

기계학습 개론

- Support Vector Machine (SVM)

정보통신공학과

Prof. Jinkyu Kang



MYONGJI
UNIVERSITY

이번주 수업의 목차

- 선형 SVM
 - 선형 분리가 가능한 상황
 - 선형 분리가 불가능한 상황
- SVM 구현
 - 학습과 인식
 - M 부류 SVM
- SVM의 특성

들어가는 말

- SVM의 차별성
 - 기존 분류기는 ‘오류율을 최소화’
 - SVM은 ‘여백을 ^{margin} 최대화’하여 일반화 능력의 극대화 꾀함

SVM의 발상

- 분류기의 일반화 능력
 - ②보다 ③이 여백이 더 크다.
 - 즉 ③이 ②보다 일반화 능력이 뛰어나다.
 - 신경망은 초기값 ①에서 시작하여 ②를 찾았다면 거기서 멈춘다. 왜?
 - SVM은 ③을 찾는다.
- 중요한 문제
 - 여백이라는 개념을 어떻게 공식화할 것인가?
 - 여백을 최대로 하는 결정 초평면을 어떻게 찾을 것인가?

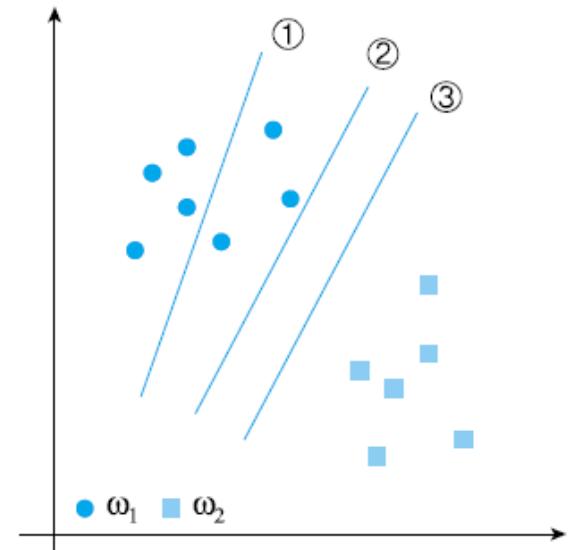


그림 5.1 분류기의 일반화 능력



선형 SVM

- 이진 분류를 위한 결정 초평면과 그것의 수학적 특성

$$d(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \quad (5.1)$$

1. $d(\mathbf{x})$ 는 전체 특징 공간을 두 영역으로 분할하며 한 쪽 영역에 속하는 점 \mathbf{x} 는

$d(\mathbf{x}) > 0$ 이고 다른 쪽에 있는 점은 $d(\mathbf{x}) < 0$ 이다.

2. 하나의 초평면을 표현하는 식은 여럿 있다. (5.1)에 0이 아닌 임의의 상수 c 를 곱하여도 같은 초평면을 나타낸다.

3. \mathbf{w} 는 초평면의 법선 벡터로서 normal vector 초평면의 방향을 나타내고 b 는 위치를 나타낸다.

4. 임의의 점 \mathbf{x} 에서 초평면까지의 거리는 (5.2)와 같다.

$$h = \frac{|d(\mathbf{x})|}{\|\mathbf{w}\|} \quad (5.2)$$



선형 SVM

- 결정 초평면의 수학적 특성 예제

그림 5.2에 있는 결정 직선의 수학적 특성을 살펴보자. 이 직선의 매개 변수는 $\mathbf{w} = (2,1)^T$ 이고 $b = -4$ 이다.

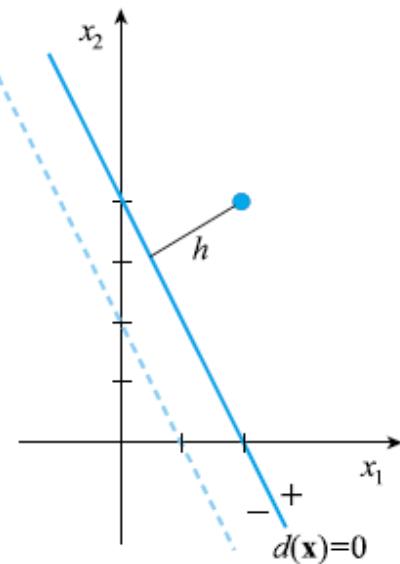


그림 5.2 직선의 수학적 특성

아래는 모두 같은 직선

$$d(\mathbf{x}) = 2x_1 + x_2 - 4 = 0$$

$$d(\mathbf{x}) = x_1 + 0.5x_2 - 2 = 0$$

$$d(\mathbf{x}) = 6x_1 + 3x_2 - 12 = 0$$

점 $\mathbf{x} = (2,4)^T$ 에서 직선까지 거리

$$h = \frac{|2 \times 2 + 1 \times 4 - 4|}{\sqrt{2^2 + 1^2}} = \frac{4}{\sqrt{5}} = 1.78885$$

