A Revisit to Support Vector Data Description (SVDD)

Wei-Cheng Chang B99902019@CSIE.NTU.EDU.TW
Ching-Pei Lee R00922098@CSIE.NTU.EDU.TW
Chih-Jen Lin CJLIN@CSIE.NTU.EDU.TW
Department of Computer Science, National Taiwan University, Taipei 10617, Taiwan

Abstract

Support vector data description (SVDD) is a useful method for outlier detection. Its model is obtained by solving the dual optimization problem. In this paper, we point out that the existing derivation of the dual problem contains several errors. This issue causes an incorrect dual problem under some parameters. Given the wide use of SVDD, it is important to correct these mistakes. We provide a rigorous derivation of the dual problem, discuss additional properties, and investigate some extensions of SVD.

1 Introduction

Support vector data description (SVDD) by Tax and Duin (2004) is a method to find the boundary around a data set. SVDD has been successfully applied in a wide variety of application domains such as handwritten digit recognition (Tax and Duin, 2002), face recognition (Lee et al., 2006), pattern denoising (Park et al., 2007) and anomaly detection (Banerjee et al., 2007).

Given a set of training data $\boldsymbol{x}_i \in \mathbf{R}^n$, $i = 1, \ldots, l$, Tax and Duin (2004) solve the following optimization problem.

$$\begin{aligned} \min_{R, \boldsymbol{a}, \boldsymbol{\xi}} \quad & R^2 + C \sum_i^l \xi_i \\ \text{subject to} \quad & \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \le R^2 + \xi_i, i = 1, \ldots, l, \\ & \xi_i \ge 0, i = 1, \ldots, l, \end{aligned} \tag{1}$$

where ϕ is a function mapping data to a higher dimensional space, and $C > 0$ is a user-specified parameter. After (1) is solved, a testing instance $\boldsymbol{x}$ is detected as an outlier if

$$\|\phi(\boldsymbol{x}) - \boldsymbol{a}\|^2 > R^2.$$

Because of the large number of variables in $\boldsymbol{a}$ after data mapping, Tax and Duin (2004)

considered solving the following Lagrange dual problem.

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{l} \alpha_i Q_{i,i} - \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{e}^T \boldsymbol{\alpha} = 1, \\ & 0 \le \alpha_i \le C, i = 1, \ldots, l, \end{aligned} \tag{2}$$

where $\boldsymbol{e} = [1, \cdots, 1]^T, \boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_l]^T$, and Q is the kernel matrix such that

$$Q_{i,j} = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j), \forall 1 \le i, j \le l.$$

This problem is very similar to the support vector machine (SVM) dual problem (Boser et al., 1992; Cortes and Vapnik, 1995), and can be solved by existing optimization methods for SVM.

In this paper, we point out that the approach of Tax and Duin has the following problems.

(a) The primal problem (1) has feasible solutions for any $C > 0$.[1] However, the dual problem (2) is infeasible if $C < 1/l$. This fact implies that the primal-dual relationship does not hold.

(b) In fact, problem (1) is not convex, so the commonly used duality theory of convex programming is not applicable.

(c) In deriving the Lagrange dual problem, they did not check whether (1) satisfies the constraint qualifications needed in ensuring strong duality (i.e., primal and dual optimal values are equal).

The aim of this paper is to fix these problems and derive a valid dual optimization problem. This paper is organized as follows. In Section 2 we give details of the above-mentioned problems in Tax and Duin (2004). Section 3 then proposes a correction for these defectiveness. Section 4 further discusses some extended cases. Section 5 concludes this work.

2 Problems in the Existing Derivation of SVDD

In this Section, we detailedly discuss each problem in Tax and Duin (2004) that is mentioned in Section 1.

2.1 Convexity of (1)

For minimizing an optimization problem, before deriving the Lagrange dual problem and checking whether the dual optimal value is identical to the primal optimal value, one should make sure that the original problem is convex. Otherwise, the duality theory of convex

[1] For example, $\boldsymbol{a} = \boldsymbol{0}, R = 0$, and $\xi_i = \|\phi(\boldsymbol{x}_i)\|^2, \forall i$.

programming may not be applicable. Here we show that problem (1) is non-convex. The following function of $\boldsymbol{a}, R, \xi_i$

$$\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 - R^2 - \xi_i$$

is concave with respect to R. Therefore, problem (1) has a non-convex feasible region, so it is not a convex optimization problem.

2.2 Strong Duality and Constraint Qualification

When deriving the Lagrange dual problem of a convex optimization problem, an important property that we hope to have is that the dual optimal value is equal to the primal value. This property, referred to as the strong duality, ensures that we can solve the original problem (called primal) through the dual problem. The following theorem is commonly used to check if the strong duality holds.

Theorem 1 (Boyd and Vandenberghe 2004, Section 5.2.3)**.** *Consider the following primal problem*

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & f_0(\boldsymbol{w}) \\ \textit{subject to} \quad & f_i(\boldsymbol{w}) \leq 0, i = 1, \ldots, m, \\ & h_i(\boldsymbol{w}) = 0, i = 1, \ldots, p, \end{aligned} \tag{3}$$

where $f_0, f_1, \ldots, f_m$ are convex, and $h_1, \ldots, h_p$ are affine.

If this problem either has linear constraints or satisfies some constraint qualifications such as Slater's condition, and the Lagrange dual problem is defined as

$$\max_{\boldsymbol{\alpha} \geq \boldsymbol{0}, \boldsymbol{\nu}} \quad g(\boldsymbol{\alpha}, \boldsymbol{\nu})$$

where

$$g(\boldsymbol{\alpha}, \boldsymbol{\nu}) = \inf_{\boldsymbol{w}} \quad \left(f_0(\boldsymbol{w}) + \sum_{i=1}^{m} \alpha_i f_i(\boldsymbol{w}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{w}) \right),$$

and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_m], \boldsymbol{\nu} = [\nu_1, \ldots, \nu_p]$, then strong duality holds.

