*Proceedings of Joint Conference of Korean Data And Information Science Society and Korean Data Analysis Society, June 21-22, 2002, pp. 23-28*

# A Recent Development in Support Vector Machine Classification

*Dug Hun Hong*[1], *Changha Hwang*[2], *Eunyoung Na*[3]

## Abstract

Support vector machine(SVM) has been very successful in classification, regression, time series prediction and density estimation. In this paper, we will propose SVM for fuzzy data classification.

Keywords : Fuzzy input, Support vector machine, Classification

## 1. 서론

SVM은 원래 분류(classification)를 위해 Vapnik(1995, 1998)과 공동연구자들에 의해 개발되어 문자인식, 얼굴인식 등의 응용분야에서 좋은 결과를 보여주고 있다. 최근 분류를 위한 SVM 이론이 함수근사로 확장되어 많이 활용되고 있다. SVM은 VC-이론을 이용하여 예측(prediction)을 잘하고 벌칙항(penalty term)을 사용하여 과대적합(overfitting)을 피하려는데 초점을 두고 있다. 여기서 VC는 이론 창시자들의 이름인 Vapnik 과 Chervonenkis 첫 글자를 따서 표현된 것이다. 한편 SVM은 고도의 비선형 구조를 파악할 수 있어, 분류와 관련된 문자인식, 얼굴인식 등의 응용분야에서 많은 성공사례를 보여주었다. 본 논문에서는 퍼지자료의 분류를 SVM을 제안하고 자료분석을 통해 그 성능을 평가한다.

## 2. 퍼지자료의 분류를 위한 SVM

훈련자료 $\{(\mathbf{x}_i, y_i), i=1,\cdots \ell\}$는 퍼지입력벡터 $\mathbf{x}_i = ((m_{x_{i1}}, \alpha_{x_{i1}}), \cdots, (m_{x_{id}}, \alpha_{x_{id}})) \in T(R)^d$인 퍼지수(fuzzy number)이고 $y_i \in \{\pm 1\}$이다. 퍼지분류를 위해 다음과 같표기한다.

$$\underset{\sim}{\alpha} = [\alpha_{11}\ \alpha_{12} \cdots \alpha_{1\ell} : \alpha_{21}\ \alpha_{22} \cdots \alpha_{2\ell}] = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix},$$

$$W = (W_1, \cdots, W_d), \ ||W||^2 = ||m_W||^2 + ||\alpha_W||^2 + ||\alpha_B||^2.$$

---

[1]경북 경산시 하양읍 대구가톨릭대학교 기계자동차공학부 교수
[2]경북 경산시 하양읍 대구가톨릭대학교 정보통계학과 교수
[3]경북 경산시 하양읍 대구가톨릭대학교 전산통계학과 박사과정

여기서 $W_i = (m_W, \alpha_W)$, $m_W = (m_{W_1}, \cdots, m_{W_i})$, $\alpha_W = (\alpha_{W_1}, \cdots, \alpha_{W_{d-1}}, 0)$ 이다. 그러면

$$\begin{aligned} &\text{minimize} \quad \|m_W\|^2 + \frac{1}{2}\|\alpha_W\|^2 + C_1\sum\xi_{1i}^2 + C_2\sum\xi_{2i}^2 \\ &\text{subject to} \quad y_i(<m_W, \mathbb{X}_i> + m_B) \geq 1 - \xi_{1i}, \quad i = 1, \cdots, \ell \\ &\qquad (<m_W, \mathbb{X}_i> + m_B) - y_{i'}(<\alpha_W, |\mathbb{X}_i|> + \alpha_B) \leq \xi_{2i} \\ &\text{where} \quad \xi_{1i} \geq 0 \ \& \ \xi_{2i} \geq 0 \end{aligned} \tag{2-1}$$