선형 SVM | 선형 분리 가능한 상황

- w (직선의 방향)가 주어진 상황에서,
 - '두 부류에 대해 직선으로부터 가장 가까운 샘플까지의 거리가 같게 되는' b 를 결정 (①과 ②는 그렇게 얻은 직선)
 - **여백**은 그런 직선에서 가장 가까운 샘플까지 거리의 두 배로 정의함
 - 가장 가까운 샘플을 **서포트 벡터**라 부름

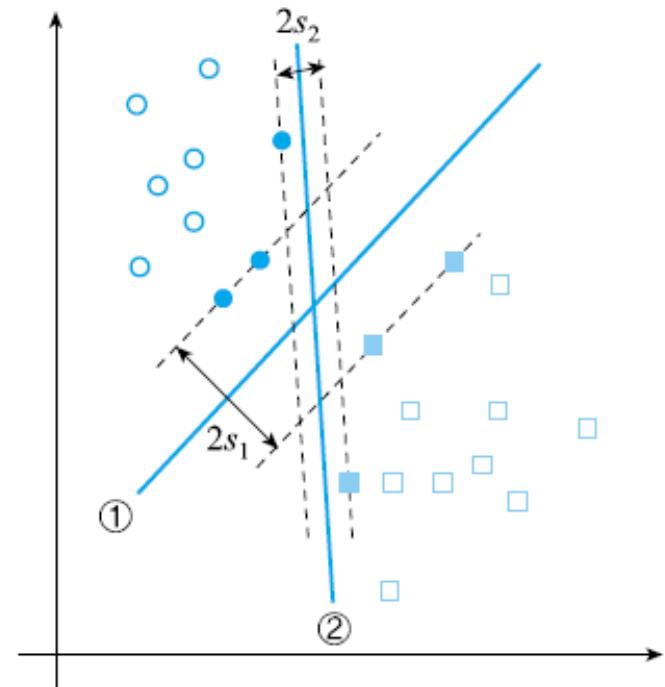


그림 5.3 선형 분리 가능한 상황



선형 SVM | 선형 분리 가능한 상황

- 이제 문제를 공식화해 보자.
 - 여백을 가장 크게 하는 결정 초평면의 방향, 즉 \mathbf{w} 를 찾아라. $(5.3)^2$
 - 그림 5.3에서 ①과 ②는 어느 것이 최적에 가까운가?
 - ①보다 나은 것이 있나?
 - 전형적인 최적화 문제임



선형 SVM | 선형 분리 가능한 상황

- 여백은 아래와 같이 공식화

$$\text{여백} = 2h = \frac{2|d(\mathbf{x})|}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (5.4)$$

- 이제 문제를 조건부 최적화 문제로 공식화

- 조건부 최적화 문제

$$\left. \begin{array}{l} \text{아래 조건 하에,} \\ \mathbf{w}^T \mathbf{x}_i + b \geq 1, \forall \mathbf{x}_i \in \omega_1 \\ \mathbf{w}^T \mathbf{x}_i + b \leq -1, \forall \mathbf{x}_i \in \omega_2 \\ \frac{2}{\|\mathbf{w}\|} \text{를 최대화하라.} \end{array} \right\} \quad (5.5)$$

- 훈련 집합 $X = \{(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_N, t_N)\}$

선형 SVM | 선형 분리 가능한 상황

- 라그랑제 승수 방법을 기반으로 수식 정리를 하면,

- 조건부 최적화 문제

아래 조건 하에,

$$\sum_{i=1}^N \alpha_i t_i = 0$$

$$\alpha_i \geq 0, i = 1, \dots, N$$

$$\tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j$$
 를 최대화하라.

$$\mathbf{w} = \sum_{i=1}^N \alpha_i t_i \mathbf{x}_i \quad (5.8)$$

$$\alpha_i (t_i (\mathbf{w}^T \mathbf{x}_i + b) - 1) = 0, i = 1, \dots, N \quad (5.11)$$