In spite of the non-convexity of (1), Tax and Duin (2004) did not check any constraint qualification. Thus they are not in a position to apply the above theorem.

An interesting difference is that it is not necessary to check constraint qualification for SVM. The constraints in (1) are nonlinear, while SVM involves only linear constraints.

2.3 Issues in Deriving the Lagrange Dual Problem

The Lagrangian of (1) is

$$L(\boldsymbol{a}, R, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = R^2 + C \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i \left(R^2 + \xi_i - \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \right) - \sum_{i=1}^{l} \gamma_i \xi_i,$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are Lagrange multipliers. The Lagrange dual problem is

$$\max_{\boldsymbol{\alpha}\geq\mathbf{0},\boldsymbol{\gamma}\geq\mathbf{0}}\left(\inf_{\boldsymbol{a},R,\boldsymbol{\xi}} \quad L(\boldsymbol{a},R,\boldsymbol{\xi},\boldsymbol{\alpha},\boldsymbol{\gamma})\right).$$

To obtain the infimum, Tax and Duin set both the partial derivatives of L with respect to R and $\boldsymbol{a}$ to be zero.

$$\frac{\partial L}{\partial R}=0 \quad \Rightarrow \quad R(1-\sum_{i=1}^{l}\alpha_i)=0, \tag{4}$$

$$\frac{\partial L}{\partial \boldsymbol{a}}=0 \quad \Rightarrow \quad \sum_{i=1}^{l}\alpha_i\phi(\boldsymbol{x}_i)-\boldsymbol{a}\sum_{i=1}^{l}\alpha_i=0. \tag{5}$$

From (4), they then derive

$$\sum_{i=1}^{l}\alpha_i=1, \tag{6}$$

and obtain

$$\boldsymbol{a}=\sum_{i=1}^{l}\alpha_i\phi(\boldsymbol{x}_i)$$

by combining (5) and (6). They finish deriving the dual problem (2) based on the above results. However, (6) is incorrect because when $R=0$ in (4), (6) does not necessarily hold. Thus any further derivations based on (6) are wrong. Finally, problem (2) does not have any feasible solution when $0<C<1/l$. In contrast, the primal problem (1) is feasible for any $C>0$. Thus strong duality is clearly violated.

3 Convexity and the Dual Problem of SVDD

In this section, we carefully address all issues mentioned in Section 2. To apply Theorem 1, we begin with reformulating (1) to a convex problem. We then check its constraint qualification before deriving the dual problem rigorously.

3.1 A Convex Equivalent of Problem (1)

As shown before, (1) is not a convex problem. Nevertheless, by defining

$$\bar{R}=R^2,$$

(1) is equivalent to the following convex problem.

$$\begin{aligned}\min_{\bar{R},\boldsymbol{a},\boldsymbol{\xi}} \quad & \bar{R}+C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & \|\phi(\boldsymbol{x}_i)-\boldsymbol{a}\|^2\leq\bar{R}+\xi_i, i=1,\ldots,l, \\ & \xi_i\geq 0, i=1,\ldots,l, \\ & \bar{R}\geq 0.\end{aligned} \tag{7}$$

A new constraint specifying the non-negativity of $\bar{R}$ is added. Notice that

$$\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 - \bar{R} - \xi_i = \phi(\boldsymbol{x}_i)^T\phi(\boldsymbol{x}_i) - 2\phi(\boldsymbol{x}_i)^T\boldsymbol{a} + \boldsymbol{a}^T\boldsymbol{a} - \bar{R} - \xi_i \tag{8}$$

is linear to both $\bar{R}$ and ξ_i, and is convex with respect to $\boldsymbol{a}$. Therefore, (7) is in a form needed by Theorem 1.

The objective function of (7) is convex rather than strictly convex. Thus, (7) may possess multiple optimal solutions with the same optimal function value. However, the optimal $\boldsymbol{a}$ is unique because (8) is strictly convex to $\boldsymbol{a}$.

3.2 Constraint Qualification

To apply Theorem 1 on problem (7), we must check if (7) satisfies certain constraint qualification. Many types of constraint qualification have been developed in the field of convex optimization. Here we consider Slater's condition.

Theorem 2 (Slater's condition, Boyd and Vandenberghe 2004, Section 5.2.3)**.** *For any function f_i and any set S, define*

$$\textit{dom}(f_i) \equiv \textit{The domain of } f_i,$$

and

$$\textit{relint}(S) \equiv \{\boldsymbol{w} \in S \mid \exists r > 0, B_r(\boldsymbol{w}) \cap \textit{aff}(S) \subset S\},$$

where $B_r(\boldsymbol{w})$ is a ball centered at $\boldsymbol{w}$ with radius r and aff(S) is the affine hull of S. Consider problem (3)*, if there exists $\boldsymbol{w}$ such that*

$$\begin{aligned} \boldsymbol{w} &\in \textit{relint}\left(\cap_{i=0}^{m} \textit{dom}\left(f_i\right)\right), \\ f_i(\boldsymbol{w}) &< 0, i = 1, \ldots, m, \\ h_i(\boldsymbol{w}) &= 0, i = 1, \ldots, p, \end{aligned}$$

then strong duality for (3) *holds.*

For any data $\boldsymbol{x}_i, i = 1, \ldots, l$, we can let $\boldsymbol{a} = \boldsymbol{0}$, then find $\bar{R} > 0$ and $\boldsymbol{\xi} > \boldsymbol{0}$ large enough such that

$$\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 - \bar{R} - \xi_i < 0, i = 1, \ldots, l.$$

Therefore, Slater's condition is satisfied and thus strong duality for (7) holds.

3.3 The Dual Problem of (7)

Recall that a difficulty of deriving (6) from (4) is that R may be zero. Now R^2 is replaced by $\bar{R}$, but a related difficulty is whether the non-negativity constraint $\bar{R} \geq 0$ is active. We handle this difficulty by splitting the derivation to two cases as described in the following theorem.

