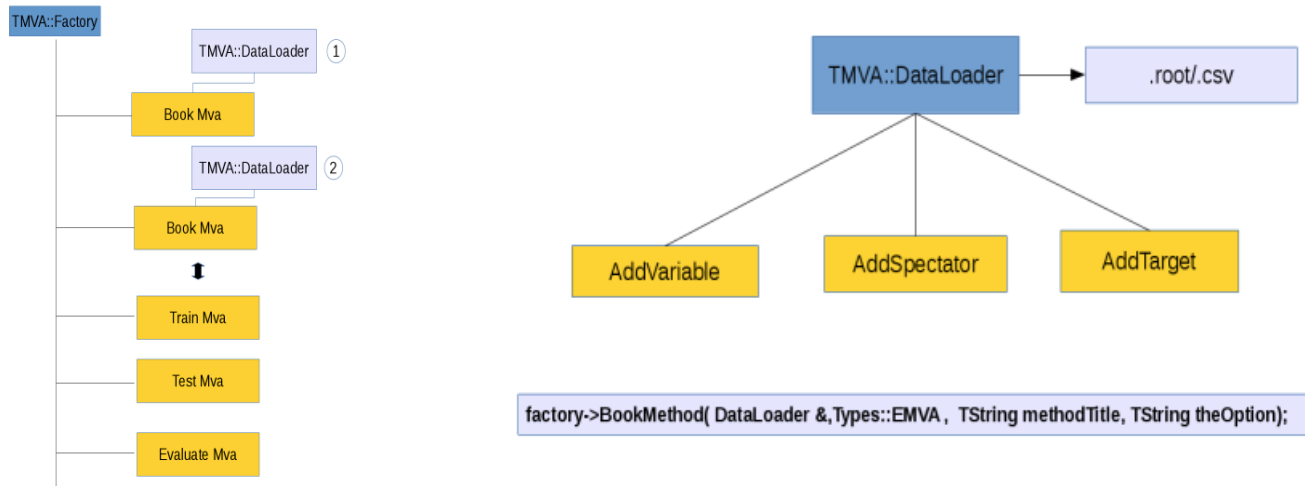


TMVA New Design

- Greater flexibility
 - **Modularization**
- **New** techniques
 - data storage and manipulation
 - parallelization
- **Feature Selection**
- Python and R integration
 - PyMVA/RMVA

TMVA::DataLoader class allows greater flexibility in dealing with data

- Connection of different **features** to different classifier methods: Useful for optimization



Proceed to Tutorial (TMVA)

