Linear and Nonlinear Analysis
Yokohama Publishers
ISSN 2188-8167
Volume 5, Number 1, 2019, 105–111

ON SUPPORT VECTOR MACHINE CLASSIFIERS WITH UNCERTAIN KNOWLEDGE SETS

GUE MYUNG LEE

This paper is dedicated to Professor Jong Soo Jung on occasion of his 65-th birthday.

Abstract. In this paper, following the approaches of Jeyakumar, Li and Suthaharan [6], we study a support vector machine (SVM) classifier with a uncertain knowledge set defined by polyhedral uncertain sets. We prove the robust Farkas lemma for our SVM, and then we get a deterministic quadratic optimization problem from the SVM. We formulate the Wolfe dual problem for our SVM, and then by using the dual problem, we obtain the convergence theorem and the algorithm for finding the solution of the quadratic optimization problem.

1. Introduction

Support vector machines (SVMs) [2, 7, 11, 14] can be presented by linear optimization problems or convex quadratic optimization problems. Incorporating prior knowledge into SVMs in the form of knowledge sets often improves corrections of the classifier or reduce the amount of training data needeed [5, 6, 8, 9, 12]. Very recently, Jeyakumar, Li and Suthaharan [6] studied SVM classifies in the face of uncertain knowledge sets, defined by interval uncertain sets, by using the robust approach (the worst approach). In this paper, following the approaches of Jeyakumar, Li and Suthaharan [6], we will study SVM classifiers in face of uncertain knowledge sets defined by polyhedral uncertain sets .

This paper is concerned with a SVM classifiers with uncertain knowledge sets, defined by polyhedral uncertainty sets, by using the approaches of Jeyakumar, Li and Suthaharan [6]. We will explain our SVM classifier with uncertain knowledge sets and show that the classifiers will be a convex quadratic optimization problem, and then we formulate its Wolfe dual problem, By using the dual problem, we get the convergence theorem and the algorithm for finding the solution of the quadratic optimization problem.

2. Support vector machine classifiers with uncertain knowledge sets

Following the approaches of Jeyakumar, Li and Suthaharan [6], we will explain our SVM classifier with uncertain knowledge sets, and related optimization problems. The classical SVM problem is formulated as discriminating between m data

2010 *Mathematics Subject Classification.* 65K10, 90C25, 90M45.

Key words and phrases. Support vector machine classifier, uncertain knowledge set, the Wolfe dual problem, quadratic optimization problems, convergence theorem, algorithm .

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MIST) (NRF-2017R1E1A1A03069931).

points in $\mathbb{R}^n$. The points are presented by an $m \times n$ matrix A, with the i-th data point a_i is at the i-th row vector of A. Each point is considered to be either class $\mathcal{A}$ or $\mathcal{B}$, which is recorded along the diagonal of the diagonal matrix $D \in \mathbb{R}^{m\times m}$. The diagonal elements $D_{ii} = +1$, if the point a_i belongs to $\mathcal{A}$ and $D_{ii} = -1$, if the point belongs to $\mathcal{B}$. We discriminate between the two data sets with the hyperplane $\{a \in \mathbb{R}^n \mid a^T w = r\}$ defined by a discriminant function $f(w) = a^T w - r$. To allow some slight error which is represented by slack variables, we consider the following classical SVM [7,13]:

$$\begin{array}{lll} \text{(SVM)} & \min_{w.r.y} & l(w) + \mu e_m^T y \\ & \text{s.t.} & D(Aw - re_m) + y \geqq e, \quad y \geqq 0, \end{array}$$

where μ is a weighting parameter and $w \in \mathbb{R}^n$, $r \in \mathbb{R}$, $y \in \mathbb{R}^m$ and $e_m \in \mathbb{R}^m$ is a vector of ones. Sometimes, $l(w) := \frac{1}{2}\|w\|_2^2$. When $l(w) := \frac{\lambda_1}{2}\|w\|_2^2 + \lambda_2\|w\|_1$, the model (SVM) becomes the doubly regularized SVM [6,15]:

$$\begin{array}{lll} \text{(KBP)} & \min_{w.r.y} & \dfrac{\lambda_1}{2}\|w\|_2^2 + \lambda_2\|w\|_1 + \mu e_m^T y \\ & \text{s.t.} & D(Aw - re_m) + y \geqq e, \quad y \geqq 0. \end{array}$$

