AlvsPK Challenge: FACT SHEET FORMAT

Title: Dr.
Name, address, email: Joerg Wichard
FMP - Molecular Modelling Group
Robert Roessle Strasse 10
13125 Berlin, Germany
joergwichard@web.de

Acronym of your best entry:

Reference: Agnostic Learning with Ensembles of Classifiers

Method:
Ensembles of Classifiers, mainly Classification and Regression Trees (CART). Ensemble building based on a modified bagging scheme.
We used a simple balancing strategy for data preprocessing.
No feature selection and no transduction method was used.
Model selection and parameter selection was combined in a modified cross-validation approach.

Results:

Table 1: Our methods best results

Dataset	Entry name	Entry ID	Test BER	Test AUC	Score	Track
ADA	Final	1039	0.181	0.906	0.1002	Agnos
GINA	boosted trees 2	1008	0.0731	0.979	0.6197	Agnos
HIVA	cv-trees	1010	0.295	0.7702	0.2008	Agnos
NOVA	boosted trees 2	1008	0.1416	0.9344	0.8462	Agnos
SYLVA	Final	1039	0.0115	0.9981	0.402	Agnos
Overall	Final	1039	0.1481	0.9137	0.4811	Agnos

Table 2: Winning entries of the AlvsPK challenge

Best results agnostic learning track						
Dataset	**Entrant name**	**Entry name**	**Entry ID**	**Test BER**	**Test AUC**	**Score**
ADA	Roman Lutz	LogitBoost with trees	13, 18	0.166	0.9168	0.002
GINA	Roman Lutz	LogitBoost/Doubleboost	892, 893	0.0339	0.9668	0.2308
HIVA	Vojtech Franc	RBF SVM	734, 933, 934	0.2827	0.7707	0.0763
NOVA	Mehreen Saeed	Submit E final	1038	0.0456	0.9552	0.0385
SYLVA	Roman Lutz	LogitBoost with trees	892	0.0062	0.9938	0.0302
Overall	Roman Lutz	LogitBoost with trees	892	0.1117	0.8892	0.1431
Best results prior knowledge track						
Dataset	**Entrant name**	**Entry name**	**Entry ID**	**Test BER**	**Test AUC**	**Score**
ADA	Marc Boulle	Data Grid	920, 921, 1047	0.1756	0.8464	0.0245
GINA	Vladimir Nikulin	vn2	1023	0.0226	0.9777	0.0385
HIVA	Chloe Azencott	SVM	992	0.2693	0.7643	0.008
NOVA	Jorge Sueiras	Boost mix	915	0.0659	0.9712	0.3974
SYLVA	Roman Lutz	Doubleboost	893	0.0043	0.9957	0.005
Overall	Vladimir Nikulin	vn3	1024	0.1095	0.8949	0.095967

- quantitative advantages: Ensembles of CART trees are robust, conceptually simple and fast. The critical issue of feature selection is done by the training algorithm, if "tree pruning" is used.

- qualitative advantages: Huge ensembles (almost like random forests) are possible with this method.

Code: Our own Matlab Ensemble Toolbox was use:
http://www.j-wichard.de/entool/