Title: Cross-indexing
Name, email: Juha Reunanen, `Juha.Reunanen@iki.fi`
Acronym of best entry: CLOP-models-only-5

Reference:
J. Reunanen (2006), *Less Biased Measurement of Feature Selection Benefits.* In C. Saunders et al. (Eds.): SLSFS 2005, LNCS 3940, pp. 198–208. *To appear.*

Method:

Cross-indexing is a recent approach for assessing the outcome of a model selection process. Compared to traditional cross-validatory model selection and assessment, using cross-indexing may in some special cases either provide less biased results in a similar amount of time, or results of similar accuracy in significantly less time (depending on whether an outer loop of cross-validation is used). The method has been described in the context of feature selection in the reference mentioned above. In this challenge, it was used to select the model architecture and the corresponding parameters, and to estimate their performance when applied together. The models compared were introduced already in the sample code: *Prepro+naiveBayes*, *PCA+kernelRidge*, *GS+kernelRidge*, *Prepro+linearSVC*, *Prepro+nonlinearSVC*, and *Relief+neuralNet*. For each model type, a couple of parameters were subjected to optimization, but in other respects the models were treated as black boxes. The final ensemble consisted of nine members for each dataset.

In more detail, the selection took place as follows: First, the data available were split into nine folds. Then, during each of the nine iterations, eight of these folds were pooled and used during the search, while the remaining kth fold was utilized as a validation set, using which the optimal model and the corresponding parameters for the kth ensemble member were chosen. The union of the eight folds was further divided into only three folds (to save some time) in order to facilitate standard cross-validation to guide a simple stochastic search for the optimal parameters. The search was interleaved to give equal possibilities for all the model architectures being considered: the execution scheduler basically tried to round-robin the time spent (instead of the number of evaluations), with the exception that more time was allocated to the optimization of those models that were able to demonstrate good performance estimates.

The performance estimate obtained for the optimal parameter set using the remaining fold was potentially overfitted when a large number of comparisons had been performed. Thus, this score was not used as such to assess the performance of the corresponding ensemble member - instead, the cross-indexing approach was adopted to recall the estimated performance on the other folds after a similar number of iterations. This score was not used to select the model, and thus it had not been overfitted due to a multiple-selection process. The final performance guess was obtained as the median of these nine guesses. This might have introduced a pessimistic bias, as the ensemble can be expected to perform better than its individual members, but based on the results, it looks like this did not really happen. However, the variance of these nine guesses could have been used to estimate *the accuracy of the BER guess*, had that been the goal of the challenge.

Results:

While no competitive BER was obtained for any of the datasets, the guess error remains at an acceptable level, and the AUC is good. Moreover, with respect to the test score, the method beats the reference entries that were using the same CLOP models, although such a comparison is hardly a fair one, as the reference models were probably trained without the validation labels. Still, it can be said that the selection of the final model, and the estimation of its performance using cross-indexing, were performed successfully.

Table 1: Comparison between *CLOP-models-only-5* and the winner of the challenge, *LB tree mix cut adapted* by Roman Lutz

Dataset	Our best entry					The challenge best entry				
	Test AUC	Test BER	BER guess	Guess error	Test score	Test AUC	Test BER	BER guess	Guess error	Test score
ADA	0.8825	0.2037	0.1884	0.0153	0.2190	0.8304	0.1696	0.1550	0.0146	0.1843
GINA	0.9631	0.0980	0.0840	0.0141	0.1121	0.9639	0.0361	0.0388	0.0027	0.0386
HIVA	0.7392	0.3088	0.3172	0.0084	0.3148	0.7129	0.2871	0.2700	0.0171	0.3029
NOVA	0.9874	0.0892	0.0917	0.0025	0.0907	0.9542	0.0458	0.0503	0.0045	0.0499
SYLVA	0.9971	0.0341	0.0320	0.0021	0.0357	0.9937	0.0063	0.0058	0.0005	0.0067
overall	0.9138	0.1468	0.1427	0.0085	0.1545 (45.8)	0.8910	0.1090	0.1040	0.0079	0.1165 (6.2)