(5.13)

- (5.11)에 의하면 모든 샘플이 $\alpha_i=0$ 또는 $t_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 = 0$ 이어야 함. $\alpha_i \neq 0$ 인 샘플이 서포트 벡터임
- 흥미로운 특성들
 - 2차 함수의 최대화 문제임
 - \mathbf{w} 와 b 가 사라졌다. (α 를 찾는 문제가 되었다.)
 - 특징 벡터 \mathbf{x}_i 가 내적 형태로 나타난다. (비선형으로 확장하는 발판)
 - 목적 함수의 두 번째 항은 N^2 개의 항을 갖는다. (여전히 풀기 어려운 문제)

선형 SVM | 선형 분리 가능한 상황

- 예제: 두 개 샘플을 가진 경우의 문제 (5.13)의 풀이

- 훈련집합 $\mathbf{x}_1 = (2,3)^T, t_1 = 1$
 $\mathbf{x}_2 = (4,1)^T, t_2 = -1$

- (5.13)은

아래 조건 하에,

$$\alpha_1 t_1 + \alpha_2 t_2 = 0$$

$$\alpha_1 \geq 0, \alpha_2 \geq 0$$

$$\tilde{L}(\mathbf{a}) = (\alpha_1 + \alpha_2)$$

$$-\frac{1}{2}(\alpha_1 \alpha_1 t_1 t_1 \mathbf{x}_1^T \mathbf{x}_1 + \alpha_1 \alpha_2 t_1 t_2 \mathbf{x}_1^T \mathbf{x}_2 + \alpha_2 \alpha_1 t_2 t_1 \mathbf{x}_2^T \mathbf{x}_1 + \alpha_2 \alpha_2 t_2 t_2 \mathbf{x}_2^T \mathbf{x}_2)$$
 를 최대화하라.

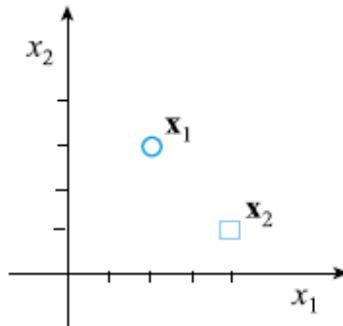
- 실제 값을 대입하면,

아래 조건 하에,

$$\alpha_1 - \alpha_2 = 0$$

$$\alpha_1 \geq 0, \alpha_2 \geq 0$$

$$\tilde{L}(\mathbf{a}) = (\alpha_1 + \alpha_2) - \frac{1}{2}(13\alpha_1^2 + 17\alpha_2^2 - 22\alpha_1\alpha_2)$$
 를 최대화하는 $\mathbf{a} = (\alpha_1, \alpha_2)^T$ 를 찾아라.



(a) 문제

선형 SVM | 선형 분리 가능한 상황

- 예제: 두 개 샘플을 가진 경우의 문제 (5.13)의 풀이 (cont.)
 - 정리하여 풀면,

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

$(1/4, 1/4)^T$ 에서 최대값을 가짐

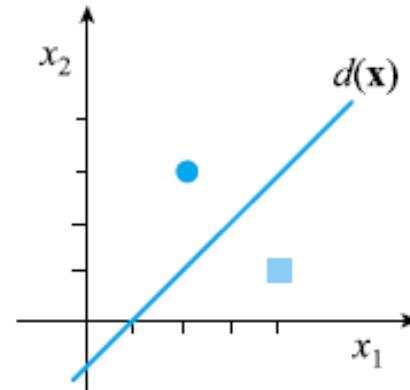
- (5.8)로 \mathbf{w} , (5.11)로 b 를 구하면

$$\mathbf{w} = \sum_{i=1}^2 \alpha_i t_i \mathbf{x}_i = \frac{1}{4}(2, 3)^T - \frac{1}{4}(4, 1)^T = \left(-\frac{1}{2}, \frac{1}{2}\right)^T$$

$$b = \frac{1}{2}$$

- 결국 결정 직선은

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



(b) SVM (속이 찬 샘플이 서포트 벡터)

