# Analysis of support vector machines

Abe, Shigeo

# ANALYSIS OF SUPPORT VECTOR MACHINES

Shigeo Abe
Graduate School of Science and Technology
Kobe University, Kobe, Japan
Phone/Fax: +81 78 803 6101
E-mail: abe@eedept.kobe-u.ac.jp
Web: www.eedept.kobe-u.ac.jp/eedept_english.html

**Abstract.** **We compare L1 and L2 soft margin support vector machines from the standpoint of positive definiteness, the number of support vectors, and uniqueness and degeneracy of solutions. Since the Hessian matrix of L2 SVMs is positive definite, the number of support vectors for L2 SVMs is larger than or equal to the number of L1 SVMs. For L1 SVMs, if there are plural irreducible sets of support vectors, the solution of the dual problem is non-unique although the primal problem is unique. Similar to L1 SVMs, degenerate solutions, in which all the data are classified into one class, occur for L2 SVMs.**

## INTRODUCTION

As opposed to L2 soft margin support vector machines (L2 SVMs), L1 soft margin support vector machines (L1 SVMs) are widely used for pattern classification and function approximation. And much effort has been done to clarify the properties of L1 SVMs [1, 2, 3, 4]. Pontil and Verri [1] clarified dependence of the L1 SVM solutions on the margin parameter $C$. Rifkin, Pontil, and Verri [2] showed degeneracy of L1 SVM solutions, in which any data are classified into one class. Fernández [3] also proved the existence of degeneracy without mentioning it. Burges [4] discussed non-uniqueness of L1 SVM primal solutions, and uniqueness of L2 SVM solutions. But except for [4], little comparison has been made between L1 and L2 SVMs [5].

In this paper, we compare L1 SVMs with L2 SVMs from the standpoint of positive definiteness, the number of support vectors, and uniqueness and degeneracy of solutions. Since the Hessian matrix of L2 SVMs is positive definite, the solutions are unique. For the L1 SVMs, we introduce the concept of irreducible set of support vectors and show that if there are plural irreducible sets, the dual solutions are non-unique. Finally, we show that L2 SVMs have degenerate solutions similar to L1 SVMs.

In the following, first we summarize L1 and L2 SVMs and discuss the Hessian matrices for L1 and L2 SVMs. Then we discuss non-uniqueness of

L1 SVM dual solutions. Finally, we prove the existence of degeneracy for L2 SVMs.

## SOFT MARGIN SUPPORT VECTOR MACHINES

In soft margin support vector machines, we consider the linear decision function

$$D(\mathbf{x}) = \mathbf{w}^t \mathbf{g}(\mathbf{x}) + b \tag{1}$$

in the feature space, where $\mathbf{w}$ is the weight vector, $\mathbf{g}(\mathbf{x})$ is the mapping function that maps the $m$-dimensional input $\mathbf{x}$ into the $l$-dimensional feature space, and $b$ is a scalar. We determine the decision function so that the classification error for the training data and unknown data is minimized. This can be achieved by minimizing

$$\frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{M} \xi_i^p \tag{2}$$

subject to the constraints

$$y_i \left( \mathbf{w}^t \mathbf{g}(\mathbf{x}_i) + b \right) \geq 1 - \xi_i \qquad \text{for} \qquad i = 1, \ldots, M, \tag{3}$$

where $\xi_i$ are the positive slack variables associated with the training data $\mathbf{x}_i$, $M$ is the number of training data, $y_i$ are the class labels (1 or $-1$) for $\mathbf{x}_i$, $C$ is the margin parameter, and $p$ is either 1 or 2. When $p = 1$, we call the support vector machine L1 soft margin support vector machine (L1 SVM) and when $p = 2$, L2 soft margin support vector machine (L2 SVM).

