응용통계연구 제 18권 2호, 2005년, pp. 421-434

수리계획법을 이용한 서포트 벡터 기계 방법에 관한 연구*

윤 민[1] 이학배[2]

요 약

기계학습은 패턴분류의 한 도구로써 광범위하게 연구되고 있다. 기계학습 방법들 중에서 서포트 벡터 기계(Support Vector Machines)는 많은 분야에서 연구되어지는 것으로 이진 패턴 분류문제에서 고차원의 특징공간에서 두 집합들 사이에 가장 큰 분리를 제공하는 최대 여유도(margin)를 가지는 분리 초평면을 찾는 것이다. 최대 여유도의 분리의 개념에 기초하여 Mangasarian(1968)은 다중-표면 방법(multi-surface method)을 제안하였고, 1980년대에 목적 계획법을 이용한 방법들이 광범위하게 개발되었다. 본 논문에서는 다목적 계획법과 목적 계획법을 이용한 수리계획법인 서포트 벡터 기계의 두가지 방법들을 제안하고 수치 예제들을 통하여 효용성에 대하여 논의하고자 한다.

주요용어: 서포트 벡터 기계, 최대 여유도 분류기, 다목적 최적화, 목적 계획법.

1. 소개

서포트 벡터 기계(Support Vector Machine, SVM)는 많은 유용한 특징들과 실제적인 수행능력의 장점들을 가지고 있기 때문에 기계학습에 있어서 여러 분야에서 연구되고 있다 (Cherkassky와 Mulier, 1998). 서포트 벡터 기계의 주요한 특징중의 하나는 표본점들에서 분리 초평면 사이의 최소 거리를 최대화함으로서 얻어지는 최대 여유도(maximal margin)를 가지는 선형분류기(linear classifier)에 기초하고 있다.

반응변수가 두개의 값을 갖는 이진(binary)일 경우($y_i = +1$ 혹은 -1)의 패턴 분류문제를 고려한다. 예를들어 판별분석에서 선형판별함수는 다음과 같다.

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0$$

여기서 $\boldsymbol{w}$는 가중치 벡터이고 w_0는 편의(bias)이다. 패턴분류를 위하여 판별분석을 이용하면 아래와 같은 기준에 따른다.

$$\begin{aligned} \boldsymbol{w}^T \boldsymbol{x} + w_0 \geq 0 \quad \text{일 때}, &\qquad y = +1 \\ \boldsymbol{w}^T \boldsymbol{x} + w_0 < 0 \quad \text{일 때}, &\qquad y = -1 \end{aligned}$$

---

* 본 논문은 2004년도 BK21 학술연구비 지원에 의하여 수행되었음.
본 논문은 2004학년도 연세대학교 학술연구비의 (부분적인) 지원에 의하여 이루어진 것임.
1) (120-749) 서울시 서대문구 신촌동 134, 연세대학교 통계연구소, 책임연구원
E-mail: myoon@yonsei.ac.kr
2) (120-749) 서울시 서대문구 신촌동 134, 연세대학교 응용통계학과, 조교수
E-mail: hblee@yonsei.ac.kr

판별분석은 이와같이 설명변수 x_i들의 선형결합에 의하여 패턴분류를 결정한다.

이러한 패턴 분류문제를 해결하기 위한 한 해결책으로 인공 신경망(neural network) 방법이 있다. 인공 신경망 방법중 널리쓰이는 역전파(back propagation)방법은 다중 국소(multiple local) 최적값을 가지는 비선형 최적화의 문제로 약화되는 한편 대규모 자료에 적용 또한 어렵게 된다. 역전파 방법은 강화학습(incremental learning)에서 환경의 변화에 따라서 대안적인 구조를 변화시키기 어려운 또 다른 약점이 있다.

본 논문에서는 패턴분류를 위하여 최적화에 의한 접근방법인 기계학습(machine learning)방법중에 서포트 벡터 기계(Support Vector Machines)에 대하여 논의하고자 한다. 기계학습의 목적은 주어진 훈련 자료집합(training data set) $(\boldsymbol{x}_i, y_i)$ $(i = 1, \ldots, \ell)$을 기초로 새롭게 관측된 특성들이 어느 집합에 속할 것인지를 예측하는데 있다.

Novikoff(1962)는 퍼셉트론(perceptron)의 개념을 이용하여 선형 분류기들을 추출하기 위하여 최대 여유도에 관한 논의를 하였다. 한편 Mangasarian(1968)은 선형 분류기를 추출하기 위하여 다중-표면 방법을 이용하였다. 최대 여유도를 이용한 선형 분류기들은 1980년대에서 1990년 대 초반에 이르는 동안 선형 목적계획법이 광범위 하게 연구되어왔다(Freed와 Glover, 1981, Erenguc와 Koehler, 1990). 본 논문에서는 다목적 계획법(Multi-objective Programming)과 목적 계획법(Goal Programming)을 이용하여 서포트 벡터 기계를 논의하고 이러한 방법들에 의하여 서포트 벡터 기계 알고리즘을 확장하고자 한다.

