[SWCON253] Machine Learning – Lec.15b

Support Vector Machine

Fall 2025

김 휘 용

hykim.v@khu.ac.kr

경희대학교
KYUNG HEE UNIVERSITY

Contents

References

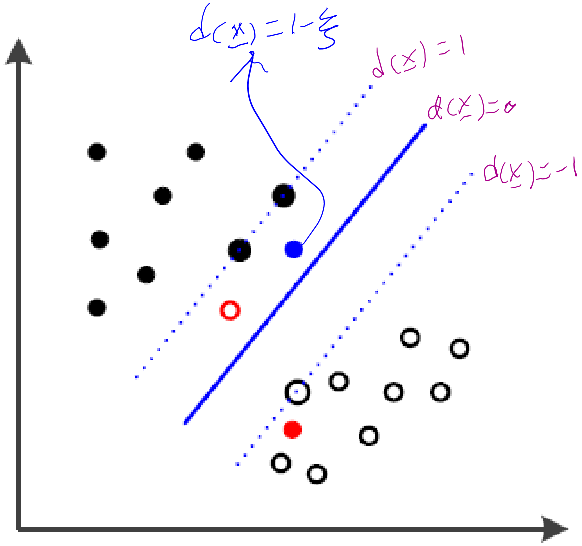
- *기계 학습* by 오일석, *패턴 인식* by 오일석

3. Linear SVM with Soft Margin

- ✓ Hard vs. Soft Margin
- ✓ Linear SVM with Soft Margin
- ✓ Illustrative Example

Hard vs. Soft Margin

- 앞의 Linear SVM은 ***Hard Margin***을 사용
  - Hard Margin 내에는 샘플이 존재하지 않는다.
  - Hard Margin을 사용하면 선형분리가 불가능한 상황에서는 해를 구할 수 없다.



- ***Soft Margin & Slack Variable***
  - Soft Margin이란 분할 띠 안에도 샘플을 허용하는 것을 말한다.
  - 각 샘플 ($\mathbf{x}_i$, y_i)은 분할 띠 와의 위치 관계에 따하 다음 표의 세 가지 case 중 하나에 속하며, **슬랙 변수(*Slack Variable*)** ξ를 도입하여 세 case를 하나로 표현할 수 있다.

Case	분류 결과	샘플 위치	그림 표시	$y_i(\mathbf{w}^\mathrm{T}\mathbf{x}_i + b)$	슬랙 변수
1	옳게 분류	분할 띠 바깥	● o	$1 \le y_i(\mathbf{w}^\mathrm{T}\mathbf{x}_i + b)$	$\xi_i = 0$
2	옳게 분류	분할 띠 안쪽	●	$0 \le y_i(\mathbf{w}^\mathrm{T}\mathbf{x}_i + b) < 1$	$0 < \xi_i \le 1$
3	틀리게 분류	결정 경계 넘음	● o	$y_i(\mathbf{w}^\mathrm{T}\mathbf{x}_i + b) < 0$	$1 < \xi_i$