Theorem 3. *Consider problem* (7)*.*

(a) For any $C > 1/l$, the constraint $\bar{R} \geq 0$ in (7) *is not necessary. That is, without this constraint, any optimal solution still satisfies $\bar{R} \geq 0$.*

(b) For any $C < 1/l$, $\bar{R} = 0$ is uniquely optimal. If $C = 1/l$, then at least one optimal solution has $\bar{R} = 0$.

The proof is in Appendix A. With Theorem 3, we derive the dual problem by considering $C > 1/l$ and $C \leq 1/l$ separately.

Case 1: $C > 1/l$**.**
The Lagrangian of (7) is

$$\begin{aligned} L(\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) &= \bar{R} + C \sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l} \alpha_i \left(\bar{R} + \xi_i - \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \right) - \sum_{i=1}^{l} \gamma_i \xi_i \\ &= \bar{R} \left(1 - \sum_{i=1}^{l} \alpha_i \right) + \sum_{i=1}^{l} \xi_i \left(C - \alpha_i - \gamma_i \right) + \sum_{i=1}^{l} \alpha_i \left(\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \right), \end{aligned} \tag{9}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are Lagrange multipliers. The Lagrange dual problem is

$$\max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\gamma} \geq \mathbf{0}} \left(\inf_{\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}} \quad L(\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \right). \tag{10}$$

Clearly, if $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ satisfies

$$1 - \boldsymbol{e}^T \boldsymbol{\alpha} \neq 0,$$

or

$$C - \alpha_i - \gamma_i \neq 0 \text{ for some } i,$$

then

$$\inf_{\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}} L(\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = -\infty.$$

Such $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ should not be considered because of the maximization over $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ in (10). This leads to the following constraints in the dual problem.

$$1 - \boldsymbol{e}^T \boldsymbol{\alpha} = 0, \tag{11}$$

$$C - \alpha_i - \gamma_i = 0, i = 1, \ldots, l. \tag{12}$$

Substituting (11) and (12) into (10), and taking $\gamma_i \geq 0, \forall i$ into account, the dual problem (10) is reduced to

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \left(\inf_{\boldsymbol{a}} \sum_{i=1}^{l} \alpha_i \left(\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \right) \right) \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C, i = 1, \ldots, l, \\ & \boldsymbol{e}^T \boldsymbol{\alpha} = 1. \end{aligned} \tag{13}$$

Because

$$\sum_{i=1}^{l} \alpha_i \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2$$

is strictly convex with respect to an unbounded variable $\boldsymbol{a}$, the infimum occurs at the point that the derivative is zero.

$$\boldsymbol{a}\sum_{i=1}^{l}\alpha_i = \sum_{i=1}^{l}\alpha_i\phi(\boldsymbol{x}_i). \tag{14}$$

By the constraint (11), (14) is equivalent to

$$\boldsymbol{a} = \frac{\sum_{i=1}^{l}\alpha_i\phi(\boldsymbol{x}_i)}{\boldsymbol{e}^T\boldsymbol{\alpha}} = \sum_{i=1}^{l}\alpha_i\phi(\boldsymbol{x}_i). \tag{15}$$

We then obtain the following dual problem for $C > 1/l$.

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{l}\alpha_i Q_{i,i} - \boldsymbol{\alpha}^T Q\boldsymbol{\alpha} \\ \text{subject to} \quad & 0 \le \alpha_i \le C, i = 1,\ldots,l, \\ & \boldsymbol{e}^T\boldsymbol{\alpha} = 1, \end{aligned} \tag{16}$$

which is the same as (2).

Note that if we do not apply Theorem 3 to remove the constraint $\bar{R} \ge 0$, the Lagrangian has an additional term $-\beta\bar{R}$, where β is the corresponding Lagrange multiplier. Then the constraint (11) becomes

$$1 - \boldsymbol{e}^T\boldsymbol{\alpha} - \beta = 0 \quad \text{and} \quad \beta \ge 0.$$

The situation becomes complicated because we must check if $\boldsymbol{e}^T\boldsymbol{\alpha} > 0$ or not before dividing $\boldsymbol{e}^T\boldsymbol{\alpha}$ from both sides of (14). In Section 4.1, we will see an example that must check the situation of $\boldsymbol{e}^T\boldsymbol{\alpha} = 0$.

We discuss how to obtain the primal optimal solution after solving the dual problem. Clearly, the optimal $\boldsymbol{a}$ can be obtained by (15). Tax and Duin (2004) find $\bar{R}$ by identifying an optimal α_i with $0 < \alpha_i < C$. From the KKT optimality condition, primal and dual optimal solutions satisfy the following slackness conditions.

$$\gamma_i\xi_i = 0 \quad \text{and} \quad \alpha_i\left(\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 - \bar{R} - \xi_i\right) = 0, i = 1,\ldots,l. \tag{17}$$

With (12), an index i with $0 < \alpha_i < C$ satisfies

$$\xi_i = 0 \quad \text{and} \quad \bar{R} = \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2. \tag{18}$$

However, it is possible that all α_i values are bounded, so their method is not always applicable. We show that the optimal $\bar{R}$ can be obtained by the following theorem.

Theorem 4. *Any optimal* $(\bar{R}, \boldsymbol{a})$ *of* (7) *and optimal* $\boldsymbol{\alpha}$ *of* (16) *satisfy*

$$\max_{i:\alpha_i<C}\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \le \bar{R} \le \min_{i:\alpha_i>0}\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2. \tag{19}$$

The proof is in Appendix B. If there exists an index i with $0 < \alpha_i < C$, (19) is reduced to (18) and the optimal $\bar{R}$ is unique. Otherwise, if every α_i is bounded (i.e., 0 or C), then (19) indicates that any $\bar{R}$ in an interval is optimal. Interestingly, (19) is similar to the inequality for the bias term b in SVM problems; see, for example, Fan et al. (2005). For the practical implementation we may adopt the following setting in LIBSVM (Chang and Lin, 2011) to calculate $\bar{R}$.