2절에서는 서포트 벡터 기계에 관한 소개를 하고, 패턴분류에서 목적 계획법의 접근은 3절에서 논의한다. 소프트 마진을 이용한 서포트 벡터 기계는 4절에서 논의하고, 본 논문의 제안방법으로 다목적 및 목적 계획법에 의한 서포트 벡터 기계를 소개하고자 한다. 6절은 예제를 이용한 비교 분석을 하고 7절은 결론이다.

2. 서포트 벡터 기계

서포트 벡터 기계는 Vapnik과 Cortes(1995)등에 의하여 개발 되었고 중요한 특징들은 다음과 같다. 서포트 벡터 기계는 특징공간(feature space)에서 최대 여유도를 가지는 선형 분류기에 기초하여, 특징공간에서 내적(inner product)을 계산함에 있어 커널의 형태들이 유지되는 점을 이용한다. 서포트 벡터 기계는 VC(Vapnik-Chervonenkis)차원을 이용하여 일반화 능력을 평가할 수 있다.

훈련 자료집합 X를 선형으로 분리할 수 없는 경우에 ℓ차원의 원자료 X를 어떤 비선형 함수 $\phi(\cdot)$에 의하여 특징공간 Z로 변환한다. 특징공간의 차원이 증가할수록 변환된 자료집합은 선형으로 분리될 가능성이 커진다(Schölkopf와 Smola, 2002). 비선형 함수를 $\boldsymbol{z}_i = \phi(\boldsymbol{x}_i)$ 라 하면, 최대 여유도를 가지는 분리 초평면(separating hyperplane)은 최소 내부편차(interior deviation)를 갖는 점에서 정규화 즉, $\boldsymbol{w}^T\boldsymbol{z} + w_0 = \pm 1$ 문제를 해결함으로서 얻게된다.

$$\begin{array}{ll} \text{Minimize} & ||\boldsymbol{w}|| \\ \text{subject to} & y_i\left(\boldsymbol{w}^T\boldsymbol{z}_i + w_0\right) \geq 1, \; i = 1, \ldots, \ell. \end{array} \tag{2.1}$$

여기서 $||\cdot||$는 일반적인 놈(norm)을 나타내고 2-놈 $||\boldsymbol{w}||_2$을 사용하면 2차계획문제로 축소되고, 반면에 $||\boldsymbol{w}||_1$이나 $||\boldsymbol{w}||_\infty$를 사용하면 선형계획문제로 줄일 수 있다(Mangararian, 1999).

식 (2.1)에 2-놈을 가지는 쌍대문제(dual problem)는 다음과 같이 나타낼 수 있다.

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j) \\
\text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i y_i = 0, \\
& \alpha_i \geq 0, \ i = 1, \ldots, \ell.
\end{aligned} \tag{2.2}
$$

여기서 α_i는 라그랑쥬 승수(Lagrange multiplier)이고 $\phi(\cdot)^T$는 비선형 함수 $\phi(\cdot)$의 전치(transpose)이다.

커널함수 $K(\boldsymbol{x}, \boldsymbol{z}) = \phi(\boldsymbol{x})^T \phi(\boldsymbol{z})$를 이용하면 식 (2.2)는 아래의 식 (2.3)과 같이 나타낼 수 있다.

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \\
\text{subject to} \quad & \sum_{i=1}^{\ell} \alpha_i y_i = 0, \\
& \alpha_i \geq 0, \ i = 1, \ldots, \ell.
\end{aligned} \tag{2.3}
$$

여러 종류의 커널 함수들이 소개되어 왔는데 그중에서 특히 q차 다항커널

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x}^T \boldsymbol{x}' + 1)^q$$

함수와 본 논문의 수치 실험에서 사용되는 가우지안(Gaussian) 커널함수들이 보편적으로 가장 많이 사용된다.

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{x}'||^2}{r^2}\right)$$

3. 목적 계획법을 이용한 패턴 분류

Freed와 Glover(1981)는 목적 계획법을 사용하여 가능한 오분류되는 자료들을 적게 하면서 두 집단을 분류하는 초평면을 구할 수 있는 방법을 제안하였다. 어떤 점 $\boldsymbol{x}_i$가 분리 초평면에서 오분류된 편차를 *외적편차*(exterior deviation)라 하고 ξ_i로 나타낸다. 반면 $\boldsymbol{x}_i$가 초평면에서 정확하게 분류된 것을 *내적편차*(interior deviation)라 하고 η_i로 표기한다. 이러한 접근법의 중요한 목적들은 아래와 같다:

(i) 최대의 외적편차를 최소화 한다 (가능한 많이 오차를 감소시킨다).

(ii) 최소의 내적편차를 최대화 한다 (여유도를 최대화 한다).

(iii) 내적편차의 가중 합을 최대화 한다.

(iv) 외적편차의 가중 합을 최소화 한다.