The dual problem for the L1 SVM is to maximize

$$Q(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} \alpha_i \, \alpha_j \, y_i \, y_j \, H(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

subject to the constraints

$$\sum_{i=1}^{M} y_i \, \alpha_i = 0, \qquad 0 \leq \alpha_i \leq C, \tag{5}$$

where $\alpha_i$ are the Lagrange multipliers associated with the training data $\mathbf{x}_i$ and $H(\mathbf{x}, \mathbf{x}') = \mathbf{g}(\mathbf{x})^t \mathbf{g}(\mathbf{x})$ is the kernel function.

The dual problem for the L2 SVM is to maximize

$$Q(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} \alpha_i \, \alpha_j \, y_i \, y_j \, \left( H(\mathbf{x}_i, \mathbf{x}_j) + \frac{\delta_{ij}}{C} \right) \tag{6}$$

subject to the constraints

$$\sum_{i=1}^{M} y_i \, \alpha_i = 0, \qquad \alpha_i \geq 0. \tag{7}$$

Let the solution of (4) and (5), or (6) and (7) be $\alpha_i^* \, (i = 1, \ldots, M)$.

## HESSIAN MATRIX

Rewriting (4) for the L1 SVM using the mapping function $\mathbf{g}(\mathbf{x})$, we have

$$Q(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \left( \sum_{i=1}^{M} \alpha_i \, y_i \, \mathbf{g}(\mathbf{x}_i) \right)^t \sum_{i=1}^{M} \alpha_i \, y_i \, \mathbf{g}(\mathbf{x}_i). \tag{8}$$

Solving (5) for $\alpha_s$ $(s \in \{1, \ldots, M\})$,

$$\alpha_s = -y_s \sum_{\substack{i=1 \\ i \neq s}}^{M} y_i \, \alpha_i. \tag{9}$$

Substituting (9) into (8), we obtain

$$
\begin{aligned}
Q(\alpha) &= \sum_{\substack{i=1 \\ i \neq s}}^{M} (1 - y_s \, y_i)\alpha_i \\
&\quad - \frac{1}{2} \left( \sum_{\substack{i=1 \\ i \neq s}}^{M} \alpha_i \, y_i \, (\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_s)) \right)^t \sum_{\substack{i=1 \\ i \neq s}}^{M} \alpha_i \, y_i \, (\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_s)). 
\end{aligned}
\tag{10}
$$

Thus the Hessian matrix of $-Q(\alpha)$, $H_{L1}$, which is an $(M-1) \times (M-1)$ matrix, is given by

$$
\begin{aligned}
H_{L1} &= -\frac{\partial Q(\alpha)^2}{\partial^2 \alpha'} \\
&= \left( \cdots y_i(\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_s)) \cdots \right)^t \left( \cdots y_j(\mathbf{g}(\mathbf{x}_j) - \mathbf{g}(\mathbf{x}_s)) \cdots \right),
\end{aligned}
\tag{11}
$$

where $\alpha'$ is obtained by deleting $\alpha_s$ from $\alpha$. Since $H_{L1}$ is expressed by the product of a transposed matrix and the matrix, $H_{L1}$ is positive semi-definite. Let $N_{\mathbf{g}}$ be the maximum number of independent vectors among $\{\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_s) \, | \, i \in \{i, \ldots, M, i \neq s\}$. Since $N_{\mathbf{g}}$ does not exceeds the dimension of the feature space, $l$, the rank of $H_{L1}$ is given by [6, pp. 311–312]

$$\min(l, N_{\mathbf{g}}). \tag{12}$$

Therefore, if $M > (l+1)$, $H_{L1}$ is positive semi-definite.

The Hessian matrix $H_{L2}$, in which one variable is eliminated, for the L2 SVM is expressed by

