Title: Weighted LS-SVM + Leave-One-Out Cross-Validation + Repeated Hold-Out
Name, address, email: Gavin Cawley, University of East Anglia, Norwich, United Kingdom, gcc@cmp.uea.ac.uk
Acronym of your best entry: Final #2

Reference:
G. C. Cawley, "Leave-one-out cross-validation based model selection criteria for weighted LS-SVMs", Proceedings of the International Joint Conference on Neural Networks (IJCNN-2006), to appear. http://theoval.cmp.uea.ac.uk/~gcc/publications/pdf/ijcnn2006a.pdf

Method:
Preprocessing: All continuous features are standardised, also for ADA a log transform used on features 1 and 3 and thresholding of features 4 and 5. For the SYLVA dataset, features 11 and 12 are never positive for positive examples; this observation was used to reduce the number of training patterns so that the LS-SVM could be applied directly.
Feature selection: No feature selection was used, other than deleting constant features in the NOVA benchmark. Regularisation was the only mechanism used to avoid over-fitting.
Classification: Least-squares support vector machines (LS-SVMs), with linear, quadratic, cubic, Boolean and radial basis function (RBF) kernel functions. LS-SVMs with and without a bias term were evaluated. The LS-SVMs could optionally be weighted to equalise the importance of positive and negative patterns (as the balanced error rate is used as the primary performance indicator). Validation set targets were used in training and model selection.
Model selection: The optimisation of regularisation and kernel parameters was achieved by minimising leave-one-out cross-validation based estimates of generalisation performance via the Nelder-Mead simplex method. A variety of model selection criterion were investigated including sum-of-squares error (i.e. Allen's PRESS statistic), hinge loss, squared hinge loss, a smoothed approximation of the error rate and the smoothed Wilcoxon-Mann-Whitney statistic (i.e. the area under the ROC curve). The selection criterion could optionally be weighted to compensate for the disparity in class frequencies. The final choice of model, including the choice of kernel, use of a bias term, use of weighting in training and/or model selection and the choice of model selection criterion were all made by minimising the leave-one-out BER. A total of 180 experiments were conducted; this was somewhat computationally expensive!
Performance prediction: The test BER was estimated via 100 random 90%/10% training/test partitions of the avaliable data, with model selection performed independently in each trial in order to avoid selection bias.

Results:
This entry is a joint winner of the competition, having the lowest average test score and finishing second in terms of average rank. This entry also has the lowest overall guess error of any submission and the second highest overall test AUC. It is interesting that the models performed so well on the HIVA and NOVA benchmarks, given that no feature selection was used. It is reassuring that regularisation is effective in avoiding over-fitting, given a good value for the regularisation parameter. Also, the leave-one-out procedure is often (rightly) criticised as having a high variance, so it is interesting that it performed so well. This is probably because there were relatively few degrees of freedom to be optimised in model selection. If leave-one-out cross-validation were used in feature selection, there would probably be a much higher degree

of over-fitting. Model details are as follows:

- `ADA`: Unweighted LS-SVM with bias, Radial Basis Function kernel, Wilcoxon-Mann-Whitney model selection criterion.
- `GINA`: Unweighted LS-SVM without bias, inhomogeneous cubic kernel, unweighted smoothed error rate model selection criterion.
- `HIVA`: Unweighted LS-SVM with bias, inhomogeneous quadratic kernel, Wilcoxon-Mann-Whitney model selection criterion.
- `NOVA`: Weighted LS-SVM with bias, linear kernel, weighted mean-squared error model selection.
- `SYLVA`: Weighted LS-SVM with bias, inhomogeneous cubic kernel, Wilcoxon-Mann-Whitney model selection.

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	0.8965	0.1845	0.1742	0.0103	0.1947 (13)	0.9149	0.1723	0.1650	0.0073	0.1793 (1)
GINA	0.9900	0.0461	0.0470	0.0009	0.0466 (13)	0.9712	0.0288	0.0305	0.0017	0.0302 (1)
HIVA	0.7464	0.2804	0.2776	0.0028	0.2814 (2)	0.7671	0.2757	0.2692	0.0065	0.2797 (1)
NOVA	0.9914	0.0445	0.0470	0.0025	0.0464 (3)	0.9914	0.0445	0.0436	0.0009	0.0448 (1)
SYLVA	0.9990	0.0067	0.0065	0.0002	0.0067 (8)	0.9991	0.0061	0.0060	0.0001	0.0062 (1)
overall	0.9246	0.1124	0.1105	0.0034	0.1152 (7.8)	0.8910	0.1090	0.1040	0.0079	0.1165 (6.2)

Code:
The LS-SVM, leave-one-out model selection, Nelder-Mead simplex optimisation methods were all implemented by the author in MATLAB. Extensive use was made of automatically generated scripts to run the individual experiments. A demonstration of the approach used will shortly be made available from `http://theoval.cmp.uea.ac.uk/~gcc/matlab/default.html#loo` .

Keywords: standardisation, no feature selection, kernel method, least-squares support vector machine, L2 norm regularisation, leave-one-out model selection, pattern weighting, Nelder-Mead simplex, repeated hold-out validation.