많은 모형들이 제안되었지만 단지 (iii)과 (iv)만을 고려한 선형 목적 계획법에 따르면 아래의 식 (3.1)과 같이 쓸 수 있다.

$$\begin{array}{lll} \text{Minimize} & \displaystyle\sum_{i=1}^{\ell}(h_i\xi_i - k_i\eta_i) & \\ \text{subject to} & y_i(\boldsymbol{x}_i^T\boldsymbol{w} + w_0) = \eta_i - \xi_i, & (3.1) \\ & \xi_i,\ \eta_i \geq 0,\ i = 1, \ldots, \ell. & \end{array}$$

여기서 h_i와 k_i는 양의 상수이다. $\mathcal{D}_+$와 $\mathcal{D}_-$를 $\boldsymbol{x}_i \in \mathcal{D}_+$일 때 $y_i = +1$ 그리고 $\boldsymbol{x}_i \in \mathcal{D}_-$일 때 $y_i = -1$로 분류되는 자료집합이라 정의하자. $\boldsymbol{x}_i \in \mathcal{D}_+$일 때 $\boldsymbol{x}_i^T\boldsymbol{w} + w_0 = \eta_i - \xi_i$이고 $\boldsymbol{x}_i \in \mathcal{D}_-$일 때 $\boldsymbol{x}_i^T\boldsymbol{w} + w_0 = -\eta_i + \xi_i$인데 아래와 같이 하나의 방정식으로 표현할 수 있다.

$$y_i(\boldsymbol{x}_i^T\boldsymbol{w} + w_0) = \eta_i - \xi_i.$$

ξ_i와 η_i가 각각 외적편차와 내적편차의 의미를 가지기 위하여 모든 $i = 1, \ldots, \ell$에 대하여 $\xi_i\eta_i = 0$의 조건이 성립해야한다.

Lemma 3.1 (Sawaragi *et al.*, 1985) 모든 $i = 1, \ldots, \ell$에 대하여, 만약 $h_i > k_i$ 이면 식 (3.1)의 모든 해에서 $\xi_i\eta_i = 0$는 성립한다.

위의 공식으로 얻어진 해는 $\boldsymbol{w} = 0$와 같은 비적합 해(unaccepted solution)나 비유계인 해(unbounded solution)를 얻을 수 있음에 주의해야 한다. 그러므로 선형 분류기들에 목적계획법을 이용한 선형분류기의 추출하기 위한 접근법은 유계인 비자명 최적해(nontrivial solution)를 얻기 위하여 $\boldsymbol{w}$에 적절한 정규조건을 넣어야 한다. 그러한 정규조건들 중의 하나는 $\|\boldsymbol{w}\| = 1$이다.

4. 소프트 마진 서포트 벡터 기계

두 자료집합 $\mathcal{D}_+$와 $\mathcal{D}_-$를 완전하게 분리하는 방법인 하드 마진(hard margin)방법은 과적합(overlearning)을 하는 경향이 있다. 이것은 잡음에 의하여 쉽게 영향을 받을 수 있다는 것을 의미한다. 이런 어려움을 극복하기 위하여 소프트 마진(soft margin)방법이 도입되었다. 소프트 마진 방법은 slack 변수들(외적 편차) ξ_i $(i = 1, \ldots, \ell)$에 의해 나타내어지는 약간의 오차를 허용하는 방법이다. 가중치 함수(weight function) $\|\boldsymbol{w}\|$ 최소화와 외적편차의 합 $\sum_{i=1}^{\ell}\xi_i$ 최소화를 이루기 위하여 상호조정(trade-off) 모수 C를 이용하면 소프트 마진방법

에 의한 식을 아래의 (4.1)과 같이 얻을 수 있다:

$$
\begin{aligned}
\text{Minimize} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{\ell}\xi_i \\
\text{subject to} \quad & y_i\left(\boldsymbol{w}^T\boldsymbol{z}_i + w_0\right) \geq 1 - \xi_i, \\
& \xi_i \geq 0, \ \ i = 1, \ldots, \ell.
\end{aligned}
\tag{4.1}
$$

식 (4.1)의 쌍대문제에서 커널을 이용하면 식 (4.2)와 같이 표현할 수 있다.

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{i=1}^{\ell}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{\ell}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \\
\text{subject to} \quad & \sum_{i=1}^{\ell}\alpha_i y_i = 0, \\
& 0 \leq \alpha_i \leq C, \ i = 1, \ldots, \ell.
\end{aligned}
\tag{4.2}
$$

소프트 마진 방법은 선형 분류기에 대한 목적계획법의 아이디어와 동일하게 간주할 수 있다는 사실에 기초하여 Benett와 Mangasarian(1992)은 다중-표면방법(Multi-surface Method)의 확장에 대한 논의를 하였다. 외부편차뿐 만이 아니라 내부편차도 서포트 벡터 기계에서 고려할 수 있다. Nakayama와 Asada(2001)와 Asada와 Nakayama(2003)는 다목적 최적화와 목적 계획법들의 접근법을 서포트 벡터 기계에 적용하는 부분에 대하여 언급하였다. 목적 계획법을 이용한 방법을 적용할 때 비적합 해를 극복하기 위하여 적절한 정규 조건의 필요성을 3절에서 지적하였다.