(a) If some indices satisfy $0 < \alpha_i < C$, then we calculate the average of $\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2$ over all such i. The reason is that each single $\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2$ may be inaccurate because of numerical errors.

(b) If all α_i are bounded, then we choose $\bar{R}$ to be the middle point of the interval in (19).

Finally, the optimal $\boldsymbol{\xi}$ can be computed by

$$\xi_i = \max\left(\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 - \bar{R}, 0\right), i = 1, \ldots, l. \tag{20}$$

Another interesting property is that when C is large, all models of SVDD are the same. We state this result in the following theorem.

Theorem 5. *For any $C > 1$, problem (7) is equivalent to following problem.*

$$\begin{aligned} \min_{\bar{R}, \boldsymbol{a}} \quad & \bar{R} \\ \textit{subject to} \quad & \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \leq \bar{R}. \end{aligned} \tag{21}$$

The proof is in Appendix C. The relation between (7) and (21) is similar to that between soft-margin and hard-margin SVM, where the latter uses neither C nor $\boldsymbol{\xi}$ because of assuming that data are separable. For SVM, it is know that if data are separable, there is a $\bar{C}$ such that for all $C > \bar{C}$, the solution is the same as without having the loss term; see, for example, Lin (2001). This $\bar{C}$ is problem dependent, but for SVDD, we have shown that $\bar{C}$ is one.

Case 2: $C \leq 1/l$.
By Theorem 3, we can remove the variable $\bar{R}$ from problem (7). The minimum must occur when

$$\xi_i = \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \geq 0.$$

Thus, problem (7) can be reduced to

$$\min_{\boldsymbol{a}} \quad \sum_{i=1}^{l} \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2. \tag{22}$$

This problem is strictly convex to $\boldsymbol{a}$, so setting the gradient to be zero leads to

$$\boldsymbol{a} = \frac{\sum_{i=1}^{l} \phi(\boldsymbol{x}_i)}{l}. \tag{23}$$

Therefore, when $C \leq 1/l$, the optimal solution is independent of C. Further, the optimization problem has a closed-form solution in (23).

3.4 Implementation Issues

The dual problem (16) is very similar to the SVM dual problem. They both have a quadratic objective function, one linear constraint, and l bounded constraints. Therefore, existing optimization methods such as decomposition methods (e.g., Platt 1998; Joachims 1998;

Fan et al. 2005) can be easily applied. We also note that (16) is related to the dual problem of one-class SVM (Schölkopf et al., 2001), which is another method for outlier detection.

In the prediction stage, for any test instance $\boldsymbol{x}$, we must check the value

$$\|\phi(\boldsymbol{x}) - \boldsymbol{a}\|^2 - \bar{R}.$$

If it is positive, then $\boldsymbol{x}$ is considered as an outlier. If a kernel is used and $C > 1/l$, then from (15), the calculation is conducted by

$$\|\phi(\boldsymbol{x}) - \boldsymbol{a}\|^2 - \bar{R} = K(\boldsymbol{x}, \boldsymbol{x}) - 2 \sum_{i:\alpha_i > 0} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} - \bar{R},$$

where $K(\cdot, \cdot)$ is the kernel function. The $\boldsymbol{\alpha}^T Q \boldsymbol{\alpha} - \bar{R}$ term is expensive to calculate, although it is independent from test instances. A trick is to store this constant after solving the dual problem. Note that we can rewrite (19) in a way related to $\boldsymbol{\alpha}^T Q \boldsymbol{\alpha} - \bar{R}$.

$$\max_{i:\alpha_i < C} (Q_{i,i} - 2(Q\boldsymbol{\alpha})_i) \le \bar{R} - \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} \le \min_{i:\alpha_i > 0} (Q_{i,i} - 2(Q\boldsymbol{\alpha})_i).$$

Then $\bar{R} - \boldsymbol{\alpha}^T Q \boldsymbol{\alpha}$ is the Lagrange multiplier of the dual problem (16) with respect to the equality constraint $\boldsymbol{e}^T \boldsymbol{\alpha} = 1$. In our implementation based on LIBSVM, its solver handily provides this multiplier after solving the dual problem.

4 Extensions

In this section, we discuss some extensions of SVDD.

4.1 L2-Loss SVDD

One alternative of SVDD is to adopt L2 loss in the objective function. We will show that L2-loss SVDD has several differences from the L1-loss one.

The optimization problem of L2-loss SVDD is

$$\begin{aligned} \min_{\bar{R}, \boldsymbol{a}, \boldsymbol{\xi}} \quad & \bar{R} + C \sum_{i=1}^{l} \xi_i^2 \\ \text{subject to} \quad & \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \le \bar{R} + \xi_i, i = 1, \ldots, l, \\ & \bar{R} \ge 0. \end{aligned} \tag{24}$$

Note that the constraint $\xi_i \ge 0, \forall i$ appeared in (7) is not necessary for L2-loss SVDD, because if at an optimum, $\xi_i < 0$ for some i, we can then replace ξ_i with 0 so that

$$\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \le \bar{R} + \xi_i < \bar{R} + 0.$$

The constraints are still satisfied, but the objective value is smaller. This contradicts the assumption that ξ_i is optimal.

Similar to the L1-loss case, because of using $\bar{R}$ rather than R^2, (24) is a convex optimization problem. Furthermore, Slater's condition holds, and so does the strong duality.

