ISSN 1342-2804

Research Reports on Mathematical and Computing Sciences

A Modified Algorithm for Nonconvex Support Vector Classification

Akiko Takeda

August 2007, B–443

Department of
Mathematical and
Computing Sciences
Tokyo Institute of Technology

SERIES B: Operations Research

B-443 A Modified Algorithm for Nonconvex Support Vector Classification

Akiko Takeda†, August 2007

Abstract. As an extension of ν-support vector machine for classification (SVC), Extended ν-SVC was developed by Perez-Cruz et al. Their numerical experiments confirm the validity of Extended ν-SVC, but we need to solve a nonconvex QP problem for Extended ν-SVC. In the paper, we propose a modification for the existing algorithm of Extended ν-SVC, which makes possible to analyze the finite convergence and local optimality of the algorithm. The modification is theoretically rather than practically important, but experimental results also show that the modification causes the algorithm to finish faster.

Key words.

support vector machines, conditional value-at-risk, nonconvex quadratic programming, extended ν-SVM algorithm

† Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1-W8-29 Oh-Okayama, Meguro-ku, Tokyo 152-8552 Japan. *takeda@is.titech.ac.jp*

1 Introduction

The ν-support vector machines for classification (ν-SVC) proposed by Schölkopf et al. [8] has a meaningful parameter ν, which roughly represents the fraction of support vectors. The parameter ν cannot always take all possible values between 0 and 1. Therefore, to extend the permissible range of ν up to $[0, 1)$, Extended ν-SVC was developed by Perez-Cruz et al. [6]. While the modification to extend the range up to 1 is relatively simple and understandable, the modification to extend the range down to zero is far more complex. However, due to the modification for extending the range down to zero, numerical experiments of [6] confirm the validity of Extended ν-SVC.

On the other hand, Gotoh and Takeda [2] proposed a classification method by introducing a risk measure known as the conditional value-at-risk (β-CVaR) [7]. The CVaR minimization problem for the margin distribution is essentially the same as the formulation of Extended ν-SVC with $\nu = 1-\beta$. Takeda [10] further investigated the classification method based on CVaR minimization as β-SVC, and discussed theoretical aspects, mainly generalization performance, of β-SVC.

When the parameter $\beta = 1 - \nu$ is set to be less than a threshold $\bar{\beta}$, problems of β-SVC and Extended ν-SVC result in a convex quadratic programming (QP) problem of ν-SVC. For $\beta = 1 - \nu \geq \bar{\beta}$, however, we need to solve a nonconvex QP problem for β-SVC and Extended ν-SVC. If a kernel matrix is not full-rank as in linear or polynomial kernel, nonconvex β-SVC and Extended ν-SVC has a possibility to find a good classifier.

Local search algorithms were proposed in [2, 6] for β-SVC and Extended ν-SVC. The algorithm of [6] has the advantage of easy implementation with a standard linear programming (LP) solver, compared to that of [2]. In the paper, we propose a modification for the algorithm of [6], which makes possible to analyze the finite convergence and local optimality of the algorithm.

2 Nonconvex Problem based on CVaR Minimization

As one of soft margin approaches in the nonlinearly separable case, Gotoh and Takeda [2] proposed *conditional geometric score* (CGS) classification method using given training data $(\boldsymbol{x}_i, y_i) \in \chi \times \{\pm 1\}$, $i \in M := \{1, ..., m\}$. The examples $\boldsymbol{x}_i$, $i \in M$, are taken from some nonempty set $\chi \subset \mathbb{R}^n$ and the labels y_i, $i \in M$, are from binary values: -1 or 1. The CGS method regards the signed distance function defined by

$$f(\boldsymbol{w}, b; \boldsymbol{x}, y) = -\frac{y(\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b)}{\|\boldsymbol{w}\|} \tag{1}$$

as a cost function. Let $\alpha_\beta(\boldsymbol{w}, b)$ be β-percentile, which is also called the *value-at-risk* (β-VaR), for the distribution of $f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i)$, $i \in M$. β-VaR is typically used by security houses or investment banks to measure the market risk of their asset portfolios. The mean of the β-tail distribution of $f(\boldsymbol{w}, b; \boldsymbol{x}, y)$ is known as *conditional value-at-risk* (β-CVaR) [7], and denoted by $\phi_\beta(\boldsymbol{w}, b)$ (see Fig. 1).