$$H_{L2} = H_{L1} + \left\{ \frac{y_i y_j + \delta_{ij}}{C} \right\}. \tag{13}$$

Because of the term $\sum_{i=1}^{M} \alpha_i^2 / C$ in $Q(\alpha)$, $H_{L2}$ is positive definite. Thus, unlike $H_{L1}$, $H_{L2}$ is positive definite irrespective of the dimension of the feature space.
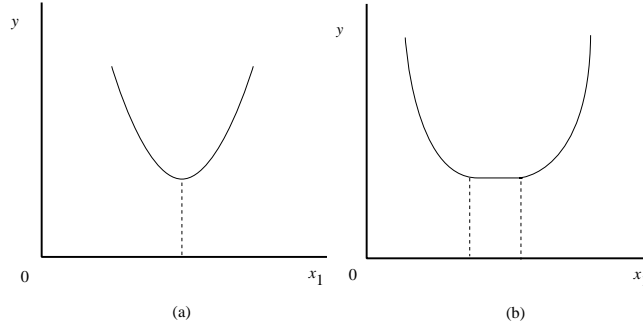
Figure 1: Convex functions. (**a**) Strictly convex function. (**b**) Convex function

Table 1: CONVEXITY OF OBJECTIVE FUNCTIONS

|        | Hard Margin           | L1 Soft Margin          | L2 Soft Margin          |
|--------|-----------------------|-------------------------|-------------------------|
| Primal | Strictly Convex       | Convex                  | Strictly Convex         |
|        | $(\mathbf{w}, b)$     | $(\mathbf{w}, b, \xi_i)$ | $(\mathbf{w}, b, \xi_i)$ |
| Dual   | Convex                | Convex                  | Strictly Convex         |
|        | $(\alpha_i)$          | $(\alpha_i)$            | $(\alpha_i)$            |

## NON-UNIQUE SOLUTIONS

If a convex function gives a minimum or maximum at a point not in an interval, the function is called strictly convex. In general, if the objective function of a quadratic programming problem constrained in a convex set is strictly convex or the associated Hessian matrix is positive (negative) definite, the solution is unique. And if the objective function is convex, there may be cases where the solution is non-unique (see Fig. 1). Convexity of objective functions for different support vector machine architectures is summarized in Table 1. The symbols in the brackets show the variables.

We must notice that since $b$ is not included in the dual problem, even if the solution of the dual problem is unique, the solution of the primal problem may not be unique [4].

Assume that the hard margin SVM has a solution, i.e., the given problem is separable in the feature space. Then, since the objective function of the primal problem is $\|\mathbf{w}\|^2/2$, which is strictly convex, the primal problem has a unique solution for $\mathbf{w}$ and $b$. But the dual solution may be non-unique because the Hessian matrix is positive semi-definite. Since the hard margin SVM for a separable problem is equivalent to the L1 SVM with an unbounded solution, we leave the discussion to that for the L1 SVM.

The objective function of the primal problem for the L2 SVM is strictly convex. Therefore, $\mathbf{w}$ and $b$ are uniquely determined if we solve the primal problem. In addition, since the Hessian matrix of the dual objective function
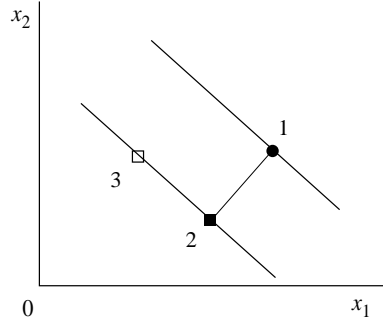
Figure 2: An example of a non-support vector

is positive definite, $\alpha_i$ are uniquely determined. And because of the uniqueness of the primal problem, $b$ is determined uniquely using the Kuhn-Tucker condition.

The L1 SVM includes the linear sum of $\xi_i$. Therefore, the primal objective function is convex. Likewise, the Hessian matrix of the dual objective function is positive semi-definite. Thus the primal and dual solutions may be non-unique.

**Theorem 1** For the L1 SVM, vectors that satisfy $y_i(\mathbf{w}^t\mathbf{g}(\mathbf{x}_i) + b) = 1$ are not always support vectors.

