An Experimental Study on Geometric Support Vector Machines

Yoshihiko Nomura   Takemasa Yamasaki   Kazushi Ikeda
Department of Systems Science
Kyoto University
Kyoto 606-8501 Japan

Abstract

The Nu support vector machine is geometrically characterized as the problem of finding the shortest segment between two reduced convex hulls, each of which is made from the set of given examples belonging to one class. This paper discusses what happens if each reduced convex hull is replaced with the set of its vertices, which may lead to a smaller complexity. Our experimental study shows this substitution makes the performance much worse, which means that the SVM solution is not a vertex in many cases.

1 Introduction

A support vector machine (SVM) is a classifier that nonlinearly maps given input vectors to feature vectors in a high-dimensional space, and that linearly separates the feature vectors with an optimal hyperplane in terms of margin [1–4]. Although an SVM has good properties such that there are no local minima in its error surface and it has a high generalization ability, it requires a high computational complexity since the problem is equivalent to a quadratic programming (QP) with variables of the same number as given examples.

To reduce the complexity, Mavroforakis and Theodoridis [5] proposed an SVM implement based on geometric properties of the nu-SVM, a variant of SVMs proposed in [6] and analyzed in [7–10]. An important property of the nu-SVM utilized in [5] as well as the others is the fact that the SVM solution is strongly related to the reduced convex hull (RCH) of given examples in the feature space [11]. In the case of homogeneous separating hyperplane being employed, especially, the problem of SVM with soft margins is equivalent to finding the point nearest the origin in the RCH.

The nearest point may be located on a surface or an edge of the RCH. The point is sometimes a vertex of the RCH. As shown in [5], any vertex of the RCH is a weighted sum of given examples where the weight takes one of the fixed three values. [5] utilized this property and reduced the complexity of the nu-SVM.

Suppose that we substitute the nearest vertex for the nearest point in the RCH. Although this operation obviously degrades the performance, it may reduce the complexity drastically at an expense of a slightly lower performance. The purpose of this study is to elucidate the trade-off experimentally.

Note that we treat only the linear kernel, that is, the input and feature spaces are identical, through this paper for simplicity.

2 Geometry of the Nu Support Vector Machines

The nu support vector machines (ν-SVMs) are a variation of SVMs where the margin is not set to unity but a variable β which is maximized as much as possible, differently from the original SVMs [6]. That is, given N input vectors $\boldsymbol{f}^{(n)}$ and the corresponding outputs $y^{(n)}$, the ν-SVM is formulated as

$$\min_{\boldsymbol{w},b,\xi_n,\beta} \left[\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{n=1}^{N}\xi_n - \beta\right]$$
$$\text{s.t. } y^{(n)}(\boldsymbol{w}'\boldsymbol{f}^{(n)} + b) \geq \beta - \xi_n, \quad \xi_n \geq 0, \tag{1}$$

where ξ_n are slack variables for the soft margin technique.

If we define $\tilde{\boldsymbol{w}} = (\boldsymbol{w}; b) \in \tilde{F}$ and $\tilde{\boldsymbol{f}} = (\boldsymbol{f}; 1) \in \tilde{F}$ where h is a positive constant and $\tilde{F}$ is the augmented input space $F \times R$, the separating hyperplane $\boldsymbol{w}'\boldsymbol{f} + b = 0$ is expressed as a simple inner product $\tilde{\boldsymbol{w}}'\tilde{\boldsymbol{f}} = 0$, that is, the hyperplane is homogeneous. This operation is called lifting-up. The ν-SVM with homogeneous hyperplanes seems equivalent to but differs from the original ν-SVM (1), since the former also penalizes the offset in the cost function due to

$$\|\tilde{\boldsymbol{w}}\|^2 = \|\boldsymbol{w}\|^2 + b^2. \tag{2}$$

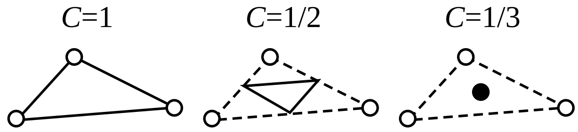

Figure 1: The reduced convex hull of examples. When C is the reciprocal to the size of the example set, it reduces to the centroid of the examples.

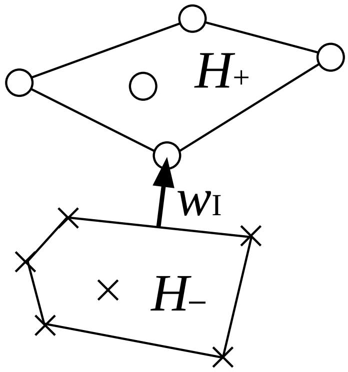

Figure 2: A geometrical view of the solution of a ν-SVM.