선형 SVM | 선형 분리 가능한 상황

- 예제: 세 개 샘플을 가진 경우의 문제 (5.13)의 풀이

- 훈련집합 $\mathbf{x}_1 = (2,3)^T, t_1 = 1$

$$\mathbf{x}_2 = (4,1)^T, t_2 = -1$$

$$\mathbf{x}_3 = (5,1)^T, t_3 = -1$$

- (5.13)에 실제 값을 대입하면,

아래 조건 하에,

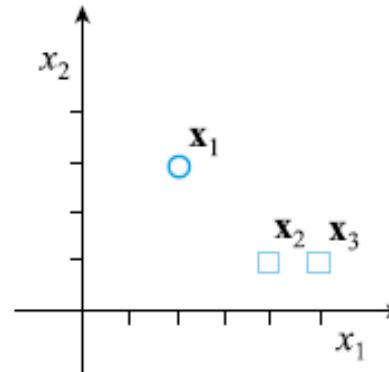
$$\alpha_1 - \alpha_2 - \alpha_3 = 0$$

$$\alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_3 \geq 0$$

$$\tilde{L}(\mathbf{a}) = (\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2}(13\alpha_1^2 + 17\alpha_2^2 + 26\alpha_3^2 - 22\alpha_1\alpha_2 - 26\alpha_1\alpha_3 + 42\alpha_2\alpha_3) \text{ 을}$$

최대화하는 $\mathbf{a} = (\alpha_1, \alpha_2, \alpha_3)^T$ 을 찾아라.

- 이 문제를 어떻게 풀 것인가?

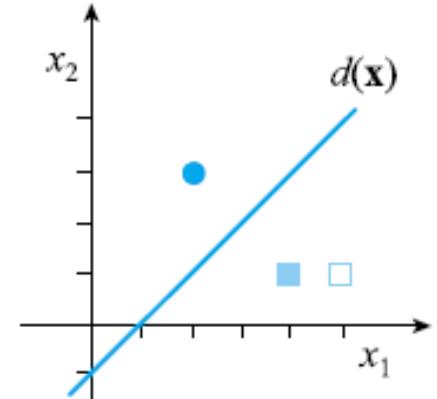


(a) 문제

선형 SVM | 선형 분리 가능한 상황

- 네 가지 경우로 나누어 분석적 풀이
 - ① $\alpha_1=0, \alpha_2\neq0, \alpha_3\neq0$
 - ② $\alpha_1\neq0, \alpha_2=0, \alpha_3\neq0$
 - ③ $\alpha_1\neq0, \alpha_2\neq0, \alpha_3=0$
 - ④ $\alpha_1\neq0, \alpha_2\neq0, \alpha_3\neq0$
- ① ② ④ 는 모순이므로 버림. 왜 모순?
- ③ $\alpha_1\neq0, \alpha_2\neq0, \alpha_3=0$ 인 경우를 풀면,
 - 등식 조건으로부터 $\alpha_1=\alpha_2$ 이므로
결국 $\alpha_1=1/4, \alpha_2=1/4, \alpha_3=0$
 - (5.8)과 (5.11)로 \mathbf{w} 와 b 를 구하면
 - 결정 직선은

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



(b) SVM (속이 찬 샘플이 서포트 벡터)

$$\tilde{L}(\mathbf{a}) = 2\alpha_1 - 4\alpha_1^2 = -4\left(\left(\alpha_1 - \frac{1}{4}\right)^2 - \frac{1}{16}\right)$$