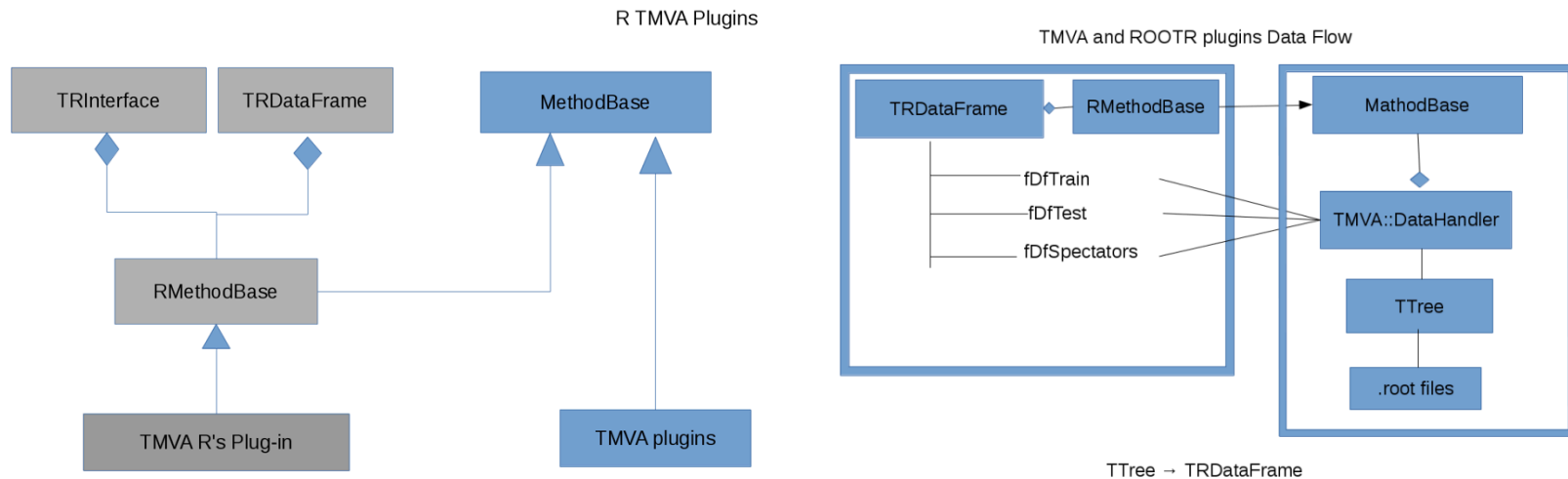
Part VI: Data Loader

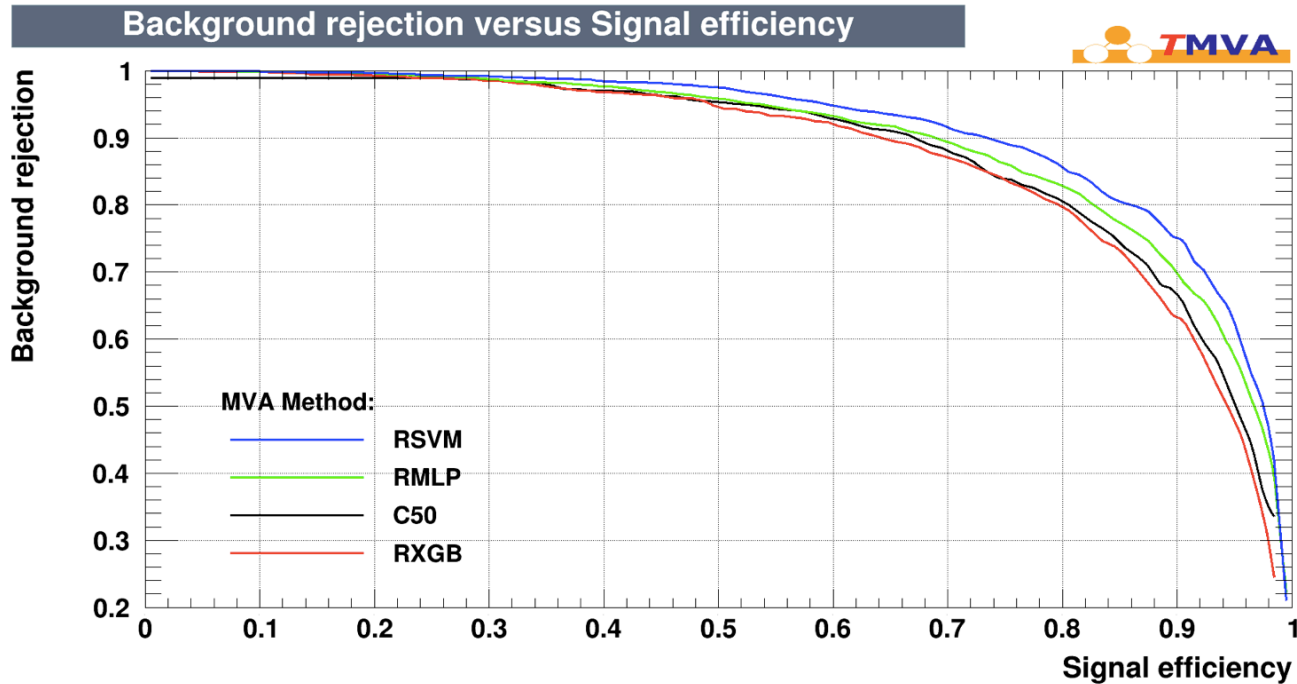
R: <https://www.r-project.org/>

Statistical software environment

- Many **ML** packages
 - Classification
 - Regression
- **Root-R** interface
 - Use **R** within root
- **RMVA**
 - Use **R ML** packages in root/TMVA