가 된다. 이 문제를 최적화하기위해 라그랑즈 함수(Lagrangian fuction)로 표현을 하면

$$\begin{aligned} L(W, B, \xi, \alpha) = &\ \frac{1}{2}\|m_W\|^2 + \frac{1}{4}\|\alpha_W\|^2 + \frac{C_1}{2}\sum\xi_{1i}^2 + \frac{C_2}{2}\sum\xi_{2i}^2 \\ &- \sum\alpha_{1i}[y_i(<m_W, \mathbb{X}_i> + m_B) - 1 + \xi_{1i}] \\ &- \sum\alpha_{2i}[y_i(<\alpha_W, |\mathbb{X}_i|> + \alpha_B) - (<m_W, \mathbb{X}_i> + m_B) + \xi_{2i}] \end{aligned} \tag{2-2}$$

이다. (2-2)식의 라그랑즈 함수를 최소화하는 $m_w$ 와 $\alpha_w$ 를 구하면

$$=> \begin{cases} \frac{\partial L}{\partial m_W} = m_W - \sum y_i\alpha_{1i}\mathbb{X}_i + \sum\alpha_{2i}\mathbb{X}_i = 0 & => \ m_W = \sum(y_i\alpha_{1i} - \alpha_{2i})\mathbb{X}_i \\ \frac{\partial L}{\partial \alpha_W} = \frac{1}{2}\alpha_W - \sum\alpha_{2i}y_i\mathbb{X}_i = 0 & => \ \alpha_W = 2\sum\alpha_{2i}y_i\mathbb{X}_i \\ \frac{\partial L}{\partial \xi_1} = C_1\xi_1 - \alpha_{1i} = 0 & => \xi_{1i} = \frac{\alpha_{i1}}{C_1} \\ \frac{\partial L}{\partial \xi_2} = C_2\xi_2 - \alpha_{2i} = 0 & => \ \xi_{2i} = \frac{\alpha_{2i}}{C_2} \\ \frac{\partial L}{\partial m_B} = -\sum\alpha_{1i}y_i + \sum\alpha_{2i} = 0\,(\text{제약}) & => \ \sum\alpha_{1i}y_i - \sum\alpha_{2i} = 0 \\ \frac{\partial L}{\partial \alpha_B} = \sum\alpha_{2i}y_i = 0 & \end{cases}$$

$m_W = \sum(y_i\alpha_{1i} - \alpha_{2i})\mathbb{X}_i$, $\alpha_W = 2\sum\alpha_{2i}y_i\mathbb{X}_i$ 가 된다. 많이 사용되고 있는 커널함수 중 본 논문에 사용된 커널은 $K(x, y) = (x^t y + 1)^2$ 이며, 다시 (2-2)에 대입하면 다음의 식(2-4)의 최대화 문제가 된다. 아래의 (2-4-1)은 선형인 경우이며 (2-4-2)는 비선형인 경우이다.

$$
\begin{aligned}
L(W,B,\xi,\alpha) = & \frac{1}{2}\sum_{i,j=1}^{\ell}(y_i\alpha_{1i}-\alpha_{2i})(y_j\alpha_{1j}-\alpha_{2j})<\mathbb{X}_i,\mathbb{X}_j> \\
& +\frac{1}{4}\times 2\times 2\sum_{i,j=1}^{\ell}y_i y_j\alpha_{2i}\alpha_{2j}<\mathbb{X}_i,\mathbb{X}_j>+\frac{1}{2C_1}\sum\alpha_{1i}^2+\frac{1}{2C_2}\sum\alpha_{2i}^2 \\
& -\sum y_i\alpha_{1i}\left(\sum_{j=1}^{\ell}(y_j\alpha_{1j}-\alpha_{2j})<\mathbb{X}_i,\mathbb{X}_j>\right)+\sum_i\alpha_{1i}-\frac{1}{C_1}\sum\alpha_{1i}^2 \\
& -\frac{1}{C_2}\sum_{i=1}^{\ell}\alpha_{2i}^2+\sum_{i=1}^{\ell}\alpha_{2i}\left(\sum_{j=1}^{\ell}(y_j\alpha_{1j}-\alpha_{2j})<\mathbb{X}_i,\mathbb{X}_j>\right) \\
& -\sum_{i=1}^{\ell}y_i\alpha_{2i}\left(2\sum_{j=1}^{\ell}\alpha_{2j}y_j<\mathbb{X}_i,\mathbb{X}_j>-\sum_{i=1}^{\ell}y_i\alpha_{2i}\right)
\end{aligned}
\tag{2-3}
$$