The CGS problem [2] minimizes β-CVaR for $f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i)$, $i \in M$ of (1) as

$$\min_{\boldsymbol{w}, b, \alpha} \alpha + \frac{1}{(1-\beta)m} \sum_{i \in M} [f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i) - \alpha]^+,$$

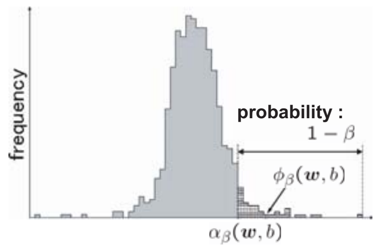

Fig. 1. Illustration of the β-tail expectation of f

where $[X]^+ := \max\{X, 0\}$. Rockafellar & Uryasev [7] have shown that the solution α^* is almost equal to β-VaR, $\alpha(\boldsymbol{w}^*, b^*)$, and the optimal value is equal to β-CVaR, $\phi(\boldsymbol{w}^*, b^*)$. In the problem, a threshold is set on β-VaR, and expected excess of $f(\boldsymbol{w}^*, b^*; \boldsymbol{x}, y)$ over β-VaR, which corresponds to β-CVaR, is regarded as the loss or risk.

The CVaR minimization problem is rewritten as

$$\begin{array}{ll} \min\limits_{\boldsymbol{w},b,\alpha,\boldsymbol{z}} & \alpha + \dfrac{1}{(1-\beta)m} \displaystyle\sum_{i \in M} z_i \\ \text{subject to} & z_i + y_i(\,\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b\,) + \alpha \geq 0, \quad i \in M, \\ & z_i \geq 0, \quad i \in M, \\ & \boldsymbol{w}^\top \boldsymbol{w} = 1. \end{array} \tag{2}$$

Let the numbers of data with positive and negative labels be m_+ and m_-, respectively, and suppose that m_+ and m_- are positive. Then Problem (2) is proved to have an optimal solution when the parameter β is chosen so that $\beta_{min} := 1 - \frac{2\min\{m_+, m_-\}}{m} \leq \beta < 1$.

Problem (2) is essentially the same as the formulation of Extended ν-SVC proposed by Perez-Cruz et al [6]. Indeed, Extended ν-SVC with $\nu = 1 - \beta$ and β-SVC generate the same classifier.

The optimal value of (2) is nondecreasing with respect to β (see [2, 10]). When the training data are linearly separable, there exists $(\boldsymbol{w}, b)$ such that $f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i) < 0$ holds for all $i \in M$. Then, at an optimal solution, β-CVaR and β-VaR obviously take negative values. In the nonlinearly separable case, however, β-CVaR and β-VaR possibly become positive especially for large β. Since the optimal value of (2), $\phi_\beta(\boldsymbol{w}^*_\beta, b^*_\beta)$, is nondecreasing with respect to β, there may exist $\bar\beta$ which induces 0 optimal value in (2), that is, $\phi_{\bar\beta}(\boldsymbol{w}^*_{\bar\beta}, b^*_{\bar\beta}) = 0$, though it is difficult to find such $\bar\beta$ exactly. With the use of $\bar\beta$, Problem (2) is classified into two cases: the convex case where the optimal value of (2) is negative for $\beta \in [\beta_{min}, \bar\beta)$, and the nonconvex case where its optimal value is nonnegative for $\beta \in [\bar\beta, 1)$.

Problem (2) is not obviously convex. But when β is in the range $[\beta_{min}, \bar\beta)$, it can be transformed into a convex problem:

$$\begin{array}{ll} \min\limits_{\boldsymbol{w},b,\alpha,\boldsymbol{z}} & \alpha + \dfrac{1}{(1-\beta)m} \displaystyle\sum_{i \in M} z_i \\ \text{subject to} & z_i + y_i(\,\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b\,) + \alpha \geq 0, \quad i \in M, \\ & z_i \geq 0, \quad i \in M, \\ & \boldsymbol{w}^\top \boldsymbol{w} \leq 1. \end{array} \tag{3}$$

The nonconvex constraint $\boldsymbol{w}^\top \boldsymbol{w} = 1$ of (2) is relaxed into a convex constraint $\boldsymbol{w}^\top \boldsymbol{w} \leq 1$, since at optimality $\boldsymbol{w}^\top \boldsymbol{w} = 1$ is attained as far as the optimal value of (2) is negative. It is easy to incorporate a kernel function into β-SVC (3). Taking dual for Problem (3) and incorporating kernels $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ to dot products $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle$ in the objective function, we have a QP problem which is exactly ν-SVC with parameter $\nu = 1 - \beta$.