Here we assume that there are uncertain knowledge sets $\text{co}\{u_1^i, \ldots, u_l^i\}$, $u_1^i, \ldots, u_l^i \in \mathbb{R}^n, i = 1, \ldots, k$, where co$A$ is the convexhull of the set A, and that the uncertain knowledge set $\{z \in \mathbb{R}^n \mid h_i^T z \leqq d_i,\ i = 1, \ldots, k\}$ lies on class $\mathcal{A}$'s side of the bounding hyperplane $w^T z = r + 1$.

(KBP) becomes the following knowledge-based support vector machine problem under data uncertainty:

$$\begin{array}{lll} \text{(UKBP)} & \min_{w.r.y} & \dfrac{\lambda_1}{2}\|w\|_2^2 + \lambda_2\|w\|_1 + \mu e^T y \\ & \text{s.t.} & D(Aw - re_m) + y \geqq e, \quad y \geqq 0, \\ & & \{z \in \mathbb{R}^n \mid h_i^T z_i \leqq d_i,\ h_i \in \text{co}\{u_1^i, \ldots, u_l^i\}, i = 1, \ldots, k\} \\ & & \quad \subset \{z \in \mathbb{R}^n \mid w^T z \geqq r + 1\}. \end{array}$$

Jeyakumar, Li and Suthaharan [6] used the interval sets $[\underline{h}_i, \bar{h}_i] \times [\underline{d}_i, \bar{d}_i]$ ($\underline{h}_i, \bar{h}_i \in \mathbb{R}^n$ with $\underline{h}_i < \bar{h}_i$ and $\underline{d}_i, \bar{d}_i \in \mathbb{R}$ with $\underline{d}_i \leqq \bar{d}_i$ for $i = 1, \ldots, k$), which was called interval uncertain knowlege sets, as uncertain sets for $(h_i, d_i, i = 1, \ldots, k)$.

Following robust optimization approaches in [1,6], we formulate the following robust counterpart of (UKBP), which is a deterministic optimization problem, given by

$$\begin{array}{lll} \text{(RKBP)} & \min_{w.r.y} & \dfrac{\lambda_1}{2}\|w\|_2^2 + \lambda_2\|w\|_1 + \mu e_m^T y \\ & \text{s.t.} & D(Aw - re_m) + y \geqq e_m, \quad y \geqq 0, \\ & & \{z \in \mathbb{R}^n \mid h_i^T z_i \leqq d_i,\ \forall h_i \in \text{co}\{u_1^i, \ldots, u_l^i\},\ i = 1, \ldots, k\} \\ & & \quad \subseteq \{z \in \mathbb{R}^n \mid w^T z \geqq r + 1\}. \end{array}$$

3. Robust Farkas lemma

Jeyakumar, Li and Suthaharan [6] proved the robust Farkas lemma for their SVM classifier with the knowledge sets defined by interval uncertain sets, and then get a convex optimization problem from the counter part of their SVM. Here we derive the robust Farkas lemma for our SVM classifier, and then following the approaches of Jeyakumar, Li and Suthaharan [6], we will convert the above robust counterpart (RKBP) into a convex quadratic optimization problem.

