Large Margin vs. Large Volume in Transductive Learning^{*}

Ran El-Yaniv¹, Dmitry Pechyony¹, and Vladimir Vapnik²

¹ Computer Science Department, Technion - Israel Institute of Technology,
Haifa, 32000, Israel

`{rani,pechyony}@cs.technion.ac.il`

² NEC Laboratories America, Princeton, NJ 08540, USA
`vapnik@att.net`

We focus on distribution-free *transductive learning*. In this setting the learning algorithm is given a ‘full sample’ of unlabeled points. Then, a training sample is selected uniformly at random from the full sample and the labels of the training points are revealed. The goal is to predict the labels of the remaining unlabeled points as accurately as possible. The full sample partitions the transductive hypothesis space into a finite number of *equivalence classes*. All hypotheses in the same equivalence class, generate the same dichotomy of the full sample. We consider a *large volume* principle, whereby the priority of each equivalence class is proportional to its “volume” in the hypothesis space.

The large volume principle was previously treated for the case of hyperplanes. In this paper, instead of hyperplanes, we consider soft classification vectors whose set of equivalence classes w.r.t. the full sample contains all possible dichotomies. Symmetry is broken by generating equivalence classes of non-uniform volume, defined via a non axis aligned data-dependent ellipsoid. Since exact or quantifiable approximate volume estimation is computationally hard, we resort to a cruder approach whereby volume is crudely related to the angles between hypotheses and the principal axes of the ellipsoid. This approach makes sense because long principal axes lie in regions of large volume. Our construction leads to a family of transductive algorithms and here we focus on one instantiation. Although the resulting algorithm is defined in terms of a non-convex optimization problem, we develop an efficient global optimum solution using a known technique. We also derive a data-dependent error bound for this algorithm.

Our experiments with the new Approximate Volume Regularization (AVR) algorithm over 31 datasets show its overwhelming advantage over TSVM and SVM in text categorization and image classification. However, on a different set of UCI datasets, TSVM and SVM are significantly superior to AVR. We identify some factors that influence the success and failure of our algorithm. One interesting observation is that AVR has significant advantage over TSVM when TSVM outperforms SVM, and vice versa.

^{*} This is an extended abstract of an article published in the Machine Learning Journal [1].

References

1. El-Yaniv, R., Pechyony, D., Vapnik, V.: Large margin vs. large volume in transductive learning. *Machine Learning* 72(3), 173–188 (September 2008)