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

$$b = \frac{1}{2}$$



선형 SVM | Python을 이용한 SVM

- 예제: 세 개 샘플을 가진 경우의 문제 (5.13)의 풀이

- 훈련집합 $\mathbf{x}_1 = (2, 3)^T, t_1 = 1$

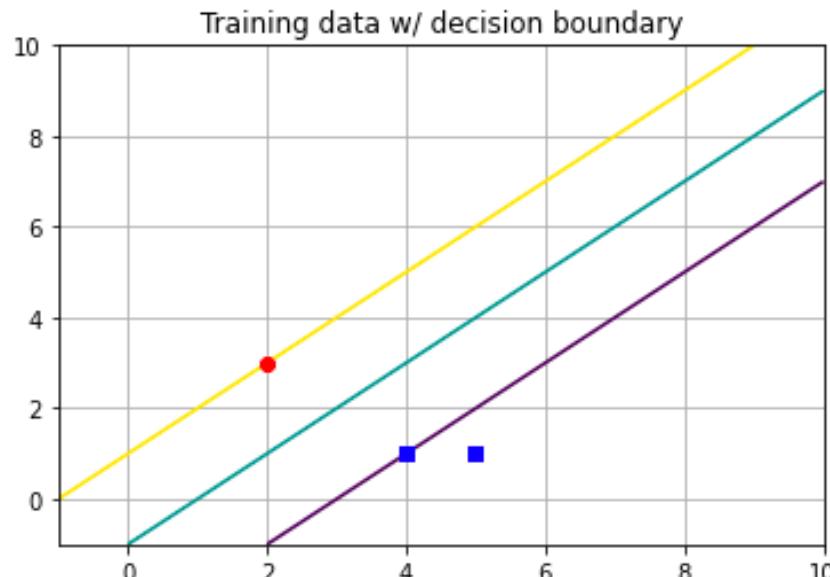
$$\mathbf{x}_2 = (4, 1)^T, t_2 = -1$$

$$\mathbf{x}_3 = (5, 1)^T, t_3 = -1$$

$\Leftrightarrow \omega_1(1)$ 에 속하는 샘플: $(2, 3)^T$

$\omega_2(-1)$ 에 속하는 샘플: $(4, 1)^T, (5, 1)^T$

- 테스트집합: $(0, 0)^T, (3, 1)^T$



선형 SVM | 선형 분리 불가능한 상황

- 선형 분리 불가능한 상황

- 샘플 (x, t) 의 세 가지 상황

- 경우 1: 분할 띠의 바깥에 있다. $1 \leq t(\mathbf{w}^T \mathbf{x} + b)$ 를 만족한다.
 - 경우 2: 분할 띠의 안쪽에 있는데 자기가 속한 부류의 영역에 있다. $0 \leq t(\mathbf{w}^T \mathbf{x} + b) < 1$ 을 만족한다.
 - 경우 3: 결정 경계를 넘어 자신이 속하지 않은 부류의 영역에 놓여 있다. $t(\mathbf{w}^T \mathbf{x} + b) < 0$ 을 만족한다.

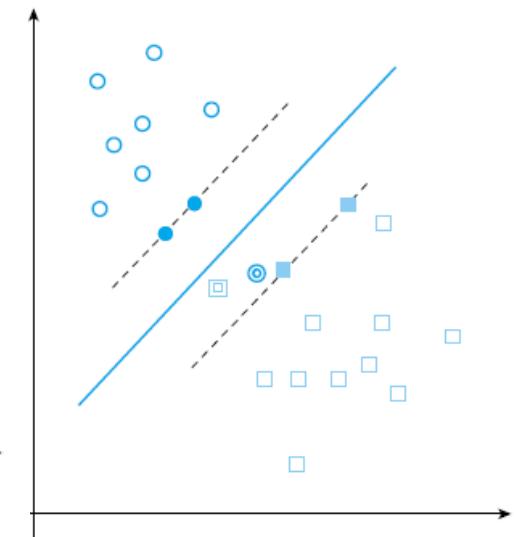


그림 5.6 선형 분리 불가능한 상황의 SVM



선형 SVM | 선형 분리 불가능한 상황

- 슬랙 변수 ξ 를 도입하여 하나의 식으로 쓰면,

$$t(\mathbf{w}^T \mathbf{x} + b) \geq 1 - \xi \quad (5.14)$$

표 5.2 선형 분리 불가능한 상황에서 샘플의 세 가지 경우

	샘플 위치	분류	$t(\mathbf{w}^T \mathbf{x} + b)$ 값	슬랙 변수	그림 5.6에서 기호
경우 1	분할 띠 바깥	옳게 분류	$1 \leq t(\mathbf{w}^T \mathbf{x} + b)$	$\xi = 0$	□○●■
경우 2	분할 띠 안쪽	옳게 분류	$0 \leq t(\mathbf{w}^T \mathbf{x} + b) < 1$	$0 < \xi \leq 1$	□
경우 3	결정 경계 넘음	틀리게 분류	$t(\mathbf{w}^T \mathbf{x} + b) < 0$	$1 < \xi$	◎