See [10] for the details on the effect of lifting-up.

The Wolfe dual problem of (1) is derived as

$$\min_{\alpha_n} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{s.t. } \boldsymbol{w} = \sum_{n=1}^{N} \alpha_n y^{(n)} \boldsymbol{f}^{(n)},$$

$$0 \le \alpha_n \le C, \quad \sum_{n=1}^{N} \alpha_n = 1, \quad \sum_{n=1}^{N} y^{(n)} \alpha_n = 0, \qquad (3)$$

where α_n are the Lagrange multipliers. Let $\tilde{\alpha}_n = 2\alpha_n$ and $\tilde{C} = 2C$. Then, (3) is written as

$$\min_{\tilde{\alpha}_n} \frac{1}{2}\|\boldsymbol{w}\|^2$$

$$\text{s.t. } \boldsymbol{w} = \frac{1}{2} \sum_{y^{(n)}=1} \tilde{\alpha}_n \boldsymbol{f}^{(n)} - \frac{1}{2} \sum_{y^{(n)}=-1} \tilde{\alpha}_n \boldsymbol{f}^{(n)}, \qquad (4)$$

$$0 \le \tilde{\alpha}_n \le \tilde{C}, \quad \sum_{y^{(n)}=1} \tilde{\alpha}_n = \sum_{y^{(n)}=-1} \tilde{\alpha}_n = 1. \quad (5)$$

Here, (5) means that the first term of (4) is a vector in the convex hull of the examples with $y^{(n)} = 1$ and the second with $y^{(n)} = -1$ since the sum of $alp\tilde{ha}_n \ge 0$ is unity. And the restriction of the weight $\tilde{\alpha}$ between 0 and C reduces the convex hull to the so-called reduced convex hull (Fig. 1). Since the distance between such vectors is minimized in (4), the ν-SVM solution has the direction vector $\boldsymbol{w}_\mathrm{I}$ parallel to the segment connecting the two reduced convex hulls H_+ and H_- and having the minimum length (Fig. 2), while the bias of the solution is a little different from the center of the segment [10–12]. This clear geometrical picture is a reason why the ν-SVM is preferred to the original SVM in theoretical studies.

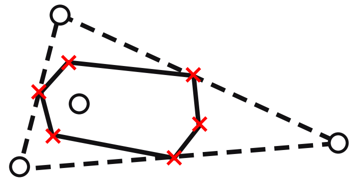

Figure 3: A reduced convex hull is the convex hull of centroids.

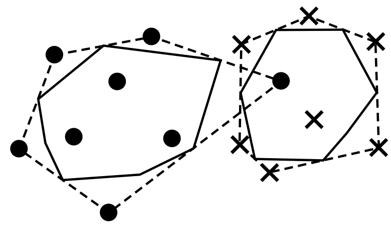

Figure 4: Soft-margin technique makes the examples linearly separable.

3 Discrete Support Vector Machines

Suppose $C = 1/M$ where M is an integer, since the generalization is straightforward. Then, the reduced convex hull of an example set is equivalent to the convex hull of the set which consists of the centroids of M distinct examples (Fig. 3). This fact clearly shows that the soft-margin technique, introducing C less than unity, is to change the distribution of the input vectors, that is, to make it milder by averaging and linearly separable (Fig. 4) [7]. Since the reduced convex hull shrinks to the centroid of all examples as $C \to 1/N$, any problem becomes separable then, except some special cases.

Suppose that we substitute the nearest vertex for the nearest point in the RCH. Although this operation obviously degrades the performance, it may reduce the complexity drastically at an expense of a slightly lower performance, since $\tilde{\alpha}_n$ takes either of the fixed values, 0 or $C = 1/M$. We term this method the discrete SVM in this paper.

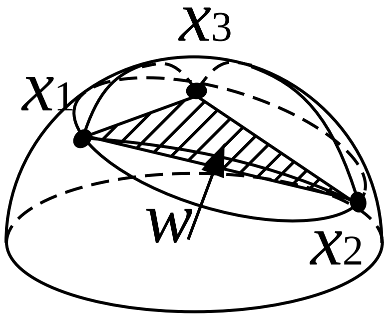

Figure 5: A geometrical view of the ν-SVM solution.

The discrete SVM produces a different hyperplane from the ν-SVM, except for special cases, e.g., when the input vectors distribute on the 1-dimensional hypersphere and the soft-margin parameter C is a half. From the geometrical viewpoint, they have different pictures: In homogeneous hyperplanes' case, the ν-SVM solution is the center of the minimum circumscribed circle including all the examples [8, 13] while the nearest vertex is the centroid of M examples (Fig. 5). We experimentally elucidate their difference in the next section.

