ALvsPK Challenge: FACT SHEET

Title: Cross-indexing
Name, email: Juha Reunanen, Juha.Reunanen@iki.fi
Acronym of best entry: cross-indexing-7a (AL), cross-indexing-prior-1 (PK)

Reference:

J. Reunanen (2006): Less Biased Measurement of Feature Selection Benefits. In C. Saunders et al. (Eds.), *Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop* (SLSFS 2005; LNCS 3940), Revised Selected Papers, pp. 198–208.

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*, *Relief+neuralNet*, *RF*, and *Boosting* (with *neuralNet*, *SVC* and *kernelRidge*). 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 three, five or nine members, depending on the dataset.

In more detail, the selection took place as follows: First, the data available were split into K (five or nine[1]) folds, depending on the dataset (no magic here: just varied it depending on the time, memory etc. available). Then, during each of the K iterations, $K-1$ 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 $K-1$ 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 for the present dataset.

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. Therefore, 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

[1] In some of my submissions, including cross-indexing-7a, the HIVA model only contains three ensemble members. This is because two of the five (that were searched for) were manually removed, due to their apparently bad performance.

estimated performance *on the other folds* after a similar number of iterations. These scores had not been used to select this model, thus they had not been overfitted due to a multiple-selection process. The final performance guess (which was not required in this challenge, but is always useful for development purposes) was obtained as the median of the K guesses. This may have introduced a pessimistic bias, as the ensemble can be expected to perform better than its individual members.

Results:

Table 1: Our methods best results

Dataset	Entry name	Entry ID	Test BER	Test AUC	Score	Track
ADA	cross-indexing-prior-2	905	0.1807	0.907	0.0961	Agnos
GINA	cross-indexing-prior-3	996	0.0236	0.997	0.1068	Prior
HIVA	cross-indexing-7a	882	0.2863	0.7662	0.1004	Agnos
NOVA	cross-indexing-prior-1	743	0.0472	0.9903	0.0769	Agnos
SYLVA	cross-indexing-prior-1	743	0.0066	0.9989	0.0603	Prior
Overall	cross-indexing-prior-1a	883	0.11	0.9312	0.1294	Prior

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

CLOP models used:

Dataset	Track	Ensemble members
ADA	AL	2*{sns,std,norm,gentleboost(neural),bias}; 2*{std,norm,gentleboost(kridge),bias}; 1*{rf,bias}
GINA	AL	6*{std,gs,svc(degree=1)}; 3*{std,svc(degree=2)}
	PK	4*{std,svc(degree=2); 1*{rf}
HIVA	AL	3*{norm,svc(degree=1),bias}
NOVA	AL	5*{norm,gentleboost(kridge),bias}
SYLVA	AL	4*{std,norm,gentleboost(neural),bias}; 4*{std,neural}; 1*{rf,bias}
	PK	3*{sns,std,norm,gentleboost(neural),bias}; 2*{rf,b}

(sns = shift'n'scale, std = standardize, norm = normalize)

The ensemble members were chosen during the model selection loop according to their estimated performance (using the cross-indexing criterion).

Keywords:

- Preprocessing: centering, scaling, standardization.
- Feature selection: Gram-Schmidt (only GINA on the AL track).
- Classifier: boosting, neural networks, ridge regression, kernel method, RF.
- Hyper-parameter selection: stochastic search, cross-validation, cross-indexing.
- Other: ensemble method.