Title: Large margin linear classifiers with bias adjustment for skewed two-class distributions.

Name, address, email: Edward F. Harrington, Defence Science and Technology Organisation (DSTO) Australia, edward.harrington@dsto.defence.gov.au

Acronym of best entry: ba4

Reference: no specific paper as yet.

Method:

The data was preprocessed with centering and scaling (standardization). In the case of the SVC sub-sampling was also done.

No feature selection was done prior to classification.

Three linear classifiers were used: SVC, Online Bayes Point Machine (OBPM) and Approximate Large Margin Algorithm (ALMA).

Model and parameter selection was done using a simple cross-validation. To select the model of OBPM and ALMA we considered the results for each random permutation of the training set. To tune the bias parameter of OBPM and ALMA we used a line search to minimize the affect of any skew between the two classes in the training set. The criteria to select the "best" model for each classifier was simply the minimum BER from the validation set. The classifier with the lowest validation BER amongst the three was selected as the "best" classifier.

Results:

Data set	***Our Best entry***					***The Challenge best entry***				
	Test AUC	Test BER	Guess BER	Guess Error	Test Score	Test AUC	Test BER	BER guess	Guess Error	Test score
ADA	0.8900	0.1964	0.1761	0.0203	0.2168	0.8304	0.1550	0.1550	0.0146	0.1843
GINA	0.9758	0.0750	0.0600	0.0086	0.0836	0.9639	0.0388	0.0388	0.0027	0.0386
HIVA	0.7043	0.3509	0.3500	0.0009	0.35	0.7129	0.2871	0.2700	0.0171	0.3029
NOVA	0.9369	0.0631	0.0500	0.0131	0.0762	0.9542	0.0458	0.0503	0.0045	0.0499
SYLVA	0.9886	0.0114	0.0042	0.0072	0.0187	0.9937	0.0063	0.0058	0.0005	0.0067
Overall	0.8991	0.1394	0.1294	0.01	0.1492 (43.8)	0.8910	0.1090	0.1040	0.0079	0.1165 (6.2)

Keywords: standardization, sub-sampling, cross-validation, permutation, bias adjustment, linear classifier, OBPM, ALMA, SVC.