*Proof.* Consider the two-dimensional case shown in Fig. 2. In the figure, $\mathbf{x}_1$ belongs to Class 1, $\mathbf{x}_2$ and $\mathbf{x}_2$ belong to Class 2, and $\mathbf{x}_1 - \mathbf{x}_2$ and $\mathbf{x}_3 - \mathbf{x}_2$ are orthogonal. The dual problem with the dot product kernel is given as follows. Maximize

$$Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3$$
$$-\frac{1}{2}(\alpha_1\mathbf{x}_1 - \alpha_2\mathbf{x}_2 - \alpha_3\mathbf{x}_3)^t(\alpha_1\mathbf{x}_1 - \alpha_2\mathbf{x}_2 - \alpha_3\mathbf{x}_3) \quad (14)$$

subject to

$$\alpha_1 - \alpha_2 - \alpha_3 = 0, \quad C \geq \alpha_i \geq 0, \quad i = 1, 2, 3. \quad (15)$$

Substituting $\alpha_3 = \alpha_1 - \alpha_2$ and $\alpha_2 = a\,\alpha_1$ $(a \geq 0)$ into (14), we obtain

$$Q(\alpha) = 2\,\alpha_1$$
$$-\frac{1}{2}\alpha_1^2(\mathbf{x}_1 - \mathbf{x}_3 - a(\mathbf{x}_2 - \mathbf{x}_3))^t(\mathbf{x}_1 - \mathbf{x}_3 - a(\mathbf{x}_2 - \mathbf{x}_3)). \quad (16)$$

Defining

$$d^2(a) = (\mathbf{x}_1 - \mathbf{x}_3 - a(\mathbf{x}_2 - \mathbf{x}_3))^t(\mathbf{x}_1 - \mathbf{x}_3 - a(\mathbf{x}_2 - \mathbf{x}_3)), \quad (17)$$

(16) becomes

$$Q(\alpha) = 2\,\alpha_1 - \frac{1}{2}\alpha_1^2 d^2(a). \quad (18)$$

When

$$C \geq \frac{2}{d^2(a)}, \tag{19}$$

$Q(\alpha)$ is maximized at $\alpha_1 = 2/d^2(a)$ and takes the maximum

$$Q\left(\frac{2}{d^2(a)}\right) = \frac{2}{d^2(a)}. \tag{20}$$

Since $\mathbf{x}_1 - \mathbf{x}_2$ and $\mathbf{x}_3 - \mathbf{x}_2$ are orthogonal, $d(a)$ is minimized at $a = 1$. Thus $Q(2/d^2(a))$ is maximized at $a = 1$. Namely, $\alpha_1 = \alpha_2 = 2/d^2(a)$ and $\alpha_3 = 0$. Since $y_3(\mathbf{w}^t\mathbf{x}_3 + b) = 1$, the theorem is proved.

**Definition 1** For L1 SVMs a set of support vectors is *irreducible* if deletion of all the non-support vectors that satisfy $y_i(\mathbf{w}^t\mathbf{g}(\mathbf{x}_i) + b) = 1$, and any support vector results in the change of the optimal hyperplane. It is *reducible* if the optimal hyperplane does not change for deletion of all the non-support vectors that satisfy $y_i(\mathbf{w}^t\mathbf{g}(\mathbf{x}_i) + b) = 1$, and some support vectors.

Deletion of non-support vectors from the training data set does not change the solution. In the following theorem, the Hessian matrix associated with a set of support vectors means that the Hessian matrix is calculated for the support vectors, not the entire training data.

**Theorem 2** For L1 SVMs, let all the support vectors be unbounded. Then the Hessian matrix associated with an irreducible set of support vectors is positive definite and the Hessian matrix associated with a reducible set of support vectors is positive semi-definite.

*Proof.* Let the set of support vectors be irreducible. Then, since deletion of any support vector results in the change of the optimal hyperplane, any $\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_s)$ cannot be expressed by the remaining $\mathbf{g}(\mathbf{x}_j) - \mathbf{g}(\mathbf{x}_s)$. Thus the associated Hessian matrix is positive definite. If the set of support vectors is reducible, deletion of some support vector, e.g., $\mathbf{x}_i$ does not cause the change of the optimal hyperplane. This means that $\mathbf{g}(\mathbf{x}_i) - \mathbf{g}(\mathbf{x}_s)$ is expressed by the linear sum of the remaining $\mathbf{g}(\mathbf{x}_j) - \mathbf{g}(\mathbf{x}_s)$. Thus the associated Hessian matrix is positive semi-definite.