To derive the dual problem, the Lagrangian of (24) is

$$L(\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \beta) = \bar{R} + C\sum_{i=1}^{l} \xi_i^2 - \sum_{i=1}^{l} \alpha_i \left(\bar{R} + \xi_i - \|\phi\left(\boldsymbol{x}_i\right) - \boldsymbol{a}\|^2\right) - \beta\bar{R}, \tag{25}$$

where $\boldsymbol{\alpha}$ and β are Lagrange multipliers. The Lagrange dual problem is

$$\max_{\boldsymbol{\alpha}\geq\mathbf{0},\beta\geq 0} \left(\inf_{\boldsymbol{a},\bar{R},\boldsymbol{\xi}} \quad L\left(\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \beta\right)\right). \tag{26}$$

Clearly, if

$$1 - \boldsymbol{e}^T\boldsymbol{\alpha} - \beta \neq 0,$$

then

$$\inf_{\boldsymbol{a},\bar{R},\boldsymbol{\xi}} L(\boldsymbol{a}, \bar{R}, \boldsymbol{\xi}, \boldsymbol{\alpha}, \beta) = -\infty.$$

Thus we have the following constraint in the dual problem.

$$1 - \boldsymbol{e}^T\boldsymbol{\alpha} - \beta = 0. \tag{27}$$

In addition, L is strictly convex to $\xi_i, \forall i$, so we have

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Rightarrow \quad \xi_i = \frac{\alpha_i}{2C}, i = 1, \ldots, l. \tag{28}$$

Substituting (27) and (28) into (26), the dual problem of (24) is

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \left(\inf_{\boldsymbol{a}} \sum_{i=1}^{l} \alpha_i \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 - \sum_{i=1}^{l} \frac{\alpha_i^2}{4C} \right) \\ \text{subject to} \quad & 0 \leq \alpha_i \leq \infty, i = 1, \ldots, l, \\ & \boldsymbol{e}^T\boldsymbol{\alpha} \leq 1. \end{aligned} \tag{29}$$

Similar to the derivation from (13) to (14), the infimum occurs when

$$\boldsymbol{a}\sum_{i=1}^{l} \alpha_i = \sum_{i=1}^{l} \alpha_i \phi(\boldsymbol{x}_i). \tag{30}$$

However, because we now have (27) rather than (6), we cannot divide $\sum_{i=1}^{l} \alpha_i$ from both sides of (30) as in (15). Instead, we must check if $\sum_{i=1}^{l} \alpha_i = 0$ can happen or not.

When $\sum_{i=1}^{l} \alpha_i = 0$, by the constraints of (29), we have $\alpha_i = 0, \forall i$. If it is an optimal solution of (29), then the objective value is zero, and by the strong duality, so is the primal optimal value. Because both $\bar{R}$ and ξ_i^2 are non-negative, this situation is possible only when

$$\bar{R} = 0, \boldsymbol{\xi} = \mathbf{0},$$

which indicates

$$\phi(\boldsymbol{x}_1) = \phi(\boldsymbol{x}_2) = \cdots = \phi(\boldsymbol{x}_l). \tag{31}$$

We can rule out this situation before solving (24).

If (31) does not occur, (29) is equivalent to the problem of adding the constraint $\sum_{i=1}^{l} \alpha_i \neq 0$. Then from (30),

$$\boldsymbol{a} = \frac{\sum_{i=1}^{l} \alpha_i \phi(\boldsymbol{x}_i)}{\boldsymbol{e}^T \boldsymbol{\alpha}}. \tag{32}$$

Finally, the dual problem is

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{l} \alpha_i Q_{i,i} - \frac{\boldsymbol{\alpha}^T Q \boldsymbol{\alpha}}{\boldsymbol{e}^T \boldsymbol{\alpha}} - \sum_{i=1}^{l} \frac{\alpha_i^2}{4C} \\ \text{subject to} \quad & 0 \leq \alpha_i \leq \infty, i = 1, \ldots, l, \\ & \boldsymbol{e}^T \boldsymbol{\alpha} \leq 1. \end{aligned} \tag{33}$$

Because the loss term $C \sum_{i=1}^{l} \xi_i^2$ is now strictly convex, we are able to prove the uniqueness of the optimum.

Theorem 6. *The optimal solutions of* (24) *and* (33) *are both unique.*

The proof is in Appendix D. We can then further simplify (33) by the following theorem.

Theorem 7. *There exists* $C^* \geq 0$ *such that*

(a) If $C > C^*$*, the optimal* $\boldsymbol{\alpha}$ *and* $\bar{R}$ *satisfy* $\boldsymbol{e}^T \boldsymbol{\alpha} = 1$ *and* $\bar{R} > 0$*.*

(b) If $C \leq C^*$*, the optimal* $\bar{R} = 0$*.*

The proof is in Appendix E. Clearly, C^* plays the same role as $1/l$ in Theorem 3 for L1-loss SVDD. The main difference, which will be discussed in detail, is that C^* is problem dependent.

Following Theorem 7, we discuss the two situations $C > C^*$ and $C \leq C^*$ in detail. First, when $C > C^*$, (29) is equivalent to

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^{l} \alpha_i Q_{i,i} - \boldsymbol{\alpha}^T Q \boldsymbol{\alpha} - \sum_{i=1}^{l} \frac{\alpha_i^2}{4C} \\ \text{subject to} \quad & 0 \leq \alpha_i \leq \infty, i = 1, \ldots, l, \\ & \boldsymbol{e}^T \boldsymbol{\alpha} = 1, \end{aligned} \tag{34}$$

which is very similar to (16). Some minor differences are that (34) has an additional $\sum_{i=1}^{l} (\alpha_i^2/4C)$ term and $\boldsymbol{\alpha}$ is unbounded.

For the situation that $C \leq C^*$, by the same explanation to derive (22), (24) can be reduced to

$$\begin{aligned} \min_{\boldsymbol{a}, \boldsymbol{\xi}} \quad & C \sum_{i=1}^{l} \xi_i^2 \\ \text{subject to} \quad & \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \leq \xi_i, i = 1, \ldots, l. \end{aligned} \tag{35}$$

Note that C in (35) is not needed. Therefore, similar to Case 2 of L1 loss, problem (35) is independent of C. However, while (22) has a simple analytic solution, (35) does not. If we consider the dual problem, then the following result can be proved.