On the other hand, if solving a convex problem (3) for β-SVC (2) with $\beta \in (\bar{\beta}, 1)$, one has the meaningless optimal solution $\boldsymbol{w} = \boldsymbol{0}$ and $b = 0$. Therefore, in this case, the nonconvex constraint $\boldsymbol{w}^\top \boldsymbol{w} = 1$ is essential. As a nonlinear kernel-based variant of β-SVC, [10] proposed the problem:

$$\begin{array}{ll} \displaystyle\min_{\boldsymbol{w}, b, \alpha, \boldsymbol{z}} & \displaystyle\alpha + \frac{1}{(1-\beta)m} \sum_{i \in M} z_i \\ \text{subject to} & z_i + y_i(\, \langle \boldsymbol{w}, \boldsymbol{v}_i \rangle + b\,) + \alpha \geq 0, \quad i \in M, \\ & z_i \geq 0, \quad i \in M, \\ & \boldsymbol{w}^\top \boldsymbol{w} = 1, \end{array} \tag{4}$$

where $\boldsymbol{v}_i$ is obtained from the decomposition of the kernel matrix such as

$$\begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots k(\boldsymbol{x}_1, \boldsymbol{x}_m) \\ & \cdots \\ k(\boldsymbol{x}_m, \boldsymbol{x}_1) & \cdots k(\boldsymbol{x}_m, \boldsymbol{x}_m) \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^\top \\ \vdots \\ \boldsymbol{v}_m^\top \end{bmatrix} [\boldsymbol{v}_1 \ \cdots \ \boldsymbol{v}_m] \, .$$

By solving the problem, we have an optimal solution $(\boldsymbol{w}^*, b^*)$ and its KKT multipliers $(\boldsymbol{\lambda}^*, \delta^*)$, where δ^* corresponds to $\boldsymbol{w}^{*\top} \boldsymbol{w}^* = 1$ and λ_i^* does to $z_i^* + y_i(\, \langle \boldsymbol{w}^*, \boldsymbol{v}_i \rangle + b^*\,) + \alpha^* \geq 0$. A detailed discussion on these KKT multipliers is shown in the next section. Using the relation $\boldsymbol{w}^* = \frac{1}{\delta^*} \sum_{i \in M} \lambda_i^* y_i \boldsymbol{v}_i$, we can estimate the label of new data point $\boldsymbol{x}$ as $h(\boldsymbol{x}) = \text{sign}(\frac{1}{\delta^*} \sum_{i \in M} \lambda_i^* y_i k(\boldsymbol{x}, \boldsymbol{x}_i) + b^*)$.

When a positive definite matrix is used for $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ as in radial basis functions (RBF), a convex QP with any $\beta \in [\beta_{min}, 1)$ provides a negative optimal value, and we need not to solve nonconvex kernelized β-SVC (4). When a positive semi-definite kernel matrix $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is chosen as in linear or polynomial kernel, the nonconvex case of β-SVC is caused with large β.

3 Modified Extended ν-SVC Algorithm for nonconvex β-SVC

Now we consider the existing solution methods for the nonconvex case of β-SVC (4) with $\beta \in [\bar{\beta}, 1)$, since we can use an efficient solution method such as the interior-point method and Sequential Minimal Optimization (SMO) algorithm for convex QPs with $\beta \in [\beta_{min}, \bar{\beta})$. The nonconvex constraint $\boldsymbol{w}^\top \boldsymbol{w} = 1$ in Problem (4) can be replaced by $\boldsymbol{w}^\top \boldsymbol{w} \geq 1$ for $\beta \in [\bar{\beta}, 1)$, and thus, the nonconvex case of β-SVC (4) is essentially a nonconvex problem. This type of nonconvex problem, whose feasible region is the intersection of a polyhedral set with a concave inequality (say, $\boldsymbol{w}^\top \boldsymbol{w} \geq 1$), is often referred to a *linear reverse convex program* (LRCP). It is known that a global optimal solution of the LRCP (4) is a *basic solution*, where $r + m + 2$ constraints including one equality-constraint are satisfied with equalities. To solve the LRCP problem, several kinds of global optimization algorithms based on cutting plane methods were proposed (see [3]), but they consume long computation time as the size of the problem becomes large. In this section, we restrict ourselves to finding a local minimizer of (4). We introduce a local solution method, Modified Extended ν-SVC Algorithm, for β-SVC, and investigate the properties of the method.

