Classification of Pairwise Proximity Data with Support Vectors

Sepp Hochreiter and Klaus Obermayer
Department of Electrical Engineering and Computer Science
Technische Universität Berlin, 10587 Berlin, Germany

We investigate the problem of learning a classification task on data which are represented in terms of their pairwise proximities. This representation does not refer to an explicit representation of the data items by feature vectors, rather the set of objects is represented by a matrix, which assigns a real number (the similarity measure) to every object pair. This representation is more general than the standard approach of using feature vectors, from which pairwise proximities can always be calculated.

We assume that there exists a training set, for which a class label is assigned to every object and for which the proximity matrix has been measured. We now interpret the entries of the matrix as being produced by an unknown kernel operating on (unknown) object feature vectors. Proximity matrices are by definition symmetric. Positive definiteness, however, cannot be assured, and the construction of a classifier using the standard support vector machine (SVM) approach fails.

In order to properly extend the standard SVM approach we first extend Mercer's theorem to indefinite kernels. We show, that - under certain mild assumptions - an indefinite kernel corresponds to a dot product between feature vectors in a Minkowski space. The optimal hyperplane is then constructed in Minkowski space using the principle of structural risk minimization. However, instead of minimizing the squared length of its normal vector we proposed to minimize the squared length of the projection of the data onto the normal vector of the separating hyperplane. This amounts to replacing the vector of ones and the kernel matrix in the dual objective function by the kernel matrix and by its square, respectively. If the kernel matrix is positive definite, both objective functions have the same minima. Optimizing the length of the projection of the data, however, leads to well defined solutions also if the kernel matrix has negative eigenvalues. In principle, there are different ways in which the dual objective function can be modified to allow for well defined solutions for indefinite kernels. We motivate our particular choice by (i) the close relationship between this kind of SVM and the recently described "Coulomb Classifiers" [2] and (ii) by the fact, that our approach allows for the construction of classifiers, for which some objects can be excluded from acting as support vectors. Those objects are used for optimization but may be omitted for classification, e.g. for the determination of similarity values with new incoming objects. This feature may be beneficial for problems for which some objects (e.g. web-pages in content-based classification) cease to exist.

Benchmarks were conducted using several real world datasets (connectivity patterns in cat cortex, sequence similarities between proteins, text relevance). Our new method was compared to (i) K-nearest-neighbors and (ii) a method which applies SVMs to the column vectors of a proximity matrix [1]. The results show, that the new method provides better classification results. It is also more efficient, because only a small subset of objects (the support vectors) of the training set must be kept in order to measure the similarities with new incoming objects for their classification.

[1] T. Graepel, R. Herbrich, P. Bollmann-Sdorra, and K. Obermayer. Classification on pairwise proximity data. In M. S. Kearns, S. A. Solla, and D. A. Cohn, ed., *Advances in Neural Information Proc. Systems 11*, pp. 438–444. MIT Press, 1999.

[2] S. Hochreiter and M. C. Mozer. Coulomb classifiers: Reinterpreting SVMs as electrostatic systems. Technical Report CU-CS-921-01, Department of Computer Science, University of Colorado, Boulder, 2001. See http://ni.cs.tu-berlin.de/~hochreit.