# Large scale multi-class classification based on linear optimisation

**Sándor Szedmák**
joint work with
**John Shawe-Taylor**, **Craig J. Saunders** and **David R. Hardoon**

ISIS Group
Electronics and Computer Science
University of Southampton
*ss03v@ecs.soton.ac.uk*
23/03/2004

## Abstract

One of the most hard tasks in image classification to find a method being applicable on large scale multi-class problems where the sample size and number of the features are huge.

Linear discriminant analysis as a classical method for multi-class classification, which was introduced by Fisher (1936) [4], plays an important role in the machine learning society recently. The kernelized version of this method are discussed in several papers, however they generally deal with the two class version of this approach. Bartlett recognised, in 1938 [2], there is strong relationship between the Fisher Discriminant and the Canonical Correlation Analysis and this statement is valid for the multi-class case as well. Based on this work Barker et al. (2003) [1] and Rosipal et al. (2003) [10] discuss the details about this relationship and show the appropriate kernel approach to this problem.

Using Canonical Correlation for multi-class classification in large scale problem suffers from the numerical difficulty to solve the generalised eigenvalue problem to provide the optimum. We present an analogue classification procedure based on linear optimisation which is able to extend the scale range of the solvable problems and to give sparse solution. Our method exploits the relationship between the L1 norm SVM and the boosting approach which were presented by Bennett et al. (2000) [3], Mangasarian (1999) [5] and Meir et al. (2003) [6]. Additionally, the formulation based on the soft margin SVM can solve the problem when the number of the features are less than the number of the observations in a given sample.

# References

[1] M. Barker and W. Rayens, "Partial least squares for discrimination", *Journal of Chemometrics*, 17:166-173, Wiley InterScience, 2003.

[2] M.S. Bartlett, "Further aspects of the theory of multiple regression", *Proceedings of the Cambridge Philosophical Society* 34:33-40, 1938.

[3] K.P. Bennett and E. Bredensteiner, "Duality and Geometry in SVM Classifiers", *Proceedings of the Seventeenth International Conference on Machine Learning, Pat Langley Editor*, Morgan Kaufmann, San Francisco, 57-64, 2000.

[4] R.A. Fisher, "The use of multiple measurement in taxonomic problems", *Annals of Eugenics*, 7:179-188, 1936.

[5] O.L. Mangasarian, "Arbitrary-norm separating plane", *Operations Research Letters* 24,1:15-23,1999.

[6] R. Meir and G. Rätsch, "An introduction to Boosting and Leveraging", *In Advanced Lectures on Machine Learning*, LNCS, 119-184, Springer, 2003

[7] G. Rätsch, S. Mika, B. Schölkopf, K.R. Müller, "Constructing Boosting Algorithms from SVMs: an Application to One Class Classification", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2002.

[8] R. Rifkin and A. Klautau, "In Defense of One-Vs-All Classification", *Journal of Machine Learning Research*, 5:101-141, 2004.

[9] R. Rosipal and L.J. Trejo, "Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space", *Journal of Machine Learning Research*, 2:97-123, 2001.

[10] R. Rosipal, L. J. Trejo and B. Matthews, "Kernel PLS-SVC for Linear and Nonlinear Classification", *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, 2003.