PERFORMANCE PREDICTION CHALLENGE: FACT SHEET FORMAT (1 to 2 pages)

Title: Artificial Contrasts with Ensembles and Regularized Least Squares Classifiers
Name, address, email: Kari Torkkola and Eugene Tuv
Acronym of your best entry: ACE+RLSC

References:

1) ACE: Feature Selection Using Ensemble Based Ranking Against Artificial Contrasts, Eugene Tuv, Alexander Borisov and Kari Torkkola, IJCNN06, to appear.
2) RLSC: Ensembles of Regularized Least Squares Classifiers for High-Dimensional Problems, Kari Torkkola and Eugene Tuv, in Feature Extraction, Foundations and Applications, Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti Zadeh (eds.), 2006

Method:

No preprocessing was used. For feature selection, we used feature ranking against artificial contrasts (ACE). ACE is an embedded feature selection method using Gradient Boosting Trees as the internal engine. For classification, we use regularized least squares classifiesr with Gaussian kernels. Hyper-parameters (kernel width, regularization coefficient) are adjusted after feature selection, using grid search and 10-fold CV with the same training data. Our test BER prediction is based also on the same 10-fold CV.

Results: In the challenge, we rank 8th as a group and our best entry is the 33rd, according to the average rank computed by the organizers. As advantages, compact feature sets could be mentioned.

Dataset	Our best entry					The challenge best entry				
	Test AUC	Test BER	BER guess	Guess error	Test score (rank)	Test AUC	Test BER	BER guess	Guess error	Test score (rank)
ADA	0.8178	0.1822	0.1673	0.0149	0.197 (24)	0.9149	0.1723	0.1650	0.0073	0.1793 (1)
GINA	0.9712	0.0288	0.0305	0.0017	0.0302 (1)	0.9712	0.0288	0.0305	0.0017	0.0302(1)
HIVA	0.6986	0.3014	0.2461	0.0553	0.3567 (47)	0.7671	0.2757	0.2692	0.0065	0.2797 (1)
NOVA	0.9256	0.0744	0.0469	0.0275	0.1018 (55)	0.9914	0.0445	0.0436	0.0009	0.0448 (1)
SYLVA	0.9911	0.0089	0.0044	0.0045	0.0134 (33)	0.9991	0.0061	0.0060	0.0001	0.0062 (1)
Overall	0.8809	0.1191	0.099	0.0208	0.1398 (32)	0.8910	0.1090	0.1040	0.0079	0.1165 (6.2)

Code: ACE experiments were run using Intel's IDEAL (not available). RLSC was written using MATLAB (one-liner), and all cross-validation experimentation was done using MATLAB (code not available).

- **Keywords:** embedded feature selection, gradient boosting trees, artificial contrast variables, Regularized Least Squares Classifier (RLSC) with Gaussian kernels, hyperparameter grid-search,10-fold cross validation.