4 Computer Simulations

We carried out some computer simulations to compare the discrete SVM with the ν-SVM. The discrete SVM is given N examples, where the input vector $\boldsymbol{x}^{(n)} \in R^K$ of the nth example $(\boldsymbol{x}^{(n)}, y^{(n)})$ obeys the normal distribution, $\mathrm{N}((10y^{(n)}, \mathbf{0}_{K-1})', I_K)$, and $y^{(n)} = 1$ for $0 < n \leq N/2$ and $y^{(n)} = 1$ for $N/2 < n \leq N$.

It is known that the ν-SVM has the average generalization error of order $1/N$ [14, 15]. Hence, we evaluate the average generalization error of the discrete SVM in our experiments, which is given from the angle between the weight vector of the discrete SVM and the true one. The results are shown in Fig. 6, where K is 2 or 3 and N is 100 to 1000. As far as we see in the results, the average generalization errors did not decrease even when the number of examples, N, increases. This phenomenon is very strange.

5 Conclusions

We proposed the discrete SVM, which substitutes the nearest vertex for the nearest point in the RCH, to reduce the computational complexity. However, our experimental study of evaluating its performance showed that this algorithm is far from practical at this moment. We will elucidate the reason why the discrete SVM has such a strange learning curve and propose an improved algorithm in the near future.

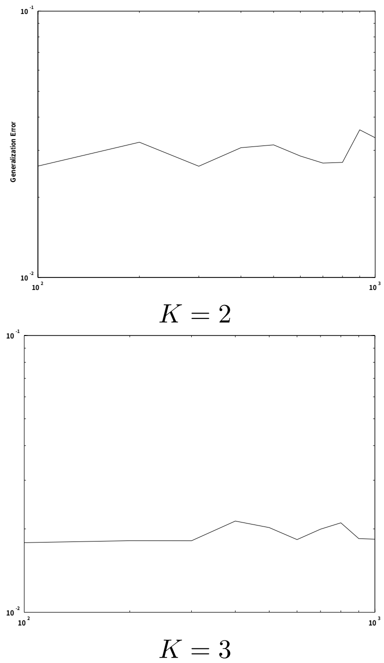

Figure 6: Generalization errors versus the number of examples.

Acknowledgments

This study is supported in part by a Grant-in-Aid for Scientific Research (15700130, 18300078) from the Japanese Government.

References

[1] V. N. Vapnik, *The Nature of Statistical Learning Theory.* Springer-Verlag, 1995.

[2] B. Schölkopf, et al., *Advances in Kernel Methods: Support Vector Learning.* Cambridge Univ. Press, 1998.

[3] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines.* Cambridge Univ. Press, 2000.

[4] A. J. Smola, et al. eds., *Advances in Large Margin Classifiers.* MIT Press, 2000.

[5] M. E. Mavroforakis and S. Theodoridis, "A geometric approach to support vector machine (SVM) classification," *IEEE Trans. NN*, vol. 3, pp. 671–682, 2006.

[6] B. Schölkopf, et al., "New support vector algorithms," *Neu. Comp.*, vol. 12, no. 5, pp. 1207–1245, 2000.

[7] K. Ikeda and T. Aoishi, "An asymptotic statistical analysis of support vector machines with soft margins," *Neu. Netw.*, vol. 18, no. 3, pp. 251–259, 2005.

[8] K. Ikeda and N. Murata, "Geometrical properties of nu support vector machines with different norms," *Neu. Comp.*, vol. 17, no. 11, pp. 2508–2529, 2005.

[9] K. Ikeda, "Effects of kernel function on nu support vector machines in extreme cases," *IEEE Trans. NN*, vol. 17, no. 1, pp. 1–9, 2006.

[10] ——, "Geometrical properties of lifting-up in the nu support vector machines," *IEICE Trans. Inf. and Sys.*, vol. E89-D, no. 2, pp. 847–852, 2006.

[11] K. P. Bennett and E. J. Bredensteiner, "Duality and geometry in SVM classifiers," *Proc. ICML*, pp. 57–64, 2000.

[12] D. Crisp and C. Burges, "A geometric interpretation of nu-SVM classifiers," *NIPS*, vol. 12, pp. 244–250, 2000.

[13] R. Herbrich, *Learning Kernel Classifiers: Theory and Algorithms.* MIT Press, 2002.

[14] K. Ikeda, "An asymptotic statistical theory of polynomial kernel methods," *Neu. Comp.*, vol. 16, no. 8, pp. 1705–1719, 2004.

[15] K. Ikeda and T. Yamasaki, "Incremental support vector machines and their geometrical analyses," *Neurocomp.*, in press.