Lemma 3.1 (Robust Farkas Lemma). *Let* $d_i \in \mathbb{R}$, $i = 1, \ldots, k$, $u^i_j \in \mathbb{R}^n$, $i = 1, \ldots, k$, $j = 1, \ldots, l$, $w \in \mathbb{R}^n$ *and* $r \in \mathbb{R}$. *Then the following are equivalent:*

(i) $h_i^T x \leqq d_i \ \forall h_i \in \mathcal{V}_i := co\{u^i_1, \ldots, u^i_l\} \Rightarrow w^T x \geqq r + 1$;

(ii) *there exist* $\lambda^i_j \geqq 0$, $i = 1, \ldots, k$, $j = 1, \ldots, l$ *such that*

$$-(r+1) - \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda^i_j d_i \geqq 0 \ \textit{and} \quad w + \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda^i_j \mu^i_j = 0.$$

Proof. $h_i^T x - d_i \leqq 0 \quad \forall h_i \in \mathcal{V}_\rangle$

$$\iff \forall \lambda_j \geqq 0 \quad \text{with} \quad \sum_{j=1}^{l} \lambda_j = 1, \ \sum_{j=1}^{l} \lambda_j [u_j^{iT} x - d_i] \leqq 0$$

$$\iff u_j^T x - d_i \leqq 0, \quad j = 1, \ldots, l.$$

Thus we have

$$(i) \iff [u_j^{iT} x - d_i \leqq 0, \ i = 1, \ldots, k, j = 1, \ldots, l \quad \Rightarrow w^T x \geqq r + 1]$$

$\iff$ (by classical Farkas Lemma in [10])

there exist $\lambda^i_j \geqq 0$, $i = 1, \ldots, k, j = 1, \ldots, l$ such that

$$w^T x - (r+1) + \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda^i_j (u_j^{iT} x - d_i) \geqq 0 \ \ \forall x \in \mathbb{R}^n$$

$\iff$ there exist $\lambda^i_j \geqq 0$, $i = 1, \ldots, k, j = 1, \ldots, l$ such that

$$-(r+1) + \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda^i_j d_i \geqq 0 \text{ and } w^T x + \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda^i_j u_i^{iT} x \geqq 0 \ \ \forall x \in \mathbb{R}^n$$

$\iff$ there exist $\lambda^i_j \geqq 0$, $i = 1, \ldots, k, j = 1, \ldots, l$ such that

$$-(r+1) - \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda^i_j d_i \geqq 0 \text{ and } w + \sum_{i=1}^{k}\sum_{j=1}^{l} \lambda^i_j u^i_i = 0.$$

□

By the Robust Farkas Lemma (Lemma 3.1), we can rewrite (RKBP) as follows;

$$\text{(RKBP)} \qquad \text{minimize}_{w.r.y} \quad \frac{\lambda_1}{2}\|w\|_2^2 + \lambda_2 \|w\|_1 + \mu e_m^T \xi$$
$$\text{s.t.} \qquad D(Aw - re_m) + \xi \geqq e_m$$

$$-w-\sum_{i=1}^{k}\sum_{j=1}^{l}\lambda_j^i u_j^i \geqq 0$$
$$w+\sum_{i=1}^{k}\sum_{j=1}^{l}\lambda_j^i u_j^i \geqq 0$$
$$-(r+1)-\sum_{i=1}^{k}\sum_{j=1}^{l}\lambda_j^i d_i \geqq 0$$
$$\xi \geqq 0,\ \lambda_j^i \geqq 0,\ i=1,\dots,k,\ j=1,\dots,l.$$

Here we insert slack variables $\xi_1, \xi_2 \in \mathbb{R}_+^m$ and $\beta \geqq 0$ in (RKBP) and let $(w_i)_+ := \max\{0, w_i\} = p_i,\ (w_i)_- := -\min\{0, w_i\} = q_i,\ p = (p_1, \dots, p_n)$ and $q = (q_1, \dots, q_n)$, and then noticing that $\|w\|_2^2 = \|p\|_2^2 + \|q\|_2^2$ and $\|w\|_1 = e_n^T(p+q)$, we get the following optimization problem (P_0) from (RKBP):