**Theorem 3** For the L1 SVM, if there is only one irreducible set of support vectors and the support vectors are all unbounded, the solution is unique.

*Proof.* Delete the non-support vectors from the training data. Then since the set of support vectors is irreducible the associated Hessian matrix is positive definite. Thus, the solution is unique for the irreducible set. Since there is only one irreducible set, the solution is unique for the given problem.

**Example 1** Consider the two-dimensional case shown in Fig. 3, in which $\mathbf{x}_1$ and $\mathbf{x}_2$ belong to Class 1 and $\mathbf{x}_3$ and $\mathbf{x}_4$ belong to Class 2. Since $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$,
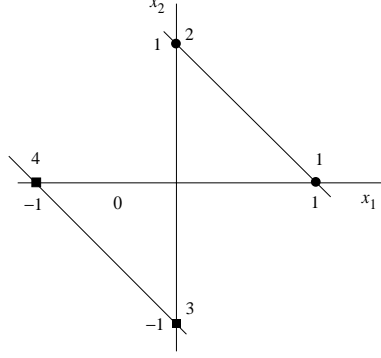
Figure 3: Non-unique solutions

and $\mathbf{x}_4$ form a rectangle, $\{\mathbf{x}_1, \mathbf{x}_3\}$ and $\{\mathbf{x}_2, \mathbf{x}_4\}$ are irreducible sets of support vectors for the dot product kernel.

The training is to maximize

$$Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}\left((\alpha_1 + \alpha_4)^2 + (\alpha_2 + \alpha_3)^2\right) \qquad (21)$$

subject to

$$\alpha_1 + \alpha_2 = \alpha_3 + \alpha_4, \quad C \geq \alpha_i \geq 0, \quad i = 1, \ldots, 4. \qquad (22)$$

For $C \geq 1$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (1, 0, 1, 0)$ and $(0, 1, 0, 1)$ are two solutions. Thus,

$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (\beta, 1 - \beta, \beta, 1 - \beta), \qquad (23)$$

where $0 \leq \beta \leq 1$, is also a solution. Then, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.5, 0.5, 0.5, 0.5)$ is a solution.

For the L2 SVM, the objective function becomes

$$\begin{aligned} Q(\alpha) = {} & \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \\ & - \frac{1}{2}\left((\alpha_1 + \alpha_4)^2 + (\alpha_2 + \alpha_3)^2 + \frac{\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \alpha_4^2}{C}\right). \end{aligned} \qquad (24)$$

Then for $\alpha_i = 1/(2 + 1/C)$ $(i = 1, \ldots, 4)$, (24) becomes

$$Q(\alpha) = \frac{1}{1 + \dfrac{1}{2C}}. \qquad (25)$$

For $\alpha_1 = \alpha_3 = 1/(1 + 1/C)$ and $\alpha_2 = \alpha_4 = 0$, (24) becomes

$$Q(\alpha) = \frac{1}{1 + \dfrac{1}{C}}. \qquad (26)$$

Thus, for $C > 0$, $Q(\alpha)$ given by (26) is smaller than that by (25). Therefore, $\alpha_1 = \alpha_3 = 1/(1 + 1/C)$ and $\alpha_2 = \alpha_4 = 0$ or $\alpha_2 = \alpha_4 = 1/(1 + 1/C)$ and $\alpha_1 = \alpha_3 = 0$ are not optimal, but $\alpha_i = 1/(2 + 1/C)$ $(i = 1, \ldots, 4)$ are.

In general, the number of support vectors for L1 SVMs is larger than that for L2 SVMs.

The above example shows non-uniqueness of the dual problem but the primal problem is unique since there are unbounded support vectors. Non-unique solutions occur when there are no unbounded support vectors. Burges and Crisp [4] derive conditions in which the dual solution is unique but the primal solution is non-unique.