$$\boxed{y_i(\mathbf{w}^\mathrm{T}\mathbf{x}_i + b) \ge 1 - \xi_i}$$

Linear SVM with Soft Margin – *Formulation*

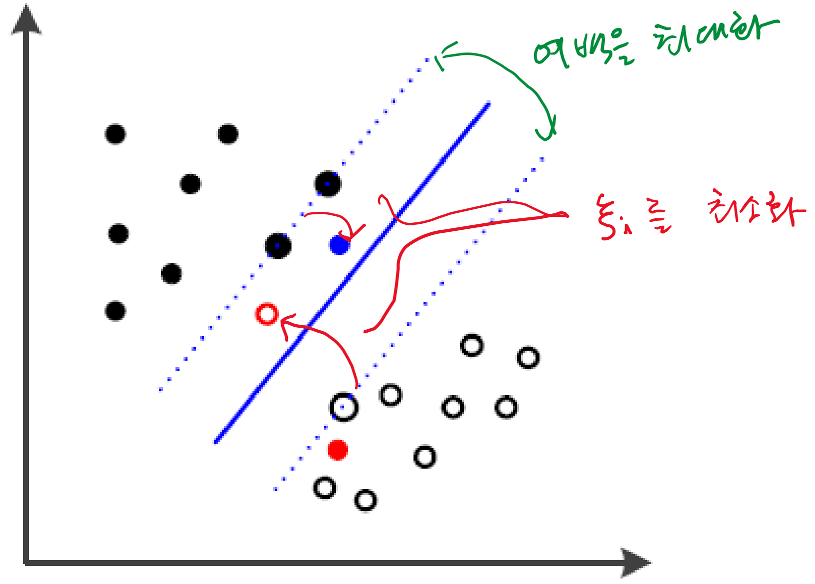
◆ Problem Definition

여백을 될 수 있는 한 크게 하면서(목적 1), $0 < \xi$인 (즉, 경우 2와 경우 3에 속하는) 샘플의 수를 될 수 있는 한 적게 하는(목적 2) 결정 초평면의 방향 $\mathbf{w}$를 찾아라.

Min. $J(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n} \xi_i$ (C는 hyper parameter)

s.t. $y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 - \xi_i, i = 1,2,\cdots,n$

$0 \leq \xi_i, i = 1,2,\cdots,n$



Linear SVM with Soft Margin – *Lagrange Method*

- Lagrangian

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\xi_i\right) - \left(\sum_{i=1}^{n}\alpha_i\left(y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1 + \xi_i\right) + \sum_{i=1}^{n}\beta_i\xi_i\right)$$

- KKT Condition → 결국 α_i, β_i를 구하는 문제

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{0} \quad \Longrightarrow \quad \mathbf{w} = \sum_{i=1}^{n}\alpha_i\, y_i\mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = 0 \quad \Longrightarrow \quad \sum_{i=1}^{n}\alpha_i\, y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = 0 \quad \Longrightarrow \quad C = \alpha_i + \beta_i, i = 1,2,\cdots,n$$

$$0 \le \alpha_i, i = 1,2,\cdots,n$$

$$0 \le \beta_i, i = 1,2,\cdots,n$$

$$\alpha_i\left(y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1 + \xi_i\right) = 0, i = 1,2,\cdots,n$$

$$\beta_i\xi_i = 0, i = 1,2,\cdots,n$$

Linear SVM with Soft Margin – *Dual Problem*

- **Wolfe Dual Problem**

Max. $$\tilde{L}(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i{}^{\mathrm{T}} \mathbf{x}_j$$

s.t. $$\sum_{i=1}^{n} \alpha_i \, y_i = 0$$

$$0 \le \alpha_i \boxed{\le C,} \qquad i = 1,2,\cdots,n$$

Linear SVM with Soft Margin – *Example*

◆ 선형분리가 불가능한 경우, Wolfe Dual Problem의 해를 구해 보자.

- **훈련 집합:** $\mathbf{x}_1 = \binom{2}{3},\ \mathbf{x}_2 = \binom{4}{1},\ \mathbf{x}_3 = \binom{5}{1},\ y_1 = 1,\ y_2 = -1,\ y_3 = 1$

- **Wolfe Dual Problem:**

Maximize: $\tilde{\mathcal{L}}(\boldsymbol{\alpha}) = (\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2}(13\alpha_1^2 + 17\alpha_2^2 + 26\alpha_3^2 - 22\alpha_1\alpha_2 + 26\alpha_1\alpha_3 - 42\alpha_2\alpha_3)$