Glover(1990)는 비적합 해를 피하기 위한 필요충분조건을 식 (4.3)과 같이 제안했다.

$$
\left(-l_{\mathcal{D}_+}\sum_{i\in I_{\mathcal{D}_-}}\boldsymbol{x}_i + l_{\mathcal{D}_-}\sum_{i\in I_{\mathcal{D}_+}}\boldsymbol{x}_i\right)^T \boldsymbol{w} = 1. \tag{4.3}
$$

여기서 $l_{\mathcal{D}_+}$와 $l_{\mathcal{D}_-}$는 각각의 범주 $\mathcal{D}_+, \mathcal{D}_-$에 속하는 자료의 수이다. Erenguc와 Koehler(1990)는 기하학적으로 $l_{\mathcal{D}_+}$와 $l_{\mathcal{D}_-}$에 대하여 자료의 중심을 통과하는 두 개의 분리 초평면들 사이의 거리를 의미하고 그 크기는 $l_{\mathcal{D}_+} \times l_{\mathcal{D}_-}$로 나타내었다.

3절에서 서술한 목적 계획법의 (ii)와 (iii)의 목적들을 고려하여 Schölkopf와 Smola(1998)는 ν-서포트 벡터 기계를 제안한 라그랑쥬 쌍대문제는 식 (4.4)와 같이 표현된다.

$$
\begin{aligned}
\text{Maximize} \quad & -\frac{1}{2}\sum_{i,j=1}^{\ell} y_i y_j \alpha_i \alpha_j K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \\
\text{subject to} \quad & \sum_{i=1}^{\ell} y_i\alpha_i = 0, \quad \sum_{i=1}^{\ell}\alpha_i \geq \nu, \\
& 0 \leq \alpha_i \leq \frac{1}{\ell}, \ \ i = 1, \ldots, \ell.
\end{aligned}
\tag{4.4}
$$

5. 다목적 계획법과 목적 계획법에 의한 서포트 벡터 기계의 확장

오분류된 점들을 나타내는 slack 변수 ξ_i들과 정분류된 점들을 나타내는 surplus 변수 η_i를을 동시에 고려하는 서포트 벡터 기계의 두 가지 알고리즘을 제안한다.

5.1. Total Margin Algorithm

오분류된 slack 변수들의 최소화와 정분류된 surplus변수들의 최대화를 위하여 아래의 식 (5.1)을 고려하자.

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C_1 \sum_{i=1}^{\ell} \xi_i - C_2 \sum_{i=1}^{\ell} \eta_i \\ \text{subject to} \quad & y_i \left(\boldsymbol{w}^T \boldsymbol{z}_i + w_0\right) \ge 1 - \xi_i + \eta_i, \\ & \xi_i \ge 0, \ \ \eta_i \ge 0, \ \ i = 1, \ldots, \ell. \end{aligned} \tag{5.1}$$

여기서 C_1와 C_2는 ξ_i와 η_i중에서 적어도 하나는 0이라는 것을 보장하기 위하여 $C_1 > C_2$의 조건을 만족하는 값들을 선택해야한다. 식 (5.1)의 라그랑쥬 함수는 다음과 같다.

$$\begin{aligned} L(\boldsymbol{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = {} & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C_1 \sum_{i=1}^{\ell} \xi_i - C_2 \sum_{i=1}^{\ell} \eta_i \\ & - \sum_{i=1}^{\ell} \alpha_i \left[y_i \left(\boldsymbol{w}^T \boldsymbol{z}_i + w_0\right) - 1 + \xi_i - \eta_i\right] - \sum_{i=1}^{\ell} \beta_i \xi_i - \sum_{i=1}^{\ell} \gamma_i \eta_i. \end{aligned}$$

여기서 $\alpha_i \ge 0, \ \beta_i \ge 0$ 그리고 $\gamma_i \ge 0$이다. $\boldsymbol{w}, \ w_0, \ \boldsymbol{\xi}$ 그리고 $\boldsymbol{\eta}$ 에 대하여 라그랑쥬 함수를 미분하여 얻은 정상조건들을 라그랑쥬 함수 L에 대입하고 커널을 이용하여 나타내면 식 (5.2)의 쌍대 최적화 문제를 얻을 수 있다.

$$\begin{aligned} \text{Maximize} \quad & \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} y_i y_j \alpha_i \alpha_j K\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \\ \text{subject to} \quad & \sum_{i=1}^{\ell} y_i \alpha_i = 0, \\ & C_2 \le \alpha_i \le C_1, \ \ i = 1, \ldots, \ell. \end{aligned} \tag{5.2}$$

$\boldsymbol{\alpha}^*$를 식 (5.2)의 최적해라 하면 분류기는 다음과 같다.

$$f(\phi(\boldsymbol{x})) = \sum_{i=1}^{\ell} \alpha_i^* y_i K\left(\boldsymbol{x}, \boldsymbol{x}_i\right) + w_0.$$

분계점(threshold) w_0^*는 아래와 같이 얻을 수 있다. 여기서 $C_2 < \alpha_j^* < C_1$와 $y_j = +1$를 만족하는 $\boldsymbol{x}_j$의 갯수를 n_+라 하고, $C_2 < \alpha_j^* < C_1$와 $y_j = -1$를 만족하는 $\boldsymbol{x}_j$의 갯수를 n_-라 하자.