$$(\text{P}_0) \qquad \text{minimize}_{(p,q,\xi_1,\xi_2,r,\lambda_i^j)} \quad \frac{\lambda_1}{2}(\|p\|_2^2+\|q\|_2^2)+\lambda_2 e_n^T(p+q)+\mu e_m^T\xi+e_n^T(\xi_1+\xi_2)+\beta$$
$$\text{s.t.} \quad D(A(p-q)-re_m)+\xi \geqq e_m$$
$$-(r+1)-\sum_{i=1}^{k}\sum_{j=1}^{l}\lambda_j^i d_i+\beta \geqq 0$$
$$-p+q-\sum_{i=1}^{k}\sum_{j=1}^{l}\lambda_j^i u_j^i+\xi_1 \geqq 0$$
$$p-q+\sum_{i=1}^{k}\sum_{j=1}^{l}\lambda_j^i u_j^i+\xi_2 \geqq 0$$
$$p,q \geqq 0,\ \xi \geqq 0,\ \xi_1,\xi_2 \geqq 0,\ \beta \geqq 0$$
$$\lambda_i^j \geqq 0,\ i=1,\dots,k,\ j=1,\dots,l.$$

Now we try to change (P_0) into a quadratic optimization problem in matrix form. Let

$$y=\begin{pmatrix} p\\ q\\ \lambda^1\\ \vdots\\ \lambda^k\end{pmatrix}\in\mathbb{R}^{2n+kl},\quad \tilde{b}=\lambda_2\begin{pmatrix} e_n\\ e_n\\ 0_{kl}\end{pmatrix}\in\mathbb{R}^{2n+k},\quad v=\begin{pmatrix}\xi\\ \xi_1\\ \xi_2\\ \beta\end{pmatrix}\in\mathbb{R}^{m+2n+1},$$

$$\widetilde{e}=\begin{pmatrix} e_m\\ 0_n\\ 0_n\\ 1\end{pmatrix}\in\mathbb{R}^{m+2n+1},\quad \widetilde{C}=\lambda_1\begin{pmatrix} I_{n\times n} & 0_{n\times n} & 0_{n\times kl}\\ 0_{n\times n} & I_{n\times n} & 0_{n\times kl}\\ 0_{kl\times n} & 0_{kl\times n} & 0_{kl\times kl}\end{pmatrix}\in\mathbb{R}^{(2n+kl)\times(2n+kl)},$$

$$U^i=(u_1^i \quad \dots \quad u_l^i)\in\mathbb{R}^{n\times l},\ i=1,\dots,k,$$

$$\widetilde{A} = \begin{pmatrix} A & -A & 0_{m\times kl} \\ -I_{n\times n} & I_{n\times n} & -U^1 - U^2 - \cdots - U^k \\ I_{n\times n} & -I_{n\times n} & U^1 \; U^2 \; \ldots \; U^k \\ 0\ldots 0 & 0\ldots 0 & -d_1 \cdots - d_1 \cdots - d_k \cdots - d_k \end{pmatrix} \in \mathbb{R}^{(m+2n+1)\times(2n+kl)}$$

$$\widetilde{D} = \begin{pmatrix} D & 0_{m\times(2n+1)} \\ 0_{(2n+1)\times m} & I_{(2n+1)\times(2n+1)} \end{pmatrix} \in \mathbb{R}^{(m+2n+1)\times(m+2n+1)}.$$

Then we can rewrite (P_0) as the following quadratic optimization problem:

$$(\mathrm{P}_0) \qquad \text{minimize}_{(y,v,r)\in\mathbb{R}^{(2n+kl)}\times\mathbb{R}^{(m+2n+1)}\times\mathbb{R}} \quad \frac{1}{2}y^T\widetilde{C}y + \tilde{b}^Ty + e_{m+2n+1}^T v$$
$$\text{s.t.} \quad \widetilde{D}(\widetilde{A}y - r\widetilde{e}) + v \geqq \widetilde{e}$$
$$y \geqq 0, \quad v \geqq 0.$$