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i=1}^{\ell}\alpha_{1i}-\frac{1}{2}\sum_{i,j=1}^{\ell}(y_i\alpha_{1i}-\alpha_{2i})(y_j\alpha_{1j}-\alpha_{2j})<\mathbb{X}_i,\mathbb{X}_j> \\
& -\sum_{i,j=1}^{\ell}y_i y_j\alpha_{2i}\alpha_{2j}<\mathbb{X}_i,\mathbb{X}_j>-\frac{1}{2C_1}\sum\alpha_{1i}^2-\frac{1}{2C_2}\sum\alpha_{2i}^2
\end{aligned}
\tag{2-4-1}
$$

Kaursh-Kuhn-Tucker **조건을 이용하여**

$$For \ \ \alpha_{1i}\neq 0,\ \alpha_{1i}[y_i(<m_W,\mathbb{X}>+m_B)-1+\xi_{1i}]=0$$

$$For \ \ \alpha_{2i}\neq 0,\ \alpha_{2i}[\xi_{2i}-<m_W,\mathbb{X}>-m_B)+y_i(<\alpha_W,\mathbb{X}_i>+\alpha_B)=0$$

**로부터** $m_B, \alpha_B$ **를 구한다. 비선형인경우는**

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i=1}^{\ell}\alpha_{1i}-\frac{1}{2}\sum_{i,j=1}^{\ell}(y_i\alpha_{1i}-\alpha_{2i})(y_j\alpha_{1j}-\alpha_{2j})K<\mathbb{X}_i,\mathbb{X}_j> \\
& -\sum_{i,j=1}^{\ell}y_i y_j\alpha_{2i}\alpha_{2j}K<\mathbb{X}_i,\mathbb{X}_j>-\frac{1}{2C_1}\sum\alpha_{1i}^2-\frac{1}{2C_2}\sum\alpha_{2i}^2
\end{aligned}
\tag{2-4-2}
$$

**이다. 이제** $f(x)$ **를 다음과 같이 표기한다.**

$$
\begin{aligned}
f(x) &= (m_{f(x)},\ \alpha_{f(x)}) \\
&= \left(\sum_{i=1}^{\ell}(y_i\alpha_{1i}-\alpha_{2i})K(x_i,x)+m_B^*,\ 2\sum_{i=1}^{\ell}\alpha_{2i}y_iK(x_i,x)+\alpha_B^*\right)
\end{aligned}
$$

**여기서 먼저,** $m_B^*$ **다음 관계식**

$$y_i m_{f(x_i)} = 1-\frac{\alpha_{1i}^*}{C_1} \ \text{ for any i with } \ \alpha_{1i}^*\neq 0$$

$$y_i\alpha_{f(x)} = m_{f(x_i)}-\frac{\alpha_{2i}^*}{C_2} \ \text{ for any i with } \ \alpha_{2i}^*\neq 0$$

으로부터 구해지며

$$m_B^* = \frac{\left\{1 - \frac{\alpha_{1i}^*}{C_1}\right\}}{y_i} - \sum_{j=1}^{\ell}(y_j\alpha_{1j} - \alpha_{2j})K(x_j, x_i)$$

가 된다. 그리고, 0이 아닌 임의의 양의 SVM에 의한 값을 대입하여 다음과 같이 $\alpha_B^*$를 구한다.

$$\alpha_B^* = \frac{\left\{\left(\sum_{k=1}^{\ell}(y_k\alpha_{1k} - \alpha_{2k})K(x_k, x_i) + m_B^*\right) - \frac{\alpha_{2i}^*}{C_2}\right\}}{y_i} - 2\sum_{j=1}^{\ell}\alpha_{2j}y_jK(x_j, x_i)$$

다음 절에서는 예제를 통하여 퍼지자료의 분류를 위해 제안된 SVM 알고리즘의 성능을 연구하려 한다.