Theorem 8. *For any $C \le C^*$, the dual optimal solution of* (24) *is a linear function of C. That is,*

$$\boldsymbol{\alpha} = \frac{C}{C^*} \cdot \boldsymbol{\alpha}^*, \tag{36}$$

where $\boldsymbol{\alpha}^*$ and $\boldsymbol{\alpha}$ are the optimal solutions for (33) *with parameters C^* and C, respectively.*

The proof is in Appendix F. Note that the objective function of (33) can be written as

$$C\left(\sum_{i=1}^{l}(\frac{\alpha_i}{C})Q_{i,i} - \frac{(\frac{\boldsymbol{\alpha}}{C})^T Q(\frac{\boldsymbol{\alpha}}{C})}{\boldsymbol{e}^T(\frac{\boldsymbol{\alpha}}{C})} - \frac{1}{4}\sum_{i=1}^{l}(\frac{\alpha_i}{C})^2\right).$$

Without considering the coefficient C, from Theorem 8, the optimal value in the parentheses is a constant. It corresponds to the optimal $\sum_{i=1}^{l}\xi_i^2$ value in the primal problem (35).

The remaining task is to compute the value of C^*. Unfortunately, it is problem dependent. We show this result by considering the following two examples. The first problem consists of two instances $\boldsymbol{x}_1 = 1$ and $\boldsymbol{x}_2 = -1$. Clearly, the optimal $\boldsymbol{a}$ for any C is $\boldsymbol{a} = 0$. From (20), the primal problem is then equivalent to

$$\min_{\bar{R}\ge 0} \quad \bar{R} + 2C(1-\bar{R})^2.$$

The optimal $\bar{R}$ is

$$\bar{R} = \begin{cases} \frac{4C-1}{4C}, & \text{if } 4C-1 \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

Thus $C^* = 1/4$. For the other example, we consider $\boldsymbol{x}_1 = 0.1$ and $\boldsymbol{x}_2 = -0.1$. By the same derivation, $C^* = 5/2$.

The problem-dependent C^* makes (24) more difficult to solve. In contrast, $C^* = 1/l$ is a constant for L1-loss SVDD. Thus L2-loss SVDD is not recommended.

4.2 Smallest Circle Encompassing the Data

The radius of the smallest circle encompassing all training instances is useful for evaluating an upper bound of leave-one-out error for SVMs (Vapnik and Chapelle, 2000; Chung et al., 2003). It can be computed by a simplified form of (7) without considering $\boldsymbol{\xi}$.

$$\begin{aligned} \min_{\bar{R},\boldsymbol{a}} \quad & \bar{R} \\ \text{subject to} \quad & \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \le \bar{R}. \end{aligned}$$

Note that this is identical to (21). Past works have derived the dual problem of (21). As expected, it is (16) without the constraint $\alpha_i \le C, \forall i$. A practical issue is that for applying an optimization procedure for (16) to solve the dual problem here, replacing C with ∞ may cause numerical issues. We address this issue by applying Theorem 5. That is, to solve (21), all we have to do is to solve (7) with any $C > 1$.

5 Conclusions

In this paper, we pointed out several problems in the existing derivation of SVDD. We make thorough corrections by rigorously following the theory of convex optimization. The extension of using L2-loss is also studied, but we conclude that it is more complicated to use than the standard L1-loss SVDD. Based on this work, we have released an extension of LIBSVM for SVDD at LIBSVM Tools.[2]

A Proof of Theorem 3

For any $C > 1/l$, assume $(\bar{R}, \boldsymbol{a}, \boldsymbol{\xi})$ is an optimum with $\bar{R} < 0$. We consider a new point $(0, \boldsymbol{a}, \boldsymbol{\xi} + \bar{R}\boldsymbol{e})$, where $\boldsymbol{e}$ is the vector of all ones. This point is feasible because

$$0 \le \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \le \bar{R} + \xi_i = 0 + (\xi_i + \bar{R})$$

and therefore

$$\xi_i + \bar{R} \ge 0.$$

Because $C > 1/l$ and $\bar{R} < 0$, the new objective function satisfies

$$0 + C\sum_{i=1}^{l}(\xi_i + \bar{R}) = C\sum_{i=1}^{l}\xi_i + lC\bar{R} < C\sum_{i=1}^{l}\xi_i + \bar{R},$$

a contradiction to the assumption that $(\bar{R}, \boldsymbol{a}, \xi)$ is optimal.

For any $C \le 1/l$, assume $(\bar{R}, \boldsymbol{a}, \boldsymbol{\xi})$ is an optimum with $\bar{R} > 0$. We consider a new point $(0, \boldsymbol{a}, \boldsymbol{\xi} + \bar{R}\boldsymbol{e})$, where $\boldsymbol{e}$ is the vector of all ones. This point is feasible because

$$0 \le \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \le \bar{R} + \xi_i = 0 + (\xi_i + \bar{R})$$

and

$$\xi_i + \bar{R} \ge 0.$$

Because $C \le 1/l$ and $\bar{R} > 0$, the new objective function satisfies

$$0 + C\sum_{i=1}^{l}(\xi_i + \bar{R}) = C\sum_{i=1}^{l}\xi_i + lC\bar{R} \le C\sum_{i=1}^{l}\xi_i + \bar{R}. \tag{37}$$

Along with the constraint $\bar{R} \ge 0$, $\bar{R} = 0$ is optimal when $C \le 1/l$. Furthermore, when $C < 1/l$, (37) becomes a strict inequality. This contradicts the assumption that $(\bar{R}, \boldsymbol{a}, \xi)$ is optimal, so the optimal $\bar{R}$ must be zero.

B Proof of Theorem 4

From the KKT conditions (17) and (12), at an optimum we have for all i,

$$\bar{R} \ge \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2, \text{ if } \alpha_i < C,$$
$$\bar{R} \le \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2, \text{ if } \alpha_i > 0.$$

The inequality (19) immediately follows.