Karush-Kuhn-Tucker(1951)의 보충(complementary)조건에 의하면 만약 $C_2 < \alpha_j^* < C_1$이면 $\beta_j > 0$와 $\gamma_j > 0$이 된다. 다시 말해 $\xi_j = \eta_j = 0$를 의미한다.

$$w_0{}^* = \frac{1}{n_+ + n_-}\left((n_+ - n_-) - \sum_{j=1}^{n_+ + n_-}\sum_{i=1}^{\ell} y_i \alpha_i^* K(\boldsymbol{x}_i, \boldsymbol{x}_j)\right).$$

5.2. μ- 서포트 벡터 기계

오분류된 자료들 중에서 가장 크게 오분류된 slack 변수를 최소화하고 surplus의 합을 최대화하면, 오분류된 자료점들과 분리 초평면 사이의 최대 거리를 나타내는 새로운 변수 σ를 도입하면 식 (5.3)을 얻게된다.

$$\begin{aligned} &\text{Minimize} && \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \mu\sigma - \frac{1}{\ell}\sum_{i=1}^{\ell}\eta_i \\ &\text{subject to} && y_i\left(\boldsymbol{w}^T \boldsymbol{z}_i + w_0\right) \geq \eta_i - \sigma, \\ &&& \sigma \geq 0,\ \ \eta_i \geq 0,\ \ i = 1, \ldots, \ell. \end{aligned} \tag{5.3}$$

여기서 μ는 σ와 η_i들의 합 사이의 상호조정(trade-off)을 반영하는 모수이다.

식 (5.3)에 대한 라그랑쥬 함수는 다음과 같다.

$$\begin{aligned} L(\boldsymbol{w}, w_0, \boldsymbol{\eta}, \sigma, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = &\frac{1}{2}\|\boldsymbol{w}\|_2^2 + \mu\sigma - \frac{1}{\ell}\sum_{i=1}^{\ell}\eta_i \\ &- \sum_{i=1}^{\ell}\alpha_i\left[y_i\left(\boldsymbol{w}^T \boldsymbol{z}_i + w_0\right) - \eta_i + \sigma\right] - \sum_{i=1}^{\ell}\beta_i\eta_i - \gamma\sigma, \end{aligned}$$

여기서 $\alpha_i \geq 0$, $\beta_i \geq 0$그리고 $\gamma \geq 0$ 이다. 위의 라그랑쥬 함수를 결합하면 다음의 식 (5.4)의 쌍대 최적화 문제인 μ–서포트 벡터 기계를 얻는다.

$$\begin{aligned} &\text{Maximize} && -\frac{1}{2}\sum_{i,j=1}^{\ell}\alpha_i\alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &\text{subject to} && \sum_{i=1}^{\ell}\alpha_i y_i = 0, \quad \sum_{i=1}^{\ell}\alpha_i \leq \mu, \\ &&& \alpha_i \geq \frac{1}{\ell},\ \ i = 1, \ldots, \ell. \end{aligned} \tag{5.4}$$

$\boldsymbol{\alpha}^*$를 식 (5.4)의 최적해라 하자. 분계점 w_0^*를 계산하기 위하여 $\frac{1}{\ell} < \alpha_j^*$를 만족하는 동일한 크기의 n을 $\boldsymbol{x}_j$인 집합 $\mathcal{D}_+$에서 얻는다. Karush-Kuhn-Tucker의 보충조건에서, 만약 $\frac{1}{\ell} < \alpha_j^*$이고, $\beta_j > 0$이면 $\eta_j = 0$임을 의미하므로 아래의 w_0^*를 얻는다.

$$w_0{}^* = -\frac{1}{2n}\sum_{\boldsymbol{x}_j \in \mathcal{D}_+}\sum_{i=1}^{\ell}\alpha_i^* y_i K(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

6. 수치 예제들

본 논문에서 논의한 방법의 수행능력을 조사하기 위하여 세 개의 자료집합들에 대한 정분류율의 결과를 비교하고자 한다.

첫 번째 자료는 Cleveland Clinic Foundation의 심장질환 관한 자료이다. 이 자료는 13개의 독립변수들과 303명의 환자들의 관측자료이다. 두 번째 자료는 BUPA Medical Research Ltd.의 간장질환(liver-disorder) 자료로 7개의 독립변수와 345명의 환자를 측정한 것이다. 세 번째 자료는 당뇨병의 원인으로 생각되는 8개의 독립변수들에 기초하여 환자들이 세계보건기구에서 정한 기준에 의하여 당뇨병의 징후를 보이는가의 여부에 따라서 두 집단으로 나누었고 이 자료의 대상은 Pima Indian족의 21세 이상의 가임 여성 768명을 대상으로 조사한 자료이다. 위의 자료들은 http://www.ics.uci.edu/~mlearn/MLSummary.html에서 다운로드 받을 수 있다. 첫 번째와 두 번째 자료에 대하여, 교차타당성(cross validation) 방법을

