

Title: Model Selection in an Ensemble Framework

Name: Joerg Wichard
Institute of Molecular Pharmacology
Molecular Modelling Group
Robert Rössle Straße 10,
D-13125 Berlin-Buch, Germany.

Email: JoergWichard@web.de

Acronym of best entry: submission 13

Reference: Model Selection in an Ensemble Framework,
J. Wichard, In Proceedings IJCNN06, to appear.

Method: Our method uses a simple normalizing and balancing of the data sets. In one experiment we were using a SVM-based feature selection method. We build heterogeneous ensembles of classifiers, consisting of linear models (LDA, ridge), nearest neighbor models, trees, neural networks and SVMs. We use a cross-validation approach for model selection and hyperparameter selection. The performance prediction guess was calculated by averaging the BER (balanced error rate) on the validation sets in the cross-validation folds.

Results: In the challenge, we rank 38.2 in the group rank and our best entry is the 37th according to the test score computed by the organizers.

Our best entry						Best challenge entry				
Dataset	Test AUC	Test BER	BER guess	Guess error	Test score	Test AUC	Test BER	BER guess	Guess error	Test score
ADA	0.8614	0.1801	0.1636	0.0165	0.1965	0.8304	0.1696	0.155	0.1046	0.1843
GINA	0.9866	0.0523	0.05	0.0023	0.0543	0.9639	0.0361	0.0388	0.0027	0.0386
HIVA	0.7172	0.3057	0.338	0.0323	0.3377	0.7129	0.2871	0.27	0.0171	0.03029
NOVA	0.9459	0.0611	0.08	0.0189	0.08	0.9542	0.0458	0.0503	0.0045	0.0499
SYLVA	0.9956	0.0267	0.007	0.0197	0.0464	0.9937	0.0063	0.0058	0.0005	0.0067
Overall	0.9013	0.1252	0.1277	0.0179	0.143 (38.2)	0.891	0.109	0.104	0.0079	0.1165 (6.2)

Code: Our method is implemented in MATLAB. We developed a toolbox for regression tasks and a toolbox for classification problems, both are based on an ensemble approach. The code is available at:
<http://chopin.zet.agh.edu.pl/~wichtel/>.

Keywords: Cross-Validation, Heterogeneous Ensembles, Mixed Models