[2] http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/#libsvm_for_svdd_and_finding_the_smallest_sphere_containing_all_data.

C Proof of Theorem 5

From (11), (12) and the constraint $\alpha_i \geq 0$, $\forall i$, if $C > 1$, then $\gamma_i > 0$ and the KKT optimality condition $\gamma_i \xi_i = 0$ in (17) implies that $\xi_i = 0$. Therefore, the $C \sum_{i=1}^{l} \xi_i$ term can be removed from the objective function of (7). The $\bar{R} \geq 0$ constraint is not needed because without $\boldsymbol{\xi}$, $\bar{R} \geq \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2$ has implicitly guaranteed the non-negativity of $\bar{R}$. Therefore, if $C > 1$, problems (7) and (21) are equivalent.

D Proof of Theorem 6

Similar to L1-loss SVDD, the optimal $\boldsymbol{a}$ is unique because the function of the inequality constraint is strictly convex to $\boldsymbol{a}$. Now because of using the strictly convex loss term $C \sum_{i=1}^{l} \xi_i^2$, the optimal $\boldsymbol{\xi}$ is unique. Then

$$\bar{R} = \text{Primal optimal value} - C \sum_{i=1}^{l} \xi_i^2$$

is unique because a convex programming problem has a unique optimal objective value. Finally, from the condition (28), the dual optimal $\boldsymbol{\alpha}$ is unique.

E Proof of Theorem 7

First we need the following lemma for the monotonicity of $\bar{R}$ with respect to the value of C.

Lemma 1. *Consider problem* (24)*. The optimal $\bar{R}$ is an increasing function with respect to C.*

Proof. For $C_2 > C_1 > 0$, assume $(\bar{R}_{C_2}, \boldsymbol{a}_{C_2}, \boldsymbol{\xi}_{C_2})$ and $(\bar{R}_{C_1}, \boldsymbol{a}_{C_1}, \boldsymbol{\xi}_{C_1})$ are optimal solutions of (24) with $C = C_2$ and $C = C_1$, respectively. Thus

$$\bar{R}_{C_1} + C_1 \sum_{i=1}^{l} (\boldsymbol{\xi}_{C_1})_i^2 \leq \bar{R}_{C_2} + C_1 \sum_{i=1}^{l} (\boldsymbol{\xi}_{C_2})_i^2,$$
$$\bar{R}_{C_2} + C_2 \sum_{i=1}^{l} (\boldsymbol{\xi}_{C_2})_i^2 \leq \bar{R}_{C_1} + C_2 \sum_{i=1}^{l} (\boldsymbol{\xi}_{C_1})_i^2.$$

We then obtain

$$C_1 \left(\sum_{i=1}^{l} (\boldsymbol{\xi}_{C_1})_i^2 - \sum_{i=1}^{l} (\boldsymbol{\xi}_{C_2})_i^2 \right) \leq \bar{R}_{C_2} - \bar{R}_{C_1} \leq C_2 \left(\sum_{i=1}^{l} (\boldsymbol{\xi}_{C_1})_i^2 - \sum_{i=1}^{l} (\boldsymbol{\xi}_{C_2})_i^2 \right), \tag{38}$$

which implies

$$\sum_{i=1}^{l} (\boldsymbol{\xi}_{C_1})_i^2 - \sum_{i=1}^{l} (\boldsymbol{\xi}_{C_2})_i^2 \geq 0 \tag{39}$$

because $C_2 > C_1$. Combining (38) and (39), we then have

$$\bar{R}_{C_2} \geq \bar{R}_{C_1}$$

as desired. □

Now we can proceed on the main proof. Let

$$\Delta \equiv \min_{\boldsymbol{a}} \sum_{i=1}^{l} \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 > 0.$$

Note that we have ruled out the special case $\Delta = 0$, which occurs only when $\phi(\boldsymbol{x}_1) = \phi(\boldsymbol{x}_2) = \cdots = \phi(\boldsymbol{x}_l)$. Assume $(\bar{R}_C, \boldsymbol{a}_C, \boldsymbol{\xi}_C)$ is the optimal solution of (24) corresponding to C. Then

$$\sum_{i=1}^{l} \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}_C\|^2 \geq \Delta > 0,$$

and there exists at least one j such that

$$\|\phi(\boldsymbol{x}_j) - \boldsymbol{a}_C\|^2 \geq \frac{\Delta}{l} > 0. \tag{40}$$

From (40), (28), and the constraints of (24),

$$\bar{R}_C + \frac{(\boldsymbol{\alpha}_C)_j}{2C} = \bar{R}_C + (\boldsymbol{\xi}_C)_j \geq \|\phi(\boldsymbol{x}_j) - \boldsymbol{a}_C\|^2 > 0, \forall C > 0. \tag{41}$$

We claim that if

$$2C \cdot \frac{\Delta}{l} > 1, \tag{42}$$

then $\bar{R}_C > 0$. Otherwise, from (41) and (40), $\bar{R}_C = 0$ implies

$$\frac{(\boldsymbol{\alpha}_C)_j}{2C} = (\boldsymbol{\xi}_C)_j \geq \|\phi(\boldsymbol{x}_j) - \boldsymbol{a}_C\|^2 \geq \frac{\Delta}{l}.$$

With (42),

$$(\boldsymbol{\alpha}_C)_j > 1,$$

but this violates the constraints of (29). Therefore, $\bar{R}_C > 0$ if (42) holds. From the KKT optimality condition that

$$\beta_C \bar{R}_C = 0,$$

we obtain

$$\beta_C = 0.$$

Thus from (27),

$$0 = 1 - \boldsymbol{e}^T \boldsymbol{\alpha}_C - \beta_C = 1 - \boldsymbol{e}^T \boldsymbol{\alpha}_C. \tag{43}$$

By Lemma 1, (43) and the fact that $C > 0$, there is an infimum $C^* \geq 0$ such that

$$\bar{R}_C > 0 \text{ and } \boldsymbol{e}^T \boldsymbol{\alpha}_C = 1, \forall C > C^*.$$

If $C^* > 0$, we claim that for any C with $0 < C < C^*$, $\bar{R}_C = 0$. Otherwise, if for one $\bar{C} < C^*$ we have $\bar{R}_{\bar{C}} > 0$, then Lemma 1 implies that $\bar{R}_C > 0, \forall C \geq \bar{C}$, a violation to the definition of C^*.