## DEGENERATE SOLUTIONS

Rifkin, Pontil, and Verri [2] discuss degenerate solutions, in which $\mathbf{w} = \mathbf{0}$, for L1 SVMs. Fernández [3] derived similar results for L1 SVMs, although he did not refer to degeneracy. Degeneracy occurs also for L2 SVMs. In the following we discuss degenerate solutions following the proof of [3].

**Theorem 4** Let $C = KC_0$, where $K$ and $C_0$ are positive parameters, and $\alpha^*$ be a solution of the L1 support vector machine with $K = 1$. Define

$$\mathbf{w}(\alpha) = \sum_{i=1}^{M} \alpha_i y_i \mathbf{g}(\mathbf{x}). \qquad (27)$$

Then the necessary and sufficient condition for

$$\mathbf{w}(\alpha^*) = \mathbf{0} \qquad (28)$$

is that $K\alpha^*$ is also a solution for any $K\,(>1)$.

*Proof.* We prove the theorem for L2 SVMs. The proof for L1 SVMs is given by deleting $\alpha^t \alpha/(2C)$ in the following proof.
**Necessary condition.** Let $\alpha'$ be the optimal solution for $K\,(K > 1)$ and $\mathbf{w}(\alpha') \neq \mathbf{0}$. Then, there is $\alpha''$ such that $\alpha' = K\alpha''$. Since $\alpha'$ satisfies the equality constraint for $K = 1$, it is a non-optimal solution. Then for $K = 1$,

$$
\begin{aligned}
Q(\alpha^*) &= \sum_{i=1}^{M} \alpha_i^* - \frac{\alpha^{*t}\alpha^*}{2C_0} \\
&\geq Q(\alpha'') = \sum_{i=1}^{M} \alpha_i'' - \frac{1}{2}\mathbf{w}(\alpha'')^t\mathbf{w}(\alpha'') - \frac{\alpha''^t\alpha''}{2C_0}. \qquad (29)
\end{aligned}
$$

For $K\,(K > 1)$,

$$
\begin{aligned}
Q(K\alpha^*) &= K\sum_{i=1}^{M} \alpha_i^* - \frac{K\alpha^{*t}\alpha^*}{2C_0} \leq Q(K\alpha'') \\
&= K\sum_{i=1}^{M} \alpha_i'' - \frac{K^2}{2}\mathbf{w}(\alpha'')^t\mathbf{w}(\alpha'') - \frac{K\alpha''^t\alpha''}{2C_0}. \qquad (30)
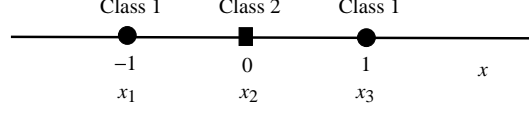\end{aligned}
$$

Figure 4: Non-separable one-dimensional case

Multiplying $K$ to all the terms in (29) and comparing it with (30), we see the contradiction. Thus, $K\alpha^*$ is the optimal solution for $K > 1$.

**Sufficient condition.** Suppose $K\alpha^*$ is the optimal solution for any $K \, (\geq 1)$. Thus for any $K \, (\geq 1)$

$$
\begin{aligned}
Q(K\alpha^*) &= \sum_{i=1}^{M} K\alpha_i^* - \frac{1}{2}K^2\mathbf{w}(\alpha^*)^t\mathbf{w}(\alpha^*) - \frac{K\alpha^{*t}\alpha^*}{2C_0} \\
&\geq Q(\alpha^*) = \sum_{i=1}^{M} \alpha_i^* - \frac{1}{2}\mathbf{w}(\alpha^*)^t\mathbf{w}(\alpha^*) - \frac{\alpha^{*t}\alpha^*}{2C_0}.
\end{aligned}
\tag{31}
$$

Rewriting (31), we have

$$
\sum_{i=1}^{M} \alpha_i^* \geq \frac{K+1}{2}\mathbf{w}(\alpha^*)^t\mathbf{w}(\alpha^*) + \frac{\alpha^{*t}\alpha^*}{2C_0}.
\tag{32}
$$

Since (32) is satisfied as $K$ approaches infinity, $\mathbf{w}(\alpha^*) = \mathbf{0}$ must be satisfied. Otherwise, $K\alpha^*$ cannot be the optimal solution.

