TMVA 4 – Toolkit for Multivariate Data Analysis in ROOT

TMVA provides a large set of sophisticated multivariate analysis techniques for both classification and regression tasks in HEP. All methods are embedded in a powerful yet user-friendly framework capable of handling the preprocessing of the input data as well as the evaluation and comparison of the MVA algorithms. TMVA is fully integrated in the popular ROOT data analysis framework.

Data input
- Supports TTree and ASCII files
- Supports arrays
- Any combination or function of input variables is possible

preprocessing
- Apply preselection
  - Individual cuts for different event classes are supported

Transformations
- Supports individual transform for each method
- Transformations can be chained
- NEW: Transformation of variable subsets
  - TMVA knows:
    - Normalisation
    - Decorrelation
    - Principal component analysis
    - Gaussianisation

Use event weights
- Supports event-by-event weights, weights for individual trees and weights for different classes

Classification
- NEW: Multiclass classification
  - to separate into classes
- …to one classifier output
- …cut on the classifier ...
- Condense all information
- Use all information...

Regression
- …to predict the value of one (or more) dependent variable(s)

evaluation & assessment
- ROC curve describes performance of a binary classifier by plotting the false positive vs. the true positive fraction
- Show average quadratic deviation of true and estimated value for both training and testing

summary & new developments
- Many MVA methods implemented
- One common platform/interface for all MVA methods
- Wide range of data preprocessing capabilities
- Common input and analysis framework (ROOT scripts)
- Train and test all methods on same data sample and evaluate consistently

TMVA classifier overview

| Criteria | TMVA classifier | CR | BDFT | MLP | QP | SVM | KNN | Random forest | IForest | BRT | M5PT
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>1. Linear combinations</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Speed</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Scalability</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
<tr>
<td>Robustness</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
<td>★</td>
</tr>
</tbody>
</table>