## 3. 모의실험

제안된 알고리즘의 성능을 분석하기위해 Fujioka & Tanaka(1993)의 데이터를 예로 들어 설명하기로 한다. 퍼지수(fuzzy number)의 “small"과 ”very large"의 소속함수(membership function)는 그림 1에 나타나 있으며 [0,20]은 패턴공간으로 지금부터 “small", "very large", 그리고 [0,20]를 각각 S, VL, U로 나타내도록 한다. 여기서 퍼지데이터(fuzzy data)는 퍼지수의 확률평균값(possibilistic mean value of fuzzy numbers)에 의해 비퍼지화(defuzzification)하여 일반데이터(crisp data)로 변경하였다.

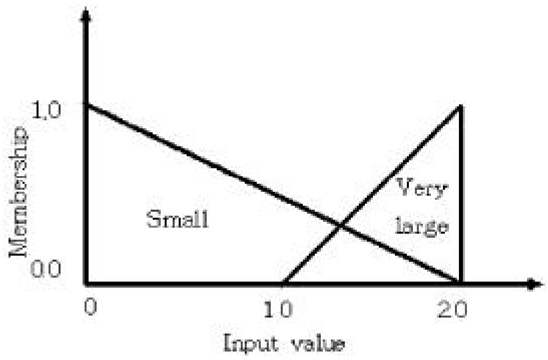


그림 1. 언어값 S와 VL의 소속 함수

$$\text{그룹}1 = \{(4,11),(8,11),(11,3),(13,4),(13,10),(S,S)\} \quad (3\text{-}1)$$
$$\text{그룹}2 = \{(2,13),(6,14),(13,2),(14,3),(14,14),(VL,U),(U,VL)\}$$

실험 결과는 그림 2에 나타나 있다. 여기서 직선 $<m_W, X> + m_B = 0$은 두 그룹의 경계선(boundary)을 나타내고, 점선은 $<\alpha_W, |X|> + \alpha_B = 0$인 fuzzy spread를 나타낸다.

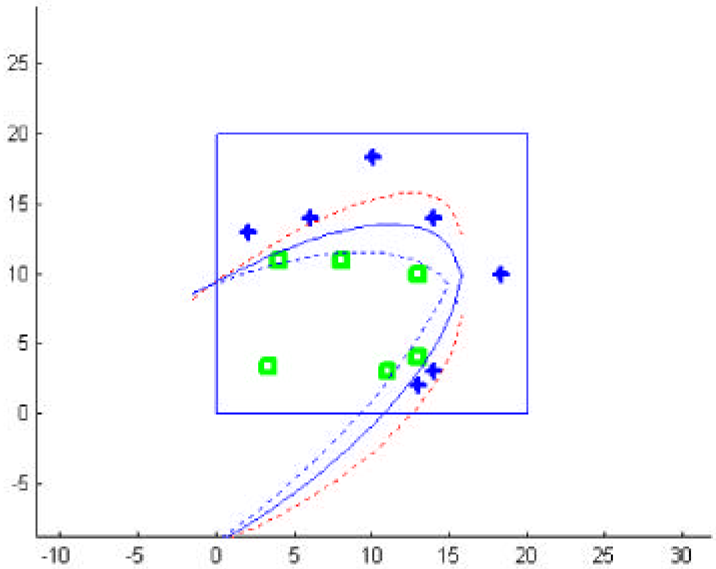


그림 2 X=[ 4 11; 8 11; 11 3; 13 4; 13 10; S S;
2 13; 6 14; 13 2 ;14 3; 14 14; VL U; U VL ];
Y=[ -1;-1;-1;-1;-1;-1; 1; 1; 1; 1; 1; 1; 1 ];

{1-,1}은 각 학습 데이터가 속하는 그룹에 대한 인덱스로 $Y_i$는 이의 부분 집합으로, 직선위의 그룹은 $Y_i=1$이고 직선아래의 class는 $Y_i=-1$로 보았을 경우 사각형형태 데이터중 4개의 데이터($Y_i=-1$), 십자가 데이터형태중 4개의 데이터($Y_i=1$)가 정확이 구분되었으며 나머지 5개의 데이터가 스프레드 선상내에 있는 것으로 나타났다.