Algorithm *(Modified Extended ν-SVC Algorithm)*

Step 0. Choose a feasible solution $\boldsymbol{w}_0$ of (4). Let $k = 0$.
Step 1. Solve the linear programming (LP) problem:

$$\begin{aligned} \min_{\boldsymbol{w},b,\alpha,\boldsymbol{z}} \quad & \alpha + \frac{1}{(1-\beta)m} \sum_{i \in M} z_i \\ \text{subject to } & z_i + y_i(\,\langle \boldsymbol{w}, \boldsymbol{v}_i \rangle + b\,) + \alpha \geq 0, \quad i \in M, \\ & z_i \geq 0, \quad i \in M, \\ & \boldsymbol{w}_k^\top \boldsymbol{w} = 1, \end{aligned} \tag{5}$$

and obtain a basic optimal solution $(\bar{\boldsymbol{w}}, \bar{b}, \bar{\alpha}, \bar{\boldsymbol{z}})$.
Step 2. If $\bar{\boldsymbol{w}} = \boldsymbol{w}_k$, terminate with $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*) := (\boldsymbol{w}_k, b_k, \alpha_k, \boldsymbol{z}_k)$. Otherwise, go to Step 3.
Step 3. Compute

$$\boldsymbol{w}_{k+1} := \frac{1}{\|\bar{\boldsymbol{w}}\|} \bar{\boldsymbol{w}}. \tag{6}$$

Let $k = k + 1$ and go to Step 1.

The major difference between the Extended ν-SVC algorithmic implementation proposed by [6] and the above modified one is the choice of $\boldsymbol{w}_{k+1}$. The Extended ν-SVC method [6] chooses $\boldsymbol{w}_{k+1}$ as a convex combination:

$$\boldsymbol{w}_{k+1} := \gamma \boldsymbol{w}_k + (1-\gamma) \bar{\boldsymbol{w}} \tag{7}$$

with $\gamma > 0$ and shows a good compromise value $\gamma = 9/10$ (it may be a mistake. see numerical results).

We can show the following properties for modified Extended ν-SVC algorithm:

- the solution $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ is a local minimizer of nonconvex β-SVC (4), and
- the algorithm terminates within a finite steps.

The change of $\boldsymbol{w}_{k+1}$ from (7) to (6) may be theoretically rather than practically important. There are not so large difference experimentally in computational efficiency if we choose a suitable parameter value for γ (see numerical results). A more efficient method was proposed in Gotoh and Takeda [2]. In the method, basic feasible solutions are followed as in the simplex method, so that the objective function value decreases. Compared to the implementation of the solution method of [2], that of modified Extended ν-SVC algorithm is much easy because we can use a standard LP solver in Step 1.

Proposition 1. *Modified Extended ν-SVC Algorithm terminates within a finite steps.*

Proof. This statement can be proved in the same way with [11]. We define

$$(\boldsymbol{w}_{k+1}, b_{k+1}, \alpha_{k+1}, \boldsymbol{z}_{k+1}) := \frac{1}{\|\bar{\boldsymbol{w}}\|} (\bar{\boldsymbol{w}}, \bar{b}, \bar{\alpha}, \bar{\boldsymbol{z}})$$

for the optimal solution $(\bar{\boldsymbol{w}}, \bar{b}, \bar{\alpha}, \bar{\boldsymbol{z}})$ of the LP solved at the kth iteration. The algorithm produces basic feasible solutions $(\boldsymbol{w}_k, b_k, \alpha_k, \boldsymbol{z}_k)$, $k = 1, 2, \ldots$, of nonconvex β-SVC (4). The number of basic solutions of (4) is finite, and thus, if $(\boldsymbol{w}_k, b_k, \alpha_k, \boldsymbol{z}_k)$, $\forall k$, are all distinct, the algorithm terminates within finite iterations. Therefore, it suffices to show that those solutions are distinct. By denoting the objective function of (4) by $q(\alpha, \boldsymbol{z})$, we have

$$q(\alpha_k, \boldsymbol{z}_k) > q(\bar{\alpha}, \bar{\boldsymbol{z}}) \geq q(\alpha_{k+1}, \boldsymbol{z}_{k+1}) \tag{8}$$

at the kth iteration. The first inequality comes from the optimality of $(\bar{\boldsymbol{w}}, \bar{b}, \bar{\alpha}, \bar{\boldsymbol{z}})$ for the LP, and the second one from the observations: $\|\bar{\boldsymbol{w}}\| > 1$ and $q(\bar{\alpha}, \bar{\boldsymbol{z}}) \geq 0$. Note that $\|\bar{\boldsymbol{w}}\| > 1$ is ensured by $\boldsymbol{w}_k^\top \bar{\boldsymbol{w}} = 1$, and $q(\bar{\alpha}, \bar{\boldsymbol{z}}) \geq 0$ is by the nonnegativity of the optimal value of nonconvex β-SVC (4) with $\beta \in [\bar{\beta}, 1)$. The relation (8) implies that $(\boldsymbol{w}_k, b_k, \alpha_k, \boldsymbol{z}_k)$, $\forall k$, are all distinct.