표 6.1: 식 (4.2) 알고리즘에 의한 정분류율

(a) 심장병 자료의 분류 결과

C	1	10	100
AVE	78.67	76.67	75.67
STDV	7.73	6.67	8.02
#. SV	95.49	95.60	95.60

(b) 간장질환 자료의 분류 결과

C	1	10	100
AVE	63.82	63.82	65.88
STDV	9.10	8.09	6.53
#. SV	82.77	71.58	62.67

사용하였다. 이것은 원래의 자료집합에서 무작위로 선택된 90%의 자료를 훈련 자료로 쓰고 나머지 10% 자료를 테스트 자료로 10번의 반복 시행의 평균을 구하였다. 표 6.1부터 표 6.4까지는 정분류율의 평균(AVE)과 표준편차(STDV)를 나타내고 기존의 (4.2)와 (4.4) 알고리즘들과 본 논문에서 제안한 (5.2)와 (5.4) 알고리즘을 이용하여 서포트 벡터의 갯수(SV)를 나타내었다. 표 6.4에서는 모수들의 값의 변화에도 불구하고 정분류율이 변하지 않기 때문에 모수들의 값을 표현하지 않았다.

심장병 자료의 경우 식 (4.2)의 알고리즘에 의한 정분류율을 나타내는 표 6.1에서 $C = 1$인 경우의 정분류율이 78.67이고 식 (4.4)를 적용한 표 6.2에서 $\nu = 0.5$, $\nu = 0.6$인 경우에 79.00의 정분류율을 얻었으나, 식 (5.2)의 알고리즘을 적용한 정분류율은 $C_1 = 100, C_2 = 1$인 경우에 80.67이고 식 (5.4)를 적용한 경우는 가장 높은 81.00의 정분류율을 나타내고 있다. 간장질환 자료의 경우도 식 (4.2)와 식 (4.4)의 알고리즘을 적용하여 얻어진 정분류율은 각각 $C = 100$인 경우 65.88와 $\nu = 0.5$의 경우 62.64라는 정분류율을 얻었다. 반면 본 논문에서 제안하는 식 (5.2)와 식 (5.4)의 알고리즘을 적용한 정분류율은 각각 $C_1 = 1, C_2 = 0.001$일 때 68.82와 70.30을 얻었다.

따라서 심장병 자료와 간장질환 자료의 경우는 표 6.1부터 표 6.4까지에서 보는 바와 같이 가장 높은 정분류율을 보이는 경우는 식 (4.2), 식 (4.4)에 의한 것 보다는 식 (5.2)나 식 (5.4)의 알고리즘에 기초한 결과라는 것을 나타내고 있다.

표 6.2: 식 (4.4) 알고리즘에 의한 정분류율

(a) 심장병 자료의 분류 결과

ν	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
AVE	76.00	76.00	77.67	78.67	79.00	79.00	78.33	76.00
STDV	4.92	4.10	3.53	4.50	5.22	3.87	5.72	9.79
#. SV	95.31	95.38	95.49	95.35	95.42	95.42	95.42	95.49

(b) 간장질환 자료의 분류 결과

ν	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
AVE	61.76	62.94	62.35	62.65	62.94	61.76	62.35	60.88
STDV	10.47	8.68	7.04	7.34	6.82	6.20	7.94	6.05
#. SV	64.50	61.22	69.77	73.34	76.40	78.97	82.57	87.20

표 6.3: 식 (5.2) 알고리즘에 의한 정분류율

(a) 심장병 자료의 분류 결과

	$C_1 = 1$			$C_1 = 10$			$C_1 = 100$		
C_2	0.001	0.01	0.1	0.01	0.1	1	0.1	1	10
AVE	78.67	78.67	78.67	76.67	77.00	78.67	75.67	80.67	78.33
STDV	7.73	7.73	7.73	6.70	6.75	6.67	8.02	8.13	7.24
#. SV	100	100	100	100	100	100	100	100	100

(b) 간장질환 자료의 분류 결과

	$C_1 = 1$			$C_1 = 10$			$C_1 = 100$		
C_2	0.001	0.01	0.1	0.01	0.1	1	0.1	1	10
AVE	68.82	63.82	64.41	63.82	63.82	67.35	65.59	68.24	66.67
STDV	9.10	9.10	8.02	8.09	8.09	8.26	6.51	6.17	5.89
#. SV	100	100	100	100	100	100	100	100	100

표 6.4: 식 (5.4) 알고리즘에 의한 정분류율

(a) 심장병 자료의 분류 결과

AVE	81.00
STDV	6.49
#. SV	100

(b) 간장질환 자료의 분류 결과

AVE	70.30
STDV	5.79
#. SV	100

표 6.5: 식 (4.2) 알고리즘에 의한 Pima Indian 자료의 정분류율

	$C = 0.01$					
	훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	65.24	0.00	100.00	64.78	0.00	100.00
STDV	0.76	0.00	0.00	1.77	0.00	0.00
	$C = 0.1$					
	훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	65.24	0.00	100.00	64.78	0.00	100.00
STDV	0.76	0.00	0.00	1.77	0.00	0.00
	$C = 1.0$					
	훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	91.56	80.36	97.52	73.57	49.81	86.54
STDV	0.50	1.86	0.58	2.79	6.14	2.84
	$C = 10$					
	훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	99.29	98.34	99.80	70.39	53.97	79.33
STDV	0.31	0.74	0.22	3.06	5.06	3.37
	$C = 100$					
	훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	100.00	100.00	100.00	69.91	52.74	79.26
STDV	0.00	0.00	0.00	2.93	4.60	3.98
	$C = 1000$					
	훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	100.00	100.00	100.00	69.91	52.74	79.26
STDV	0.00	0.00	0.00	2.93	4.60	3.98