Following the approaches of Jeyakumar, Li and Suthaharan [6], and Mangasarian and Musicant [13], we replace $\tilde{b}^T y$ by $y^T y$ and $e_{m+2n+1}^T v$ by $\frac{1}{2}v^T v$ and we append an additional $\frac{r^2}{2}$ to the objective function, we get the following quadratic optimization problem (P) which we will treat with:

$$(\mathrm{P}) \qquad \text{minimize}_{(y,v,r)\in\mathbb{R}^{(2n+k)}\times\mathbb{R}^{(m+2n+1)}\times\mathbb{R}} \quad \frac{1}{2}y^T(\widetilde{C}+\mu I)y + \frac{1}{2}r^2 + \frac{1}{2}\|v\|_2^2$$
$$\text{s.t} \quad \widetilde{D}(\widetilde{A}y - r\widetilde{e}) + v \geqq \widetilde{e}$$

where $\mu \in \mathbb{R}$ is an additional tuning parameter and $\mu > 0$.

4. Dual problem and algorithm

Following the approaches of Jeyakumar, Li and Suthaharan [6], we get the following Wolfe dual problem [10, 16] for (P):

$$(\mathrm{D}) \qquad \text{maximize}_{z\in\mathbb{R}^{m+2n+1}} \quad -\frac{1}{2}z^T\widetilde{Q}z + \widetilde{e}^T z$$
$$\text{s.t.} \quad z \geqq 0,$$

where $\widetilde{Q} = I + \widetilde{D}\widetilde{A}(\widetilde{C}+\mu I)^{-1}(\widetilde{D}\widetilde{A})^T + \widetilde{D}\widetilde{e}(\widetilde{D}\widetilde{e})^T$. Since for any $x \in \mathbb{R}^{m+2n+1}$, $x^T\widetilde{Q}x \geqq \|x\|^2$, then $\widetilde{Q}$ is a positive definite $(m+2n+1)\times(m+2n+1)$ matrix, and so (D) is a strictly concave quadratic maximization problem with non-negative constraints.

Moreover, we have the following existence theorem for the dual problem (D):

Theorem 4.1. *The dual problem (D) has a unique solution.*

Proof. Since $\|\tilde{e}\| = m+1$, we have, for any $z \in \mathbb{R}^{m+2n+1}$, $\frac{1}{2}z^T\widetilde{Q}z - \widetilde{e}^T z \geq \frac{1}{2}\|z\|^2 - (m+1)\|z\| \geq -\frac{1}{2}(m+1)^2$. Thus $\inf\{\frac{1}{2}z^T\widetilde{Q}z - \widetilde{e}^T z \mid z \geq 0\}$ is finite and so it follows from Frank-Wolfe theorem in [3] that the dual problem (D) has a solution. Moreover, since (D) is a strictly concave quadratic optimization problem, (D) has a unique solution. □

Following the proof of Theorem 5.1 of Jeyakumar, Li and Suthaharan [6] , or using the Wolfe's duality theorem and the strict converse duality theorem in page 115 and page 116 in [10], we can prove the following theorem:

Theorem 4.2. *Let $z \in \mathbb{R}^{m+2n+1}$. Then z is a solution of (D) if and only if*

$$((\widetilde{C} + \mu I)^{-1}(\widetilde{D}\widetilde{A})^T z,\ z,\ -(\widetilde{D}\widetilde{e})^T)z)$$

is a solution of (P).