다음으로 아래와 같은 퍼지입력자료에 적용하여 보자. 아래의 자료자료 $A=(a,b)_L$ 은 소속함수가 $\mu_A=\max\{1-|x-a|/b,0\}$이며 중심이 a이고 스프레드가 b인 대칭 삼각퍼지수(symmetric triangular fuzzy number)이다.

$$
\begin{aligned}
\text{그룹 } 1 &= \{((3,2)_L,(3,2)_L),((3,2)_L,(11,3)_L),\\
&\quad((9,3)_L,(5,2)_L,),((10,2)_L,(10,2)_L)\}\\
\text{그룹 } 2 &= \{((6,3)_L,(18,2)_L),((12,2)_L,(12,2)_L),\\
&\quad((14,3)_L,(17,2)_L,),((16,2)_L,(5,4)_L)\}
\end{aligned}
\qquad (3\text{-}2)
$$

결과는 그림 3에 나타나 있다. 점선바깥쪽의 아래데이터는 $Y_i=-1$인 경우이며 점선바깥쪽의 위쪽데이터는 $Y_i=1$인경우로 정확히 구분된다.

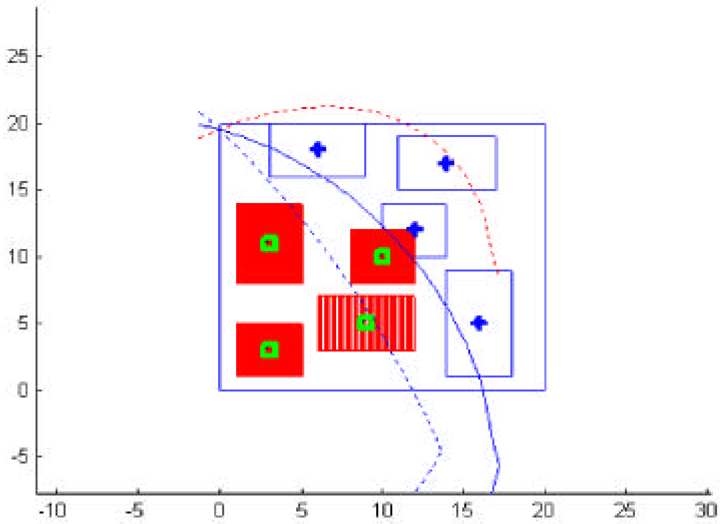


그림 3 $\text{Class } 1 = \{((3,2)_L,(3,2)_L),((3,2)_L,(11,3)_L),((9,3)_L,(5,2)_L,),((10,2)_L,(10,2)_L)\}$

$\text{Class } 2 = \{((6,3)_L,(18,2)_L),((12,2)_L,(12,2)_L),((14,3)_L,(17,2)_L,),((16,2)_L,(5,4)_L)\}$

## 결론

이상과 같이 SVM을 이용하여 퍼지자료에 대한 분류에 대해서 살펴보았다. 제안된 알고리즘은 이 논문에서 보여준 간단한 예제에는 잘 적용이 되었다. 실용적인 문제 특히 대용량 자료의 문제에 대해서 적용할 수 있도록 개발되어야 하겠다. 그리고 향후 다양한 커널함수에 대해 본 논문에서 제안된 알고리즘을 적용하여 그 결과를 비교해 보는 것이 좋을듯하다.

## 참고문헌

[1]. R. Fujioka, and H. Tanaka (1993). Neural Networks That Learn form Fuzzy If-Then Rules, *IEEE Transactions on fuzzy systems*, 1, NO2. 85-97.

[2]. S. Gunn (1998). *Support Vector Machines for Classfication and Regression*. ISIS Technical Report.

[3]. V. Vapnik (1995). *The Nature of Statistical Learning Theory*, Springer.

[4]. V. Vapnik (1998). *Statistical Learning Theory*, Springer.