표 6.6: 식 (5.2) 알고리즘에 의한 Pima Indian 자료의 정분류율

	$C_1 = 1$																	
	$C_2 = 0.001$						$C_2 = 0.01$						$C_2 = 0.1$					
	훈련집합			테스트 집합			훈련집합			테스트 집합			훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	91.86	82.23	96.98	73.83	55.63	83.79	92.04	82.82	96.95	74.04	56.38	83.73	91.97	90.12	92.96	73.52	76.01	72.14
STDV	0.51	1.99	0.61	2.43	5.35	3.26	0.51	1.78	0.59	2.28	5.23	3.28	0.84	2.24	1.88	2.08	3.46	4.46
	$C_1 = 10$																	
	$C_2 = 0.01$						$C_2 = 0.1$						$C_2 = 1$					
	훈련집합			테스트 집합			훈련집합			테스트 집합			훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	99.24	98.56	99.60	70.04	62.65	74.02	99.24	98.78	99.49	69.30	65.25	71.48	65.24	98.72	47.40	61.91	93.60	44.69
STDV	0.34	0.72	0.26	2.49	5.38	3.67	0.36	0.83	0.25	2.24	4.79	3.83	1.67	0.36	2.62	4.11	1.73	6.57
	$C_1 = 100$																	
	$C_2 = 0.1$						$C_2 = 1$						$C_2 = 10$					
	훈련집합			테스트 집합			훈련집합			테스트 집합			훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	100.00	100.00	100.00	68.13	66.28	69.14	69.65	100.00	53.48	63.17	90.92	48.12	64.26	98.98	45.75	61.39	94.20	43.54
STDV	0.00	0.00	0.00	2.94	4.12	5.32	3.22	0.00	4.82	4.04	4.03	8.18	1.80	0.38	2.64	4.23	1.41	6.73
	$C_1 = 1000$																	
	$C_2 = 1$						$C_2 = 10$						$C_2 = 100$					
	훈련집합			테스트 집합			훈련집합			테스트 집합			훈련집합			테스트 집합		
	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$	정분류율	$\mathcal{D}_+$	$\mathcal{D}_-$
AVE	69.65	100.00	53.48	63.17	90.92	48.12	66.25	100.00	48.27	61.78	93.74	44.40	64.26	98.98	45.75	61.35	94.31	43.40
STDV	3.22	0.00	4.82	4.04	4.03	8.18	2.47	0.00	3.65	3.83	2.21	6.70	1.76	0.38	2.55	4.09	1.34	6.58

세 번째 자료인 Pima Indian 자료는 두 집합의 자료들의 개수들이 편중된(unbalanced) 자료집합이다. $\mathcal{D}_+$집합(당뇨병에 대하여 양성 반응자들)은 268명이 있고, 반면에 $\mathcal{D}_-$집합(당뇨병에 대하여 음성 반응자들)은 500명이 있다. 이 자료집합의 경우는 훈련집합의 원소로서 전체 집합에서 무작위로 70%를 추출하고 나머지 30%를 테스트 자료로 선택하였다. 표 6.5와 표 6.6에서 알 수 있는 바와 같이 (4.2)와 (5.2) 알고리즘들의 분류 결과를 비교하였다. 식 (4.2)의 알고리즘을 적용한 경우 테스트 집합의 정분류율은 $C = 1.0$일 때 73.57이라는 정분류율을 얻었으나 제안된 식 (5.2)를 적용하면 $C_1 = 1, C_2 = 0.001$와 $C_1 = 1, C_2 = 0.01$인 경우에 각각 73.88와 74.04의 정분류율을 얻었다. Pima Indian 당뇨병 자료와 같은 불균형 자료에 대하여 식 (5.2)알고리즘은 원소의 개수가 적은 집합의 원소의 정분류율에 있어서 식 (4.2)알고리즘의 정분류율보다 좋은 수행능력을 보이고 있다.

7. 결론

6절의 수치 실험들을 통하여 비록 분류 결과들이 모수들의 값에 의존적이지만, Schölkopf와 Smola(1998)가 개발한 식 (4.4)와 본 논문이 제안한 식 (5.2), 그리고 식 (5.4)와 같은 다목적 계획법과 목적 계획법을 사용한 서포트 벡터 기계 알고리즘들은 Cortes과 Vapnik(1995)에 의한 식 (4.2)를 이용한 알고리즘에 비하여 정분류율을 비교하면 상대적으로 더 높은 정분류율를 얻을 수 있다는 가능성을 제시하고 있다. 본 논문의 다목적 계획법과 목적 계획법을 이용한 형태의 서포트 벡터 기계 알고리즘들은 기존의 제안된 알고리즘을 이용한 경우보다 정분류율을 개선시킬 수 있음을 알 수 있었다.