The situation at $C = C^*$ is more complicated. Because $\bar{R}_C = 0, \forall 0 < C < C^*$, (24) is reduced to

$$\begin{aligned} \min_{\boldsymbol{a},\boldsymbol{\xi}} \quad & \sum_{i=1}^{l} \xi_i^2 \\ \text{subject to} \quad & \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2 \leq \xi_i, i = 1, \ldots, l. \end{aligned} \tag{44}$$

We denote the unique optimal solution as $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\xi}})$. Let $(\bar{R}^*, \boldsymbol{a}^*, \boldsymbol{\xi}^*)$ and $(0, \boldsymbol{a}_C, \boldsymbol{\xi}_c)$ be optimal solutions at C^* and $0 < C < C^*$. Because (44) is independent of C, we have

$$\boldsymbol{a}_C = \hat{\boldsymbol{a}}, \boldsymbol{\xi}_C = \hat{\boldsymbol{\xi}}, \forall 0 < C < C^*.$$

Then

$$\begin{aligned} \bar{R}^* + C^* \sum_{i=1}^{l} (\xi_i^*)^2 &\leq 0 + C^* \sum_{i=1}^{l} \hat{\xi}_i^2, \\ 0 + C \sum_{i=1}^{l} \hat{\xi}_i^2 &\leq \bar{R}^* + C \sum_{i=1}^{l} (\xi_i^*)^2. \end{aligned}$$

Let $C \to C^*$, we have

$$C^* \sum_{i=1}^{l} \hat{\xi}_i^2 \leq \bar{R}^* + C^* \sum_{i=1}^{l} (\xi_i^*)^2 \leq C^* \sum_{i=1}^{l} \hat{\xi}_i^2.$$

Thus

$$C^* \sum_{i=1}^{l} \hat{\xi}_i^2 = \bar{R}^* + \sum_{i=1}^{l} (\xi_i^*)^2.$$

That is, $(0, \hat{\boldsymbol{a}}, \hat{\boldsymbol{\xi}})$ is an optimal solution at C^*. By Theorem 6, the optimal solution of (24) is unique at any C^*. We thus have $\bar{R}^* = 0$.

F Proof of Theorem 8

The KKT condition gives

$$\alpha_i (\frac{\alpha_i}{2C} - \|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2) = 0. \tag{45}$$

From Theorem 7, for any $C \leq C^*$, we are solving the following problem.

$$\min_{\boldsymbol{a}} \quad \sum_{i=1}^{l} (\|\phi(\boldsymbol{x}_i) - \boldsymbol{a}\|^2)^2.$$

Therefore, the optimal $\boldsymbol{a}$ is a constant. Thus,

$$\frac{\alpha_i}{2C} = \frac{\alpha_i^*}{2C^*},$$

and the proof is complete.

References

Amit Banerjee, Philippe Burlina, and Reuven Meth. Fast hyperspectral anomaly detection via SVDD. In *Proceedings of IEEE International Conference on Image Processing*, pages IV–101. IEEE, 2007.

Bernhard E. Boser, Isabelle Guyon, and Vladimir Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. ACM Press, 1992.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Kai-Min Chung, Wei-Chun Kao, Chia-Liang Sun, Li-Lun Wang, and Chih-Jen Lin. Radius margin bounds for support vector machines with the RBF kernel. *Neural Computation*, 15:2643–2681, 2003.

Corina Cortes and Vladimir Vapnik. Support-vector network. *Machine Learning*, 20:273–297, 1995.

Rong-En Fan, Pai-Hsuen Chen, and Chih-Jen Lin. Working set selection using second order information for training SVM. *Journal of Machine Learning Research*, 6:1889–1918, 2005. URL `http://www.csie.ntu.edu.tw/~cjlin/papers/quadworkset.pdf`.

Thorsten Joachims. Making large-scale SVM learning practical. In Bernhard Schölkopf, Christopher J. C. Burges, and Alexander J. Smola, editors, *Advances in Kernel Methods – Support Vector Learning*, pages 169–184, Cambridge, MA, 1998. MIT Press.

Sang-Woong Lee, Jooyoung Park, and Seong-Whan Lee. Low resolution face recognition based on support vector data description. *Pattern Recognition*, 39(9):1809–1812, 2006.

Chih-Jen Lin. Formulations of support vector machines: a note from an optimization point of view. *Neural Computation*, 13(2):307–317, 2001.

Jooyoung Park, Daesung Kang, Jongho Kim, James T. Kwok, and Ivor W. Tsang. SVDD-based pattern denoising. *Neural Computation*, 19(7):1919–1938, 2007.

John C. Platt. Fast training of support vector machines using sequential minimal optimization. In Bernhard Schölkopf, Christopher J. C. Burges, and Alexander J. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, Cambridge, MA, 1998. MIT Press.

Bernhard Schölkopf, John C. Platt, John Shawe-Taylor, Alexander J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

David M. J. Tax and Robert P. W. Duin. Uniform object generation for optimizing one-class classifiers. *Journal of Machine Learning Research*, 2:155–173, 2002.

David M. J. Tax and Robert P. W. Duin. Support vector data description. *Machine Learning*, 54(1):45–66, 2004.

Vladimir Vapnik and Olivier Chapelle. Bounds on error expectation for support vector machines. *Neural Computation*, 12(9):2013–2036, 2000. URL `citeseer.nj.nec.com/vapnik99bounds.html`.