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
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

**Background rejection versus Signal efficiency**

<table>
<thead>
<tr>
<th>MVA Method</th>
<th>Signal efficiency at bkg eff.(error): @B=0.01</th>
<th>@B=0.10</th>
<th>@B=0.30</th>
<th>ROC-integ.</th>
<th>Separation</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>RSVM</td>
<td>0.323(08)</td>
<td>0.735(08)</td>
<td>0.924(04)</td>
<td>0.913</td>
<td>0.526</td>
<td>1.355</td>
</tr>
<tr>
<td>RMLP</td>
<td>0.286(08)</td>
<td>0.689(08)</td>
<td>0.899(05)</td>
<td>0.887</td>
<td>0.481</td>
<td>1.310</td>
</tr>
<tr>
<td>C50</td>
<td>0.080(09)</td>
<td>0.671(08)</td>
<td>0.878(05)</td>
<td>0.881</td>
<td>0.462</td>
<td>1.253</td>
</tr>
<tr>
<td>RXGB</td>
<td>0.233(07)</td>
<td>0.643(08)</td>
<td>0.867(06)</td>
<td>0.875</td>
<td>0.434</td>
<td>1.194</td>
</tr>
</tbody>
</table>
PyMVA

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

Background rejection versus Signal efficiency

Evaluation results ranked by best signal efficiency and purity (area)

<table>
<thead>
<tr>
<th>MVA Method</th>
<th>Signal efficiency at bkg eff.(error):</th>
<th>Separation</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>@B=0.01</td>
<td>@B=0.10</td>
<td>@B=0.30</td>
</tr>
<tr>
<td>PyGTB</td>
<td>0.343(08)</td>
<td>0.751(07)</td>
<td>0.924(04)</td>
</tr>
<tr>
<td>PyAdaBoost</td>
<td>0.331(08)</td>
<td>0.741(07)</td>
<td>0.918(05)</td>
</tr>
<tr>
<td>PyRandomForest</td>
<td>0.245(07)</td>
<td>0.762(08)</td>
<td>0.905(05)</td>
</tr>
</tbody>
</table>
Tutorials

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

More features coming soon.
More Information

Websites: http://root.cern.ch
http://oject.org