Under a suitable constraint qualification, a local minimizer $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ of Problem (4) satisfies the Karush-Kuhn-Tucker (KKT) conditions: there is a vector $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\delta \in \mathbb{R}$ satisfying

$$\begin{cases} \delta \geq 0 \text{ (convex case)}, \quad \delta \leq 0 \text{ (nonconvex case)}, \\ 0 \leq \lambda_i \leq \dfrac{1}{(1-\beta)m}, \ i \in M, \\ \displaystyle\sum_{i \in M} \lambda_i y_i \boldsymbol{v}_i = \delta \boldsymbol{w}^*, \ \sum_{i \in M} \lambda_i y_i = 0, \ \sum_{i \in M} \lambda_i = 1, \\ \lambda_i \{ z_i^* + y_i(\langle \boldsymbol{w}^*, \boldsymbol{v}_i \rangle + b^*) + \alpha^* \} = 0, \quad i \in M, \\ z_i^* \left(\dfrac{1}{(1-\beta)m} - \lambda_i \right) = 0, \quad i \in M. \end{cases} \tag{9}$$

The vector $\boldsymbol{\lambda}$ of λ_i, $i \in M$, and δ are called KKT multipliers. The last two equations of (9) are called complementary conditions. We also call a point $(\boldsymbol{w}, b, \alpha, \boldsymbol{z})$ satisfying KKT conditions as a *KKT point*.

We can show that the classical Mangasarian-Fromovitz constraint qualification [5] holds at a local minimizer. Therefore, the KKT optimality conditions are necessary conditions for a local minimizer. Let us recall the constraint qualification. Let $q_i(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) \geq 0$, $i \in M$, be the first set of constraints in Problem (4), $r_j(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) \geq 0$, $j \in M$, be the second one, and $s(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) = 0$ be the equality-constraint. Also, define sets of active inequality-constraints as

$$\mathcal{I}(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) := \{ i \in M \ : \ q_i(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) = 0 \},$$
$$\mathcal{J}(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) := \{ j \in M \ : \ r_j(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) = 0 \}.$$

Then, for a local minimizer $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$, the constraint qualification requires that there exists a direction $\boldsymbol{d} = (\boldsymbol{d}_1, d_2, d_3, \boldsymbol{d}_4)$ such that

$$\begin{array}{ll} \nabla q_i(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)^\top \boldsymbol{d} = y_i \boldsymbol{v}_i^\top \boldsymbol{d}_1 + y_i d_2 + d_3 + \boldsymbol{e}_i^\top \boldsymbol{d}_4 > 0, & i \in \mathcal{I}(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*) \\ \nabla r_j(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)^\top \boldsymbol{d} = \boldsymbol{e}_j^\top \boldsymbol{d}_4 > 0, & j \in \mathcal{J}(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*) \\ \nabla s(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)^\top \boldsymbol{d} \ = 2 \boldsymbol{w}^{*\top} \boldsymbol{d}_1 = 0, & \end{array}$$

where $\boldsymbol{e}_i$ indicates the ith unit coordinate vector. Indeed, there exists a direction $\boldsymbol{d} = (\boldsymbol{0}, 0, 0, \boldsymbol{e})$, where $\boldsymbol{e} = (1, 1, \ldots, 1)^\top$, satisfying the above strict inequalities and one equality, and thus, Mangasarian-Fromovitz constraint qualification is satisfied at the local minimizer.

We see that $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ of modified Extended ν-SVC algorithm satisfies the KKT conditions (9). Note that $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ is an optimal solution for the LP:

$$\begin{aligned}\min_{\boldsymbol{w},b,\alpha,\boldsymbol{z}} \quad & \alpha + \frac{1}{(1-\beta)m}\sum_{i\in M} z_i \\ \text{subject to } & z_i + y_i(\,\langle \boldsymbol{w}, \boldsymbol{v}_i\rangle + b\,) + \alpha \geq 0, \quad i \in M, \\ & z_i \geq 0, \quad i \in M, \\ & \boldsymbol{w}^{*\top}\boldsymbol{w} = 1,\end{aligned}$$

and thus, $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ must satisfy the KKT conditions of the LP, which are equal to (9). If $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ is a unique nondegenerate optimal solution of the LP, it has $r+m+2$ linearly independent active constraints and satisfies strict complementarity.