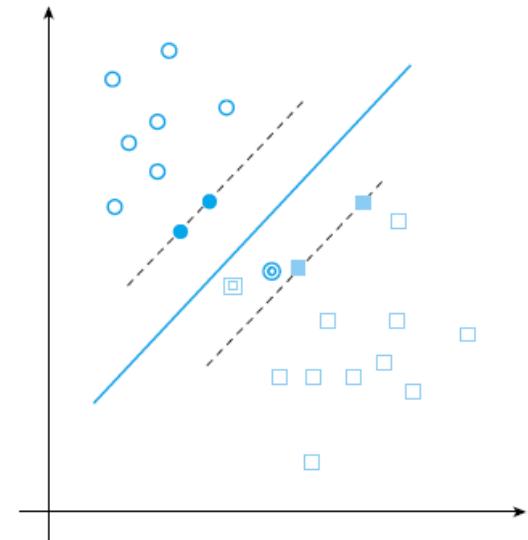


그림 5.6 선형 분리 불가능한 상황의 SVM

선형 SVM | 선형 분리 불가능한 상황

- 문제 공식화
 - 길항^{tradeoff} 관계를 갖는 두 가지 목적을 동시에 달성
 - 여백을 될 수 있는 한 크게 하며 (목적 1), 동시에 $0 < \xi$ 인 (즉 경우 2 또는 경우 3에 해당하는) 샘플의 수를 될 수 있는 한 적게 하는 (목적 2) 결정 초평면의 방향 \mathbf{w} 를 찾아라.
 - 목적 함수를 다시 쓰면

$$J(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (5.16)$$

- 첫번째 항은 목적 1, 두번째 항은 목적 2



선형 SVM | 선형 분리 불가능한 상황

- 문제 공식화

- 조건부 최적화 문제

아래 조건 하에,

$$\left. \begin{array}{l} t_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, \dots, N \\ \xi_i \geq 0, \quad i = 1, \dots, N \\ J(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \text{ 를 최소화하라.} \end{array} \right\} \quad (5.17)$$

선형 SVM | 선형 분리 불가능한 상황

- 라그랑제 승수로 풀어보면,
 - (5.17)은 아래와 같이 변한다.

- 조건부 최적화 문제 (선형 SVM)

아래 조건 하에,

$$\sum_{i=1}^N \alpha_i t_i = 0$$

$$0 \leq \alpha_i \leq C, i = 1, \dots, N$$

$$\tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j \mathbf{x}_i^T \mathbf{x}_j$$
 를 최대화하라.

} (5.27)

- (5.13)과 같다! 한 가지만 빼고.
 - $0 \leq \alpha_i$ 가 $0 \leq \alpha_i \leq C$ 로 바뀜



선형 SVM | 선형 분리 불가능한 상황

- 예제: 세 개 샘플을 가진 경우의 문제 (5.27)의 풀이
 - 훈련집합 (선형 분리 가능한가?)

$$\mathbf{x}_1 = (2, 3)^T, t_1 = 1$$

$$\mathbf{x}_2 = (4, 1)^T, t_2 = -1$$

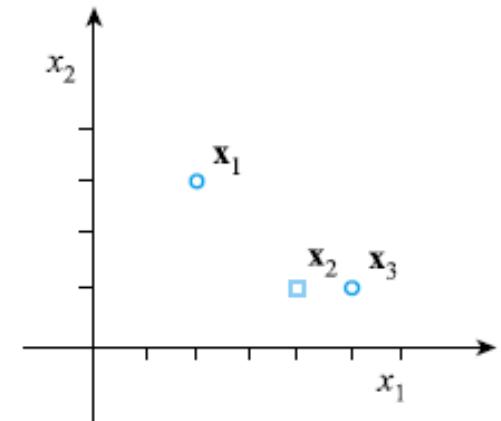
$$\mathbf{x}_3 = (5, 1)^T, t_3 = 1$$

- (5.27)에 실제 값을 대입하면,

아래 조건 하에,

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

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



(a) 문제

$$\tilde{L}(\mathbf{a}) = (\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2} (13\alpha_1^2 + 17\alpha_2^2 + 26\alpha_3^2 - 22\alpha_1\alpha_2 + 26\alpha_1\alpha_3 - 42\alpha_2\alpha_3) \text{ 를}$$

최대화하는 $\mathbf{a} = (\alpha_1, \alpha_2, \alpha_3)^T$ 을 찾아라.

- 이 문제를 어떻게 풀 것인가?