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





```

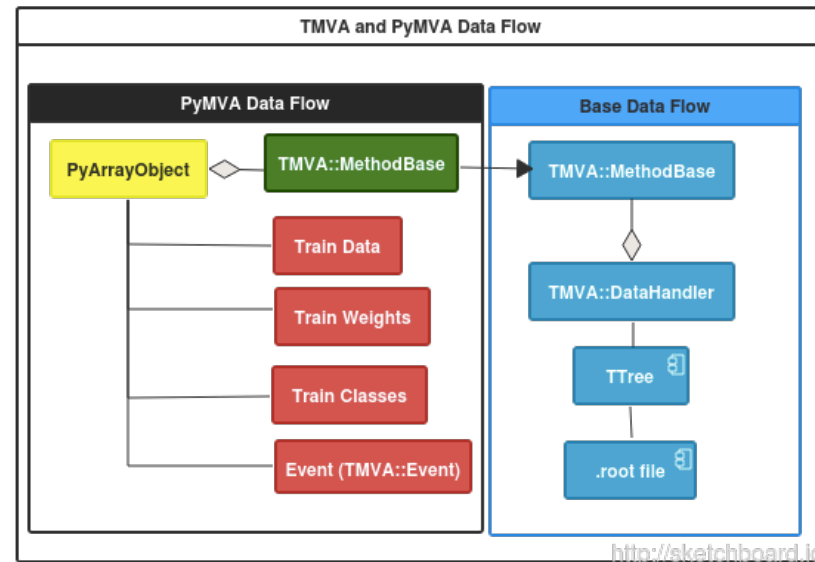
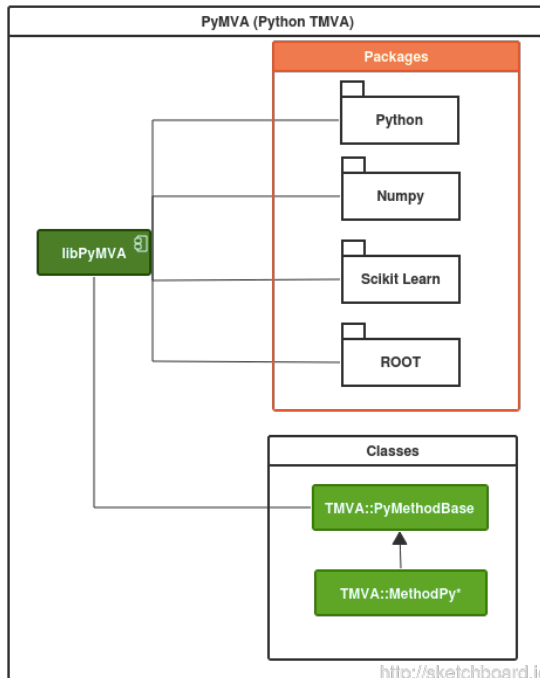
Evaluation results ranked by best signal efficiency and purity (area)
-----
MVA Method:      Signal efficiency at bkg eff.(error):  | Sepa-   Signifi-
                  @B=0.01   @B=0.10   @B=0.30   ROC-integ. | ration:  cance:
-----
RSVM              : 0.328(08) 0.735(08) 0.924(04) 0.913   | 0.526   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
-----

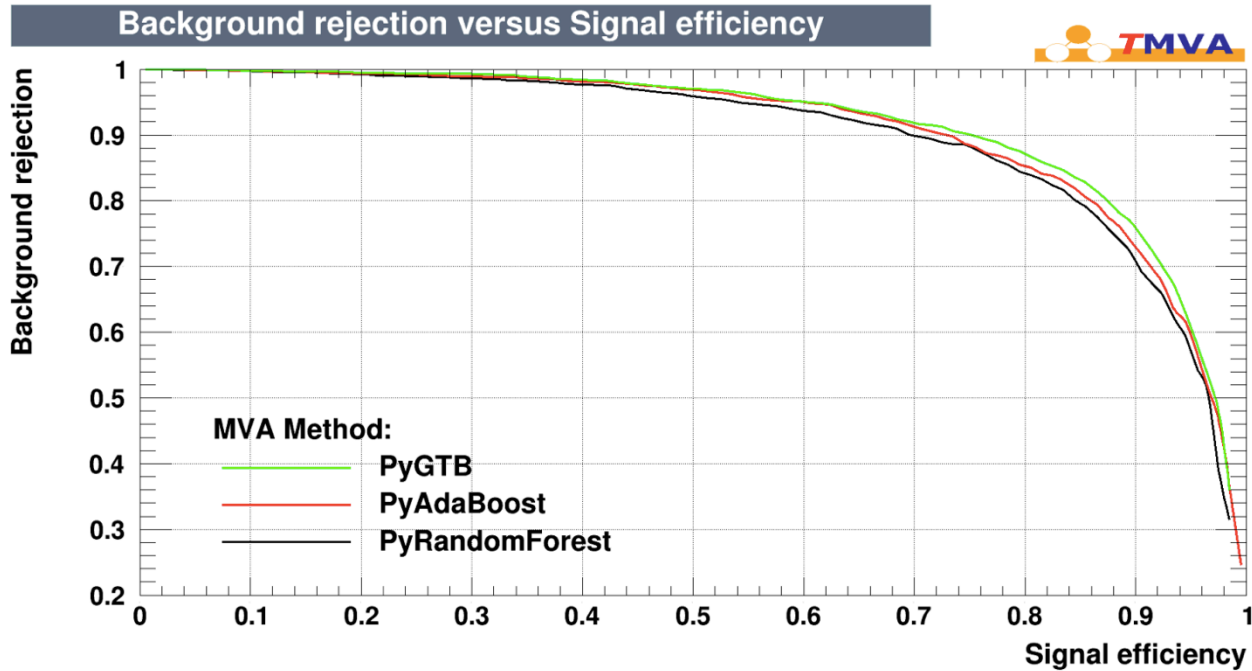
```