Proposition 2. *A solution of Modified Extended ν-SVC Algorithm is a local minimizer of nonconvex β-SVC (4), if it is unique and nondegenerate.*

Proof. The statement can be shown by mimicking the proof in [2], but let us sketch it below to make this paper self-contained. Suppose that $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ is a solution of Modified Extended ν-SVC Algorithm, and $\boldsymbol{\lambda}^*$ and $\delta^*(\leq 0)$ are corresponding KKT multipliers. Then, for any feasible perturbation $(\Delta\boldsymbol{w}, \Delta b, \Delta\alpha, \Delta\boldsymbol{z})$ of Problem (4) from $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$, the increase Δq of the objective function value is evaluated as

$$\begin{aligned}\Delta q &= \Delta\alpha + \frac{1}{(1-\beta)m}\sum_{i\in M}\Delta z_i \\ &= \underbrace{\left(\boldsymbol{\lambda}^{*\top} \middle| \boldsymbol{\mu}^{*\top}\right)}_{\boldsymbol{\eta}^{*\top}} \underbrace{\left(\begin{array}{c|c|c|c} y_1\boldsymbol{v}_1^\top & y_1 & 1 & \\ \vdots & \vdots & \vdots & \boldsymbol{I} \\ y_m\boldsymbol{v}_m^\top & y_m & 1 & \\ \hline \boldsymbol{O} & \mathbf{0} & \mathbf{0} & \boldsymbol{I}\end{array}\right)}_{\boldsymbol{V}} \underbrace{\begin{pmatrix}\Delta\boldsymbol{w} \\ \Delta b \\ \Delta\alpha \\ \Delta\boldsymbol{z}\end{pmatrix}}_{\Delta\boldsymbol{d}} - \delta^*\Delta\boldsymbol{w}^\top\boldsymbol{w}^*,\end{aligned}$$

since $\boldsymbol{\eta}^{*\top}\boldsymbol{V} - \delta^*(\boldsymbol{w}^{*\top}, 0, 0, \mathbf{0}^\top) = (\mathbf{0}^\top, 0, 1, \frac{1}{(1-\beta)m}\boldsymbol{e})$ follows from (9). The vector $\boldsymbol{\mu}^*$ in the above equation corresponds to slack variables for inequalities $\lambda_i^* \leq \frac{1}{(1-\beta)m}$, $i \in M$. Firstly, the feasible perturbation $\Delta\boldsymbol{d}$ satisfies

$$\Delta\boldsymbol{w}^\top\boldsymbol{w}^* = -\frac{1}{2}\|\Delta\boldsymbol{w}\|^2 \tag{10}$$

because of $(\boldsymbol{w}^* + \Delta\boldsymbol{w})^\top(\boldsymbol{w}^* + \Delta\boldsymbol{w}) = 1$. From the following discussion, we have the second condition:

$$\boldsymbol{\eta}^{*\top}\boldsymbol{V}\Delta\boldsymbol{d} > 0. \tag{11}$$

Using the notation of $\boldsymbol{d}^\top \equiv (\boldsymbol{w}^\top, b, \alpha, \boldsymbol{z}^\top) \in \mathbb{R}^n$, where $n = r+m+2$, we describe the linear constraints of Problem (4) as $\boldsymbol{V}\boldsymbol{d} \geq \mathbf{0}$. We separate the active constraints $\boldsymbol{V}_B\boldsymbol{d}^* = \mathbf{0}$, where $\boldsymbol{V}_B$ is a $(n-1)\times n$ submatrix of $\boldsymbol{V}$, from $\boldsymbol{V}\boldsymbol{d}^* \geq \mathbf{0}$ at $\boldsymbol{d}^{*\top} = (\boldsymbol{w}^{*\top}, b^*, \alpha^*, \boldsymbol{z}^{*\top})$. Then, for a feasible perturbation $\Delta\boldsymbol{d}$, the vector $\boldsymbol{V}_B\Delta\boldsymbol{d}$ $(\geq \mathbf{0})$ has at least one positive component, since (10) excludes the direction $\Delta\boldsymbol{d} = \gamma\boldsymbol{d}^*$ with small $\gamma \neq 0$, which is the

