Title: Dr. Erinija Pranckeviciene
Address: Institute for Biodiagnostics, NRC Canada, Ellice ave 435, Winnipeg, MB.
Email: erinija.pranckevie@nrc-cnrc.gc.ca
Method: Linear Programming SVM (Liknon) feature selection combined with state of the art classification rules.

Reference: E.Pranckeviciene and R.Somorjai, Liknon feature selection for Microarrays, in F.Masulli, S.Mitra, and G.Pasi(Eds.):WILF 2007, LNAI 4578, 580-587, 2007.

Code/engine:

1) For the Mtlab implementation of the LP SVM we refer to C. Bhattacharrya, LR Grate , A Rizki and et. al,"Simultaneous relevant feature identification and classification in high-dimensional spaces:application to molecular profiling data, Signal Processing, vol 83(4), 729-743, 2003.
2)For the Matlab implementation of the state of the art classifiers we refer to R. Duin, P. Juscak, P. Paclick, E.Penkalska , D. deRidder, D.Tax, PRTools4 A Matlab toolbox for pattern recognition, February, 2004.

Implementation strategy: The dataset first is divided into 10 parts (10 fold crossvalidation). The training set of the single fold is further subdivided into balanced training and unbalanced (the rest samples) monitoring sets. A number of the subdivisions is arbitrary, we did 31. In every subdivision a number of LP SVM models/discriminants is trained on the balanced training set and the BER -balanced error rate- of each is estimated on the monitoring set. The number of the models depends on the set of the chosen values of regularization parameter C. The data, number of features and available computational time determine the range of C values. The model with smallest monitoring BER is selected in every subdivision. As a result, in the single fold, we have an ensemble of the linear discriminants and a feature profile possibly to be investigated with the other classification rules.

Preprocessing: none.

Feature selection: Feature selection relies on the property of LP SVM to produce sparse solutions. The identities of the features, corresponding to non zero weights of the discriminants, are included in the profile. Every fold reveals slightly different feature profiles. However relevant feature identities consistently appear in many subdivisions. First it was noticed in the experiments with synthetic data and then in the experiments with microarrays.

Classification:

1) ensemble of linear discriminants;
2) the rules tested on the derived feature profile included linear and nonlinear, all available in the PRTools: fisher classifier, linear logistic classifier, subspace classifier, nearest neighbors, decision tree, neares mean classifier, quadratic classifier.
3) NO TRANSDUCTION.

Results:

Table 1: Our methods best results

Dataset	Entry name	Entry ID	Test BER	Test AUC	Score	Track

ADA	liknon feature selection + state of art (1)	1012	0.1818	0.8702	0.135	Agnos
GINA	liknon feature selection + state of the art (3)	1014	0.0533	0.974	0.3889	Agnos
HIVA	liknon feature selection+ state of art classifiers	814	0.2939	0.7589	0.1647	Agnos
NOVA	liknon feature selection + state of art 2	713	0.0725	0.9814	0.5064	Agnos
SYLVA	liknon feature selection + state of the art (2)	1013	0.019	0.9949	0.7085	Agnos
Overall	liknon feature selection + state of art (1)	1012	0.127	0.9133	0.4358	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: Simplicity and interpretability of the results in terms of feature identities, important for discrimination. The stability of the discovered feature identities in different folds suggests that the feature selection via LP SVM is robust to the sample bias. In case of high dimensional data, the discovered features provide a reduced representation of the data for testing other classifiers. A success of the suggested approach was apparent for GINA dataset. DISADVANTAGE- high computational burden.

Qualitative advantages: The sequence of values of the regularization parameter determines the increasing number of features given by the increasing number of non-zero weights of the linear discriminant. This can be considered as a feature selection structure. The sequence of the LP SVM solutions- linear discriminants of increasing complexity- forms a nested structure, where the principles of the structural risk minimization may apply.

Keywords:

- Preprocessing or feature construction: none.
- Feature selection approach: embedded feature selection.
- Feature selection engine: SVM.
- Feature selection search: feature ranking, ordered FS (ordered feature selection)
  Feature selection criterion: monitoring error
- Classifier: nearest neighbors, tree classifier, L1 norm regularization, ensemble method, bagging.
- Hyper-parameter selection: grid-search.
- Other: sample bias.