Following the proof of Theorem 5.2 of Jeyakumar, Li and Suthaharan [6], we can prove the following theorem:

Theorem 4.3. *Let $\alpha \in (0,2)$. Let $\{z_n\}$ be a sequence generated by the iteration defined below:*

$$z_0 = \widetilde{Q}^{-1}\widetilde{e} \quad and \quad z_n = \widetilde{Q}^{-1}(\widetilde{e} + (\widetilde{Q}z_{n-1} - \widetilde{e} - \alpha z_{n-1})_+),$$

where $a_+(= (a_1, \ldots, a_{m+2n+1})_+) = (max\{a_1, 0\}, \ldots, max\{a_{m+2n+1}, 0\})$. Then $\{z_n\}$ converges to the unique solution z of (D).

From Theorem 4.2 and Theorem 4.3, we obtain the following algorithm for finding a solution of (P):

Algorithm for finding a solution of (P):

1. $z_0 = \widetilde{Q}^{-1}\widetilde{e}$.
2. $z_{old} = z_0 + \widetilde{e}$.
3. While $\|z_{old} - z_n\| > tol$.
   $z_{n+1} = \widetilde{Q}^{-1}(\widetilde{e} + (\widetilde{Q}z_n - \widetilde{e} - \alpha z_n)_+)$.

Calculate $((\widetilde{C} + \mu I)^{-1}(\widetilde{D}\widetilde{A})^T z_n,\ z_n,\ -(\widetilde{D}\widetilde{e}^T z_n)$, $\mu = \frac{1}{2}$,
$n = n + 1$.

References

[1] A. Ben-Tal, L. E. Ghaoui and N. Nemirovski, *Robust Optimization*, Princeton Series in Applied Mathematics, 2009.

[2] C. J. C. Burgess, *A tutorial on support vector machines for pattern recognition*, Data Mining and Knowledge Discovery **2** (1998), 121–167.

[3] M. Frank and P. Wolfe,*An Algorithm for quadratic programming*, Naval Research Logistics Quarterly **3** (1956), 95–110.

[4] C. Fung, O.L. Mangasarian and J. Shavlik, *Knowledge-based support vector machine classifiers*, in: Newral Information Processing Systems 15, S. Thurn, S. Becker and K. Obermayer, editors, MIT Press, Cambridge, MA, 2003, pp. 521–528.

[5] V. Jeyakumar, J. Ormerod and R.S. Womersley, *Knowledge-based semidefinite linear programming classifiers*, Optim. Methods Softw. **21** (2006), 693–706.

[6] V. Jeyakumar, G. Li and S. Suthaharan, *Support vector machine classifiers with uncertain knowledge sets via robust optimization*, Optimization **63** (2014), 1099–1116.

[7] H. Nakayama, Y, Yun and M. Yoon, *Sequential Approximate Multiobjective Optimization Using Computational Intelligence*, Springer, 2009

[8] R. Khemchandani, Jeyadeva and S. Chandra, *Knowledge based proximal support vector machines*, Euro. J. Oper. Res. **195**(2009), 914–923.

[9] M. Arun Kumar, R. Khemchandani, M. Gopal and S. Chandra, *Knowledge based least squares twin support vector machines*, Information Sciences **180** (2010), 4606–4618.

[10] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, 1969.

[11] O. L. Mangasarian, *Arbitrary-norm separating plane*, Oper. Res. Lett. **24** (1999), 15–23.

[12] O. L. Mangasarian, *Knowledge-based linear programming*, SIAM J. Optim. **15** (2004/2005), 375–382.

[13] O. L. Mangasarian and David R. Musicant, *Lagrangian support vector machines*, J. Machine Learning Research **1** (2001), 161–177.

[14] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Spring-Verlag, second edition 2000.

[15] L. Wang, J. Zhu and H. Zou, The doubly regularized support vector machine, *Statistia Sinica*, **16** (2006), 589–615.

[16] P. Wolfe, *A duality theorem for nonlinear programming*, Quarterly of Applied mathematics **19** (1961), 239–244.

Manuscript received 2 March 2019
revised 20 March 2019

Gue Myung Lee
Department of Applied Mathematics, Pukyong National University, Busan 48513, Korea
E-mail address: gmlee@pknu.ac.kr