

## AlvsPK Challenge: FACT SHEET

**Title: Modified multi-class SVM formulation; Efficient LOO computation**

**Name, address, email:**

Vojtech Franc  
Fraunhofer Institut FIRST IDA  
Kekulestr. 7, 12489 Berlin  
fravoj@first.fraunhofer.de

**Acronym of your best entry:**

SVM-RBF

**Reference:**

Unpublished.

**Method:**

**Prior knowledge – GINA**

Preprocessing: Each input image was normalized such that variance of pixels was one. No feature selection was applied.

We modified multi-class SVM by Crammer et al. 2001. A decision {odd,even} is taken based on the output of the multi-class classifier, i.e. we take decision “odd” whenever the multi-class classifier returns 1,3,...9 and the decision “even” otherwise. The standard multi-class SVM minimizes the number misclassification. In our modified formulation, we minimize the number of wrong decisions made on account of the output of the multi-class classifier. For example, if the multi-class classifier returns 1 and the true label is 2 we penalize the decision by one. However, if the classifier returns 4 (or any even number) and the true label is 2 we pay no penalty since the final decision is correct. This modification leads to a QP task similar to the original multi-class SVM but the number of linear constraints is reduced. We used a cutting plane algorithm similar to Tsochantaridis et al. 2005 to solve the underlying QP task.

We implemented a kernel version of this modified SVM and used the RBF (Gaussian) kernel in all experiments. The optimal hyper-parameters (regularization constant and the kernel width) were selected from a chosen finite set by minimizing the BER estimate on the validation set.

**Agnostic learning – HIVA, NOVA, SILVA**

Features were normalized such that the covariance matrix was the identity matrix. No feature selection was applied.

We used standard binary SVM with RBF kernel in all the experiments.

The hyper-parameters were tuned using the Leave-One-Out (LOO) estimate of the BER error. To make computation of the LOO feasible we designed a new method which

allows for early stopping of the QP optimizer. The new stopping condition guarantees that the decision on the test example equals to the decision of the classifier with optimal (minimizing the QP exactly) parameters, i.e. this method gives the exact LOO estimate and it is computationally feasible. In contrast to the standard stopping conditions used in QP-SVM optimization (e.g those based on KKT conditions or duality gap) our stopping condition is directly related to the desired output of the optimization, i.e. we stop the optimization when the decision (label) on the testing example can be reliably determined.

## Results:

Table 1: Our methods best results

Dataset	Entry name	Entry ID	Test BER	Test AUC	Score	Track
ADA	RBF SVM	716	0.1833	0.8987	0.1984	Agnos
GINA	RBF SVM	933	0.023	0.9962	0.0598	Prior
HIVA	RBF SVM	734	0.2827	0.7707	0.0763	Agnos
NOVA	RBF SVM	734	0.0877	0.9834	0.6795	Agnos
SYLVA	RBF SVM	734	0.0205	0.997	0.7437	Agnos
Overall	RBF SVM	933	0.1211	0.9289	0.408	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

Add GINA: In contrast to treating the problem as standard binary classification, the proposed method exploits the information about the clustering of the data within the decision classes by modeling each digit with a single normal vector. The proposed formulation is a natural extension of the multi-class SVM when an addition information about clustering within the classes is available.

Add Agnostic Learning: Our method allows for exact and efficient computation of the LOO estimate for problems with thousands of examples.

**Code:** We implemented all the methods in Matlab. The critical time consuming parts (like QP solvers) are written in C. In principal, we can make the code available but currently it is not in the state ready for publication.

**Keywords:** Put at *least one keyword in each category*. Try some of the following keywords and add your own:

- Preprocessing or feature construction: feature normalization.
- Feature selection approach: none.
- Classifier: SVM, kernel-method, L2 norm regularization.
- Hyper-parameter selection: leave-one-out.
- Other: QP optimization.