식 (5.2)의 알고리즘은 C_1과 C_2의 값에 의존적이기는하나 작은 자료 집합의 정분류율에 있어서 식 (4.2)알고리즘의 정분류율보다 좋은 수행능력을 보인다는 것을 확인할 수 있었다. 최적의 C_1과 C_2를 찾기 위한 노력은 또 하나의 과제로 보인다. 한편 여러 종류의 자료에 대한 다양한 실험을 통하여 식 (5.4)알고리즘에서 모수의 어떤 값에 대해서는 비적합해들이 나타날 수도 있다는 사실을 알 수 있었다.

참고문헌

Asada, T. and Nakayama, H. (2003). Support vector machines using multi objective programming, in T. Tanino, T. Tanaka and M. Inuiguchi (eds.). *Multi-objective Programming and Goal Programming*, 93-98.

Bennett, K.P. and Mangasarian, O.L. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, **1**, 23-34.

Cavalier, T.M., Ignizio, J.P. and Soyster, A.L. (1989). Discriminant analysis via mathematical programming: certain problems and their causes. *Computers and Operations Research*, **16**, 353-362.

Cherkassky, V. and Mulier, F. (1998). *Learning from Data Concepts, Theory, and Methods*, John Wiley & Sons.

Cortes, C. and Vapnik, V. (1995). Support vector networks, *Machine Learning*, **20**, 273–297.

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

Erenguc, S.S. and Koehler, G.J. (1990). Survey of mathematical programming models and experimental results for linear discriminant analysis, *Managerial and Decision Economics*, **11**, 215-225.

Freed, N. and Glover, F. (1981). Simple but powerful goal programming models for discriminant problems, *European Journal of Operational Research*, **7**, 44-60.

Glover, F. (1990). Improved linear programming models for discriminant analysis, *Decision Sciences*, **21**, 771-785.

Kuhn, H.W. and Tucker, A.W. (1951). Nonlinear programming, In *Proceedings of the* 2^{nd} *Berkeley Symposium on Mathematical Statistics and Probabilistic*, University of California Press, 481-492.

Mangasarian, O.L. (1968). Multi surface method of pattern separation, *IEEE Transaction on Information Theory*, **IT-14**, 801-807.

Mangasarian, O.L. (1999). Arbitrary-norm separating plane, *Operations Research Letters*, **24**, 15-23.

Marcotte, P. and Savard, G. (1992). Novel approaches to the discrimination problem, *ZOR-Methods and Models of Operations Research*, **36**, 517-545.

Nakayama, H. and Asada, T. (2001). Support vector machines formulated as multi objective linear programming, *Proceedings of ICOTA 2001*, **3**, 1171-1178.

Novikoff, A.B. (1962). On the convergence proofs on perceptrons, In*Symposium on the Mathematical Theory of Automata*, **12**, 615–622. Polytechnic Institute of Brooklyn.

Sawaragi, Y., Nakayama, H. and Tanino, T. (1985). *Theory of Multiobjective Optimization*, Academic Press.

Schölkopf, B. and Smola, A.J. (1998). New support vector algorithms, *NeuroCOLT2 Technical Report Series*, NC2-TR-1998-031.

Yoon, M., Yun, Y.B. and Nakayama, H. (2004). Total margin algorithms in support vector machines, *IEICE Transactions on Information and Systems*, **87-D**, 1223–1230.

http://www.ics.uci.edu/m̃learn/MLSummary.html.

[2004년 7월 접수, 2005년 5월 채택]

Study on Support Vector Machines Using Mathematical Programming*

Min Yoon [1] Hakbae Lee [2]

ABSTRACT

Machine learning has been extensively studied in recent years as effective tools in pattern classification problem. Although there have been several approaches to machine learning, we focus on the mathematical programming (in particular, multi-objective and goal programming; MOP/GP) approaches in this paper. Among them, Support Vector Machine (SVM) is gaining much popularity recently. In pattern classification problem with two class sets, the idea is to find a maximal margin separating hyperplane which gives the greatest separation between the classes in a high dimensional feature space. However, the idea of maximal margin separation is not quite new: in 1960's the multi-surface method (MSM) was suggested by Mangasarian. In 1980's, linear classifiers using goal programming were developed extensively. This paper proposes a new family of SVM using MOP/GP techniques, and discusses its effectiveness throughout several numerical experiments.

Keywords: Support vector machine, Maximal margin classifier, Multi-objective optimization, Goal programming.

* This work was supported by the Brain Korea 21 Project in 2004.
This work was supported (in part) by Yonsei University Research Fund of 2004

1) Researcher-in-Charge, Yonsei Institute of Statistics Science, Shinchon-Dong 134 Seodaemun-Gu Seoul, 120-749, Korea
E-mail: myoon@yonsei.ac.kr
2) Assistant Professor, Department of Applied Statistics, Shinchon-Dong 134 Seodaemun-Gu Seoul, 120-749, Korea
E-mail: hblee@yonsei.ac.kr