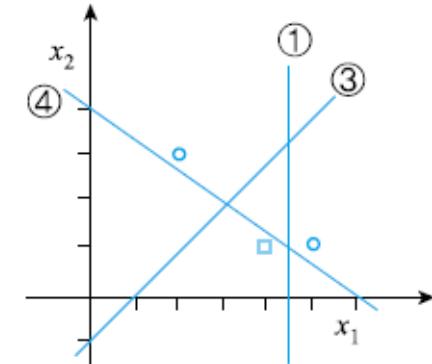
선형 SVM | 선형 분리 불가능한 상황

- 네 가지 경우로 나누어 분석적 풀이
 - ① $\alpha_1=0, \alpha_2\neq0, \alpha_3\neq0$
 - ② $\alpha_1\neq0, \alpha_2=0, \alpha_3\neq0$
 - ③ $\alpha_1\neq0, \alpha_2\neq0, \alpha_3=0$
 - ④ $\alpha_1\neq0, \alpha_2\neq0, \alpha_3\neq0$
- ① $\alpha_1=0, \alpha_2\neq0, \alpha_3\neq0$ 인 경우를 풀면,
 - 등식 조건으로부터 $\alpha_2=\alpha_3$ 이므로 $\tilde{L}(\alpha) = 2\alpha_2 - \frac{1}{2}\alpha_2^2 = -\frac{1}{2}((\alpha_2 - 2)^2 - 4)$
결국 $\alpha_1=0, \alpha_2=2, \alpha_3=2$
 - \mathbf{w} 와 b 를, 그리고 결정 직선

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

$$d(\mathbf{x}) = 2x_1 - 9 = 0$$

- 그림에서 직선 ①
 - x_1 은 오분류됨



(b) 경우 ①, ③, ④의 결정 직선

선형 SVM | 선형 분리 불가능한 상황

- ② $\alpha_1 \neq 0, \alpha_2 = 0, \alpha_3 \neq 0$ 인 경우는 모순. 왜?
- ③ $\alpha_1 \neq 0, \alpha_2 \neq 0, \alpha_3 = 0$ 인 경우를 풀면,
 - 등식 조건으로부터 $\alpha_1 = \alpha_2$ 이므로

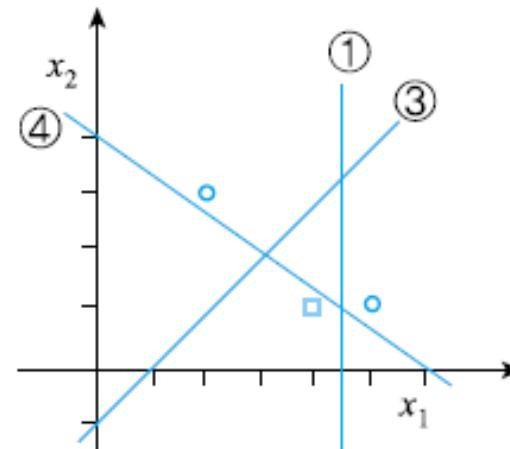
결국 $\alpha_1 = 1/4, \alpha_2 = 1/4, \alpha_3 = 0$

- \mathbf{w} 와 b 를, 그리고 결정 직선

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

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

- 그림에서 직선 ③
- \mathbf{x}_3 은 오분류됨



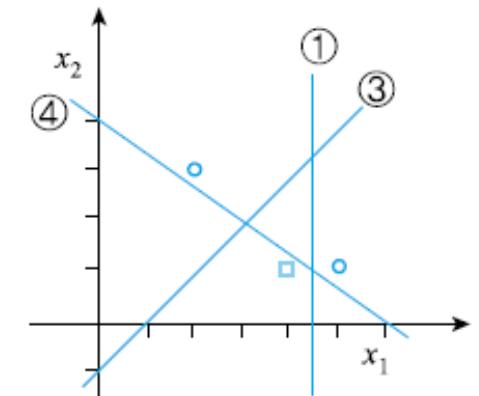
(b) 경우 ①, ③, ④의 결정 직선

선형 SVM | 선형 분리 불가능한 상황

- ④ $\alpha_1 \neq 0, \alpha_2 \neq 0, \alpha_3 \neq 0$ 인 경우를 풀면,
 - $\alpha_1 = \alpha_2 - \alpha_3$ 를 대입한 후,
 $\partial \tilde{L}(\alpha) / \partial \alpha_2 = 0$ 과 $\partial \tilde{L}(\alpha) / \partial \alpha_3 = 0$ 를 풀면
결국 $\alpha_1 = 3/2, \alpha_2 = 13/2, \alpha_3 = 5$ $\mathbf{w} = (2, 3)^T, b = -12$
 $d(x) = 2x_1 + 3x_2 - 12$