**Example 2** Consider the case shown in Fig. 4. Here, we use the dot product kernel. The inequality constraints are

$$
\begin{aligned}
-w + b &\geq 1 - \xi_1, & (33) \\
-b &\geq 1 - \xi_2, & (34) \\
w + b &\geq 1 - \xi_3. & (35)
\end{aligned}
$$

The dual problem for the L1 SVM is given as follows: Namely maximize

$$
Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(-\alpha_1 + \alpha_3)^2
\tag{36}
$$

subject to

$$
\begin{aligned}
\alpha_1 - \alpha_2 + \alpha_3 &= 0, & (37) \\
C \geq \alpha_i \geq 0 \quad &\text{for} \quad i = 1, 2, 3. & (38)
\end{aligned}
$$

From (37), $\alpha_2 = \alpha_1 + \alpha_3$. Then substituting it into (36), we obtain

$$
\begin{aligned}
Q(\alpha) &= 2\alpha_1 + 2\alpha_3 - \frac{1}{2}(-\alpha_1 + \alpha_3)^2, & (39) \\
&\quad C \geq \alpha_i \geq 0 \quad \text{for} \quad i = 1, 2, 3. & (40)
\end{aligned}
$$

Equation (39) is maximized when $\alpha_1 = \alpha_3$. Thus the optimal solution is given by

$$\alpha_1 = \frac{C}{2}, \quad \alpha_2 = C, \quad \alpha_3 = \frac{C}{2}. \tag{41}$$

Therefore, $x = -1$, 0, and 1 are support vectors. Thus from (27), $w = 0$. Substituting $w = \xi_1 = \xi_3 = 0$ into (33) and (35) and taking the equalities, we obtain $b = 1$.

The dual objective function for the L2 SVM is given by

$$Q(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(-\alpha_1 + \alpha_3)^2 - \frac{\alpha_1^2 + \alpha_2^2 + \alpha_3^2}{2C} \tag{42}$$

The objective function is maximized when

$$\alpha_1 = \alpha_2 = \frac{2C}{3}, \quad \alpha_2 = \frac{4C}{3}. \tag{43}$$

Thus, $w = 0$ and $b = 1/3$. Therefore, any datum is classified into Class 1.


## CONCLUSIONS

In this paper, we compared L1 and L2 support vector machines from the standpoint of uniqueness and degeneracy of solutions. We introduced the concept of irreducible set of support vectors and clarified the condition for non-uniqueness of L1 SVM dual solutions. We also proved that degeneracy of L2 SVM solutions occurs.


## REFERENCES

[1] M. Pontil and A. Verri, "Properties of support vector machines," **Neural Computation**, Vol. 10, No. 4, pp. 955–974, 1998.

[2] R. Rifkin, M. Pontil, and A. Verri, "A note on support vector machine degeneracy," **Proceedings of 10th International Conference on Algorithmic Learning Theory, (ALT'99). (Lecture Notes in Artificial Intelligence Vol. 1720)**, pp. 252–263, 1999.

[3] R. Fernández, "Behavior of the weights of a support vector machine as a function of the regularization parameter c," **Proceedings of the 8th International Conference on Artificial Neural Networks (ICANN'98)**, Vol. 2, pp. 917–922, 1998.

[4] C. J. C. Burges and D. J. Crisp, "Uniqueness of the SVM solution," In S. A. Solla, T. K. Leen, and K.-R. Müller, Eds., **Advances in Neural Information Processing Systems 12**, pp. 223–229, The MIT Press, 2000.

[5] N. Cristianini and J. Shawe-Taylor, **An Introduction to Support Vector Machines and Other Kernel-based Learning Methods**, Cambridge University Press, Cambridge, UK, 2000.

[6] S. Abe, **Pattern Classification: Neuro-fuzzy Methods and Their Comparison**, Springer-Verlag, London, UK, 2001.