Subject to: $\alpha_1 - \alpha_2 + \alpha_3 = 0$

$0 \le \alpha_1 \le C, 0 \le \alpha_2 \le C, 0 \le \alpha_3 \le C$

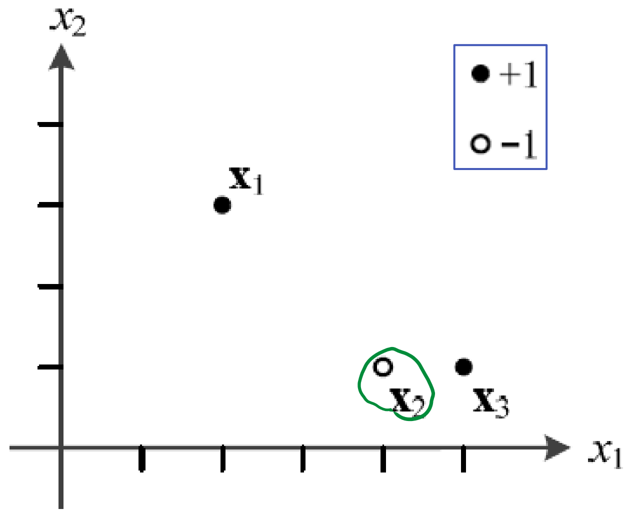
- **풀이:**

★ Class별로 Support Vector가 하나 이상 있어야 하므로,
다음의 세 가지 경우만 가능하다.

(1) $\alpha_1 = 0, \alpha_2 \ne 0, \alpha_3 \ne 0$

(2) $\alpha_1 \ne 0, \alpha_2 \ne 0, \alpha_3 = 0$

(3) $\alpha_1 \ne 0, \alpha_2 \ne 0, \alpha_3 \ne 0$



Linear SVM with Soft Margin – *Illustrative Example (cont'd)*

(1) $\alpha_1 = 0,\ \alpha_2 \neq 0,\ \alpha_3 \neq 0$

$\alpha_1 - \alpha_2 + \alpha_3 = 0$으로부터 $\alpha_2 = \alpha_3$

$$\tilde{\mathcal{L}}(\boldsymbol{\alpha}) = 2\alpha_2 - \frac{1}{2}\alpha_2^2 = -\frac{1}{2}((\alpha_2 - 2)^2 - 4)$$

$$\Longrightarrow \quad \alpha_1 = 0,\ \alpha_2 = 2,\ \alpha_3 = 2$$

$$\mathbf{w} = (2,0)^{\mathrm{T}}, b = -9$$

$$d(\mathbf{x}) = 2x_1 - 9$$

$$d(\mathbf{x}_1) = -5, d(\mathbf{x}_2) = -1, d(\mathbf{x}_3) = 1$$

(2) $\alpha_1 \neq 0,\ \alpha_2 \neq 0,\ \alpha_3 = 0$

$\alpha_1 - \alpha_2 + \alpha_3 = 0$으로부터 $\alpha_1 = \alpha_2$

$$\tilde{\mathcal{L}}(\boldsymbol{\alpha}) = 2\alpha_1 - 4\alpha_1^2 = -4\left(\left(\alpha_1 - \frac{1}{4}\right)^2 - \frac{1}{16}\right)$$

$$\Longrightarrow \quad \alpha_1 = \frac{1}{4},\ \alpha_2 = \frac{1}{4},\ \alpha_3 = 0$$

$$\mathbf{w} = \left(-\frac{1}{2}, \frac{1}{2}\right)^{\mathrm{T}}, b = \frac{1}{2}$$

$$d(\mathbf{x}) = -\frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}$$

$$d(\mathbf{x}_1) = 1, d(\mathbf{x}_2) = -1, d(\mathbf{x}_3) = -\frac{3}{2}$$

(3) $\alpha_1 \neq 0,\ \alpha_2 \neq 0,\ \alpha_3 \neq 0$

$\alpha_1 - \alpha_2 + \alpha_3 = 0$으로부터 $\alpha_1 = \alpha_2 - \alpha_3$

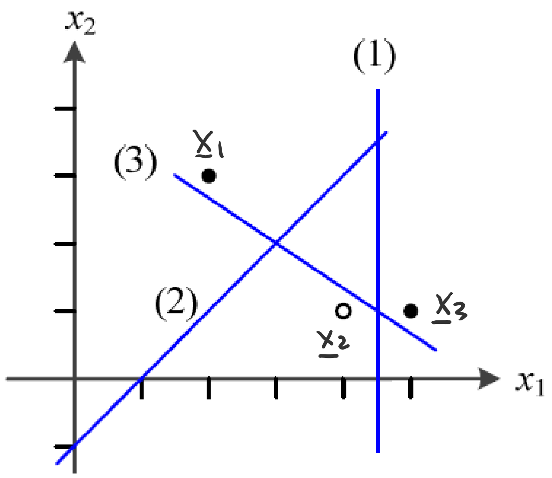
$\dfrac{\partial \tilde{\mathcal{L}}(\boldsymbol{\alpha})}{\partial \alpha_2} = 0$과 $\dfrac{\partial \tilde{\mathcal{L}}(\boldsymbol{\alpha})}{\partial \alpha_3} = 0$을 풀면