Table 1. The UCI datasets used in the experiments

	m	n	$\bar{\beta}$	β^*	1-ν^*	β^*-SVC [%]	ν^*-SVC [%]
diabetes	768	6	0.50	0.55	0.45	23.17	23.30
heart	270	13	0.70	0.70	0.65	15.93	15.93
liver-disorders	345	6	0.30	0.50	0.25	28.45	31.61
wdbc	569	*3	0.95	0.95	0.85	2.99	3.34

only one satisfying $\boldsymbol{V}_B \Delta \boldsymbol{d} = \boldsymbol{0}$, from feasible directions. Therefore, we see that a feasible perturbation $\Delta \boldsymbol{d}$ satisfies (11), that is, $\boldsymbol{\eta}^{*\top} \boldsymbol{V} \Delta \boldsymbol{d} = \boldsymbol{\eta}_B^{*\top} \boldsymbol{V}_B \Delta \boldsymbol{d} > 0$, where $\boldsymbol{\eta}_B^*$ $(> \boldsymbol{0})$ are the corresponding KKT multipliers for $\boldsymbol{V}_B \boldsymbol{d}^* = \boldsymbol{0}$. From (10) and (11), we have $\Delta q > 0$ for any sufficiently small feasible perturbation $(\Delta \boldsymbol{w}, \Delta b, \Delta \alpha, \Delta \boldsymbol{z})$ even if δ^* is negative. Consequently, $(\boldsymbol{w}^*, b^*, \alpha^*, \boldsymbol{z}^*)$ is proved to be locally optimal.

4 Numerical Results

To demonstrate the performance improvement of Modified Extended ν-SVC Algorithm, we apply it and the original Extended ν-SVC Algorithm of [6] to several datasets of the UCI Machine Learning Data Repository [1]. Performance is evaluated in terms of the average number of LPs, solved in Step 1 of those algorithms.

We used SeDuMi [9] software to solve a convex quadratic optimization problem of convex β-SVC, and LPs successively induced from nonconvex β-SVC. The SeDuMi solver is a Matlab implemented interior-point method for optimization over symmetric cones. All computations were conducted on an Opteron 850 (2.4GHz) with 8GB of physical memory.

Table 1 summarizes the UCI datasets used in the experiments. Each examples in the Wisconsin breast cancer wdbc has 30 attributes, but we use three attributes: "mean texture", "worst area" and "worst smoothness", since several researchers demonstrated highly accurate classification results using these three attributes (for example, [4]). m denotes the number of training examples in a dataset, and n does that of the attributes. The data are scaled linearly such that the values of each attribute lie between -1 and 1. Then, we solved the problem of β-SVC with a linear kernel using Modified Extended ν-SVC Algorithm, and measured test error rates using 10-fold cross-validation over different values β. $\bar{\beta}$ indicates the turning point from convexity to nonconvexity. In this numerical experiments, $\bar{\beta}$ was defined such that β-SVC is convex during $\beta \in [\beta_{min}, \bar{\beta} - 0.05]$ and nonconvex during $\beta \in [\bar{\beta}, 0.95]$. The values of β^* and ν^* are the best for nonconvex β-SVC and ν-SVC, respectively, in terms of the minimum test error rates. Those value of β^* was found by increasing 0.05 from $\bar{\beta}$ to 0.95. On the other hand, $1 - \nu^*$ were found similarly from β_{min} to $\bar{\beta} - 0.05$. β^*-SVC and ν^*-SVC indicate the minimum test error rates of β^*-SVC and ν^*-SVC, respectively. The test error rates of β-SVC with a linear kernel are not larger than those of ν-SVC in three datasets.

Table 2 shows the average number of LPs, which were solved for β^*-SVC by original Extended ν-SVC Algorithm of [6] and its modified one, over 10 runs. The best parameter values β^* of each dataset are given in Table 1. An initial feasible solution $\boldsymbol{w}_0$ in both algorithms is set to a previously obtained local minimizer of β-SVC with $\beta = \beta^* - 0.05$. The difference in the results is due to the setting of $\boldsymbol{w}_{k+1}$, (6) or (7). For $\boldsymbol{w}_{k+1}$ of (7), $\gamma = 9/10$ is recommended

Table 2. The average number of LPs for β^*-SVC

	modified	orig. ($\gamma = 1/10$)	orig. ($\gamma = 5/10$)	orig. ($\gamma = 9/10$)
diabetes	3.7	10.2	10.1	54.4
heart	3.9	32.9	11.2	63.5
liver-disorders	2.7	4.0	7.7	40.5
wdbc	2.0	32.0	6.7	36.3