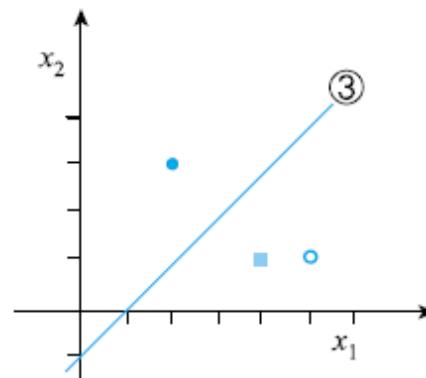
- 그림에서 직선 ④

- 세 개의 샘플 모두 서포트 벡터

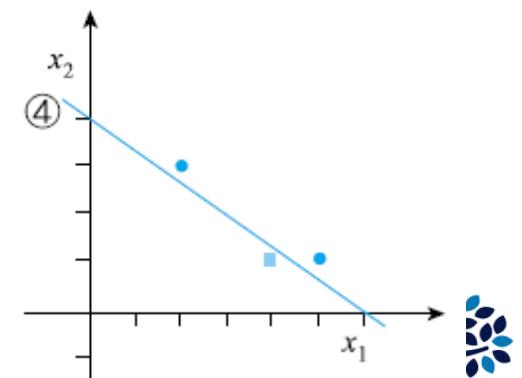


(b) 경우 ①, ③, ④의 결정 직선

- C 에 따른 유효성
 - $C < 2$ 이면 ③만 유효
 - $2 \leq C < 6.5$ 이면 ①③만 유효
 - $6.5 \leq C$ 이면 모두 유효
 - 무엇을 의미하는가?



(c) $C < 6.5$ 일 때 SVM



(d) $C \geq 6.5$ 일 때 SVM



SVM 구현 | 학습과 인식

- SVM 학습이란?
 - (5.1)의 \mathbf{w} 와 b 를 구하는 과정
 - 라그랑제 승수 α 를 구하면 된다.
- 인식은 어떻게 할 것인가?
 - 선형 SVM 선형 SVM 분류기:
$$\left. \begin{array}{l} d(\mathbf{x}) > 0 \text{ } \circ \text{ } \text{면 } \mathbf{x} \in \omega_1, d(\mathbf{x}) < 0 \text{ } \circ \text{ } \text{면 } \mathbf{x} \in \omega_2 \text{ } \text{로 } \text{ } \text{분류하라.} \\ \circ \text{ } \text{때 } d(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \end{array} \right\} \quad (5.36)$$
 - 라그랑제 승수 α 로 \mathbf{w} 구하기
 - 서포트 벡터만 필요
 - (5.37)로 \mathbf{w} 를 미리 구해 놓을 수 있다.

$$\left. \begin{array}{l} \mathbf{w} = \sum_{i=1}^N \alpha_i t_i \mathbf{x}_i \\ = \sum_{\mathbf{x}_k \in Y} \alpha_k t_k \mathbf{x}_k \end{array} \right\} \quad (5.37)$$



SVM 구현 | M 부류 SVM

- M 부류 SVM으로 확장
 - 이제까지는 이진 SVM
 - 1대 $M-1$ 방법
 - $d_j(\mathbf{x})$ 는 부류 ω_j 를 위한 분류기 (ω_j 샘플과 나머지 샘플을 분류함)

$$\left. \begin{array}{l} M \text{ 부류 SVM 분류기:} \\ \mathbf{x} \text{를 } \omega_k \text{로 분류하라.} \\ \text{이 때 } k = \arg \max_j d_j(\mathbf{x}) \end{array} \right\} \quad (5.40)$$

- 1대1 방법
 - 모든 부류 쌍에 대해 이진 분류기 만듦 $M(M-1)/2$ 개의 이진 분류기 필요)
 - 인식 결과를 가지고 투표
- 1대 $M-1$ 방법을 많이 사용



SVM의 특성

- 여백이라는 간단한 아이디어로 breakthrough 이룩함
- SVM의 특성
 - 사용자 설정 매개 변수가 적다.
 - 커널 종류와 커널에 따른 매개 변수
 - (5.15)에서 목적 1 과 목적 2의 가중치 C
 - 최적 커널을 자동 설정하는 방법 없음
 - 실험에 의한 휴리스틱한 선택
 - 일반화 능력 뛰어남
 - 구현이 까다로움
 - OSS 활용
 - SVMlight
 - LIBSVM