$$\Longrightarrow \quad \alpha_1 = \frac{3}{2}, \alpha_2 = \frac{13}{2}, \alpha_3 = 5$$

$$\mathbf{w} = (2,3)^{\mathrm{T}}, b = -12$$

$$d(\mathbf{x}) = 2x_1 + 3x_2 - 12$$

$$d(\mathbf{x}_1) = 1, d(\mathbf{x}_2) = -1, d(\mathbf{x}_3) = 1$$



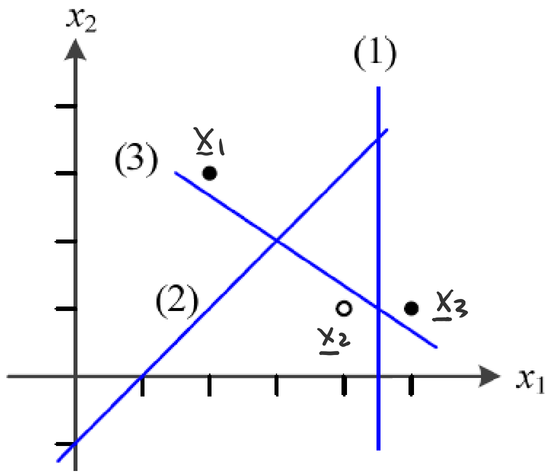
Linear SVM with Soft Margin – *Illustrative Example (cont'd)*

◆ **SVM은 (1)~(3) 중 어느 결정 직선을 선택하게 될까? → C에 따라 다름**

(1) $\alpha_1 = 0,\ \alpha_2 = 2,\ \alpha_3 = 2$ ⇒ $\tilde{L}(\alpha) = 2$ $\quad d(\mathbf{x}_1) = -5, d(\mathbf{x}_2) = -1, d(\mathbf{x}_3) = 1$ ⇒ $\frac{2}{\|W\|_2} = 1$

(2) $\alpha_1 = \frac{1}{4},\ \alpha_2 = \frac{1}{4},\ \alpha_3 = 0$ ⇒ $= \frac{1}{4}$ $\quad d(\mathbf{x}_1) = 1, d(\mathbf{x}_2) = -1, d(\mathbf{x}_3) = -\frac{3}{2}$ ⇒ $= 2\sqrt{2}$

(3) $\alpha_1 = \frac{3}{2},\ \alpha_2 = \frac{13}{2},\ \alpha_3 = 5$ ⇒ $= \frac{13}{2}$ $\quad d(\mathbf{x}_1) = 1, d(\mathbf{x}_2) = -1, d(\mathbf{x}_3) = 1$ ⇒ $= \frac{2}{\sqrt{13}}$



● ***C*에 따른 유효성:**

★ $C < 2$: 결정 직선 (2)만 유효 $\qquad 0 \le \alpha_i \le C, \qquad i = 1,2,\cdots,n$

★ $2 \le C < 6.5$: 결정 직선 (1)과 (2)가 유효

★ $6.5 \le C$: 결정 직선 (1)~(3)이 모두 유효

★ [참고] 여러 결정 직선이 유효한 경우에는, $\tilde{L}(\boldsymbol{\alpha})$ 가 큰 걸 선택함

● ***C*의 크기와 여백**

$$J(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\xi_i \qquad y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \ge 1 - \xi_i$$

★ C를 작게 하면 여백의 비중이 커진다. (여백 최대화)

★ C를 크게 하면 ξ_i의 비중이 커진다. (분할 띠 안쪽 샘플 수 최소화)

Case	분류 결과	샘플 위치	슬랙 변수
1	옳게 분류	분할 띠 바깥	$\xi = 0$
2	옳게 분류	분할 띠 안쪽	$0 < \xi \le 1$
3	틀리게 분류	결정 경계 넘음	$1 < \xi$

Q & A