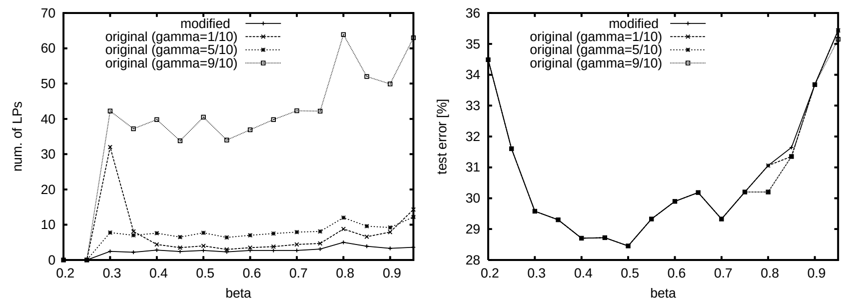

Fig. 2. Results for liver-disorders dataset (Left: the average number of LPs, Right: test error rates [%])

in [6], but we tried three values of γ, $\gamma = 1/10, 5/10$ and $9/10$. The major computations in the algorithm are for solving LPs. The average number of LPs also implies the average iteration number of (modified) Extended ν-SVC Algorithm. We see that the choice (6) of $\boldsymbol{w}_{k+1}$ causes the proposed algorithm to finish faster. By solving a few LPs constructed in the proposed modified algorithm, we obtained a local minimum for the nonconvex QP of β-SVC.

As for liver-disorders, Fig. 2 (left) shows the change in the average number of LPs with respect to $\beta \geq \bar{\beta} = 0.30$. The modified algorithm generates smaller number of LPs with any β. When $\gamma = 1/10$ is chosen in Extended ν-SVC algorithm, a peek occurs in the graph at the threshold $\beta = \bar{\beta}$, where an optimal solution of convex β-SVC with $\beta = \bar{\beta} - 0.05$ is used for $\boldsymbol{w}_0$, not only in liver-disorders but also in other datasets. Fig. 2 (right) implies that Extended ν-SVC algorithm and the modified one achieved small test error with nonconvex value $\beta \geq \bar{\beta}$. There are small differences in test error rates due to different local minimizers.

5 Concluding Remarks

We proposed a modification for the existing algorithm of Extended ν-SVC. The modification is theoretically rather than practically important. Indeed, it makes possible to analyze the finite convergence and local optimality of the algorithm. From the experimental results, we also see that the modification causes the algorithm to finish faster after solving a few LPs without choosing a suitable parameter value such as γ of Extended ν-SVC Algorithm.

References

1. C.L. Blake and C.J. Merz, UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science, 1998.

2. J. Gotoh and A. Takeda, A Linear Classification Model Based on Conditional Geometric Score, *Pacific Journal of Optimization* 1 (2005) 277-296.
3. R. Horst and H. Tuy, Global Optimization: Deterministic Approaches, 3rd edition, Springer-Verlag, Berlin, 1995.
4. O.L. Mangasarian, W.N. Street and W.H.Wolberg, Breast cancer diagnosis and prognosis via linear programming, *Operations Research* 43 (1995) 570-577.
5. O.L. Mangasarian and S. Fromovitz, The Fritz-John necessary optimality conditions in presence of equality and inequality constraints, *Journal of Mathematical Analysis and Applications* 17 (1967) 37-47.
6. F. Perez-Cruz, J. Weston, D. J. L. Hermann and B. Schölkopf, Extension of the ν-SVM Range for Classification, *Advances in Learning Theory: Methods, Models and Applications 190*, (Eds.) J.A.K. Suykens, G. Horvath, S. Basu, C. Micchelli and J. Vandewalle, IOS Press, Amsterdam (2003) 179-196.
7. R.T. Rockafellar and S. Uryasev, Conditional value-at-risk for general loss distributions, *Journal of Banking and Finance* 26 (2002) 1443-1471.
8. B. Schölkopf, A. J. Smola, R. C. Williamson and P. L. Bartlett, New support vector algorithms, *Neural Computation* 12 (2000) 1207-1245.
9. J. F. Sturm, "Using SeDuMi 1.02, a Matlab Toolbox for Optimization over Symmetric Cones," *Optimization Methods and Software* 11-12 (1999) 625-653.
10. A. Takeda, Support Vector Machine Based on Conditional Value-at-Risk Minimization, Research Report B-439; Dept. of Mathematical and Computing Sciences; Tokyo Institute of Technology; 2007. www.is.titech.ac.jp/research/research-report/B/index.html
11. A. Takeda and H. Nishino, On measuring the inefficiency with the inner-product norm in date envelopment analysis, *European J. Oper. Res.* 133 (2001) 377-393.