Proceed to Tutorial (TMVA)

Part VII: RootR + RMVA

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





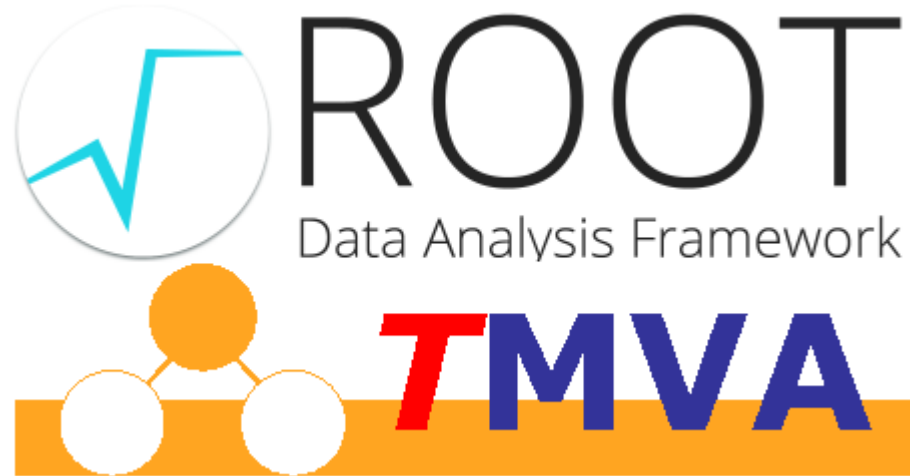
```

Evaluation results ranked by best signal efficiency and purity (area)
-----
MVA Method:      Signal efficiency at bkg eff.(error): | Sepa-   Signifi-
                  @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
-----
    
```

Proceed to Tutorial (TMVA)

Part VIII: PyMVA

More Information



Websites: <http://root.cern.ch>

<http://iml.cern.ch>

<http://oproject.org>