Sparse Kernel SVMs via Cutting-Plane Training*

Thorsten Joachims and Chun-Nam John Yu

Cornell University, Dept. of Computer Science, Ithaca, NY 14853 USA
{tj,cnyu}@cs.cornell.edu

While Support Vector Machines (SVMs) with kernels offer great flexibility and prediction performance on many application problems, their practical use is often hindered by the following two problems. Both problems can be traced back to the number of Support Vectors (SVs), which is known to generally grow linearly with the data set size [1]. First, training is slower than other methods and linear SVMs, where recent advances in training algorithms vastly improved training time. Second, since the prediction rule takes the form $h(x) = \text{sign}\left[\sum_{i=1}^{\#SV} \alpha_i K(x_i, x)\right]$ it is too expensive to evaluate in many applications when the number of SVs is large.

This paper tackles these two problems by generalizing the notion of Support Vector to arbitrary points in input space, not just training vectors. Unlike Wu et al. [2], who explore making the location of the points part of a large non-convex optimization problem, we propose an algorithm that iteratively constructs the set of basis vectors from a cutting-plane model. This makes our algorithm, called *Cutting-Plane Subspace Pursuit* (CPSP), efficient and modular. We analyze the training efficiency and the solution quality of the CPSP algorithm both theoretically and empirically. We find that its classification rules can be orders of magnitude sparser than the conventional support-vector representation while providing comparable prediction accuracy. The sparsity of the CPSP representation not only makes predictions substantially more efficient, it also allows the user to control training time. Especially for large datasets with sparse feature vectors (e.g. text classification), the CPSP methods is substantially faster than methods that only consider basis vectors from the training set.

References

1. Steinwart, I.: Sparseness of support vector machines. JMLR 4, 1071–1105 (2003)
2. Wu, M., Schölkopf, B., Bakir, G.H.: A direct method for building sparse kernel learning algorithms. JMLR 7, 603–624 (2006)
3. Joachims, T., Yu, C.-N.J.: Sparse Kernel SVMs via Cutting-Plane Training. Machine Learning (2009), DOI:10.1007/s10994-009-5126-6

---

* This is an extended abstract of an article published in the machine learning journal [3].