SVM 추가자료

(참고문헌: 패턴인식 5장)

Howon Kim

2017-11-01

들어가는 말

- SVM의 차별성
  - 기존 분류기는 ‘오류율을 최소화’
  - SVM은 ‘여백을margin 최대화’하여 일반화 능력의 극대화 꾀함

- SVM 포털
  - http://support-vector-machines.org/
  - http://support-vector.net/

5.1 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.

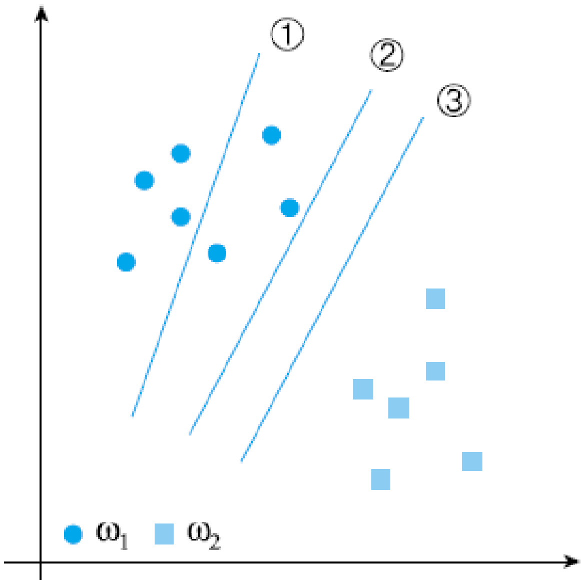


그림 5.1 분류기의 일반화 능력

- 중요한 문제
  - 여백이라는 개념을 어떻게 공식화할 것인가?
  - 여백을 최대로 하는 결정 초평면을 어떻게 찾을 것인가?

5.2 선형 SVM

- 이진 분류를 위한 결정 초평면과 그것의 수학적 특성

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

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 c를 곱하여도 같은 초평면을 나타낸다.
3. $\mathbf{w}$는 초평면의 법선 벡터로서 normal vector 초평면의 방향을 나타내고 b는 위치를 나타낸다.
4. 임의의 점 $\mathbf{x}$에서 초평면까지의 거리는 (5.2)와 같다.

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

5.2 선형 SVM

- 예제 5.1

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

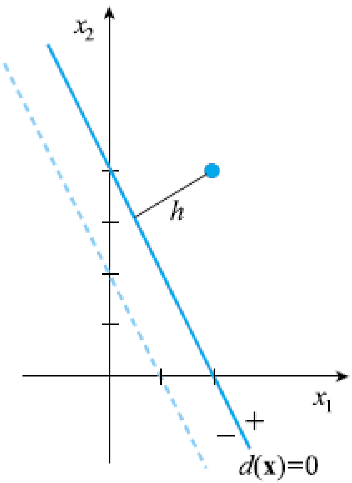


그림 5.2 직선의 수학적 특성

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 $\mathbf{x}=(2,4)^{\mathrm{T}}$에서 직선까지 거리

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

5.2.1 선형 분리 가능한 상황

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

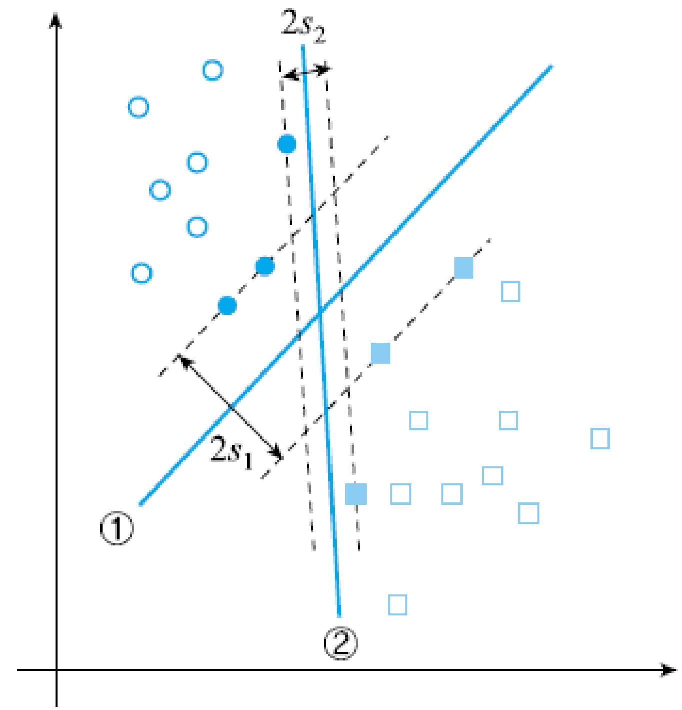


그림 5.3 선형 분리 가능한 상황

5.2.1 선형 분리 가능한 상황

- 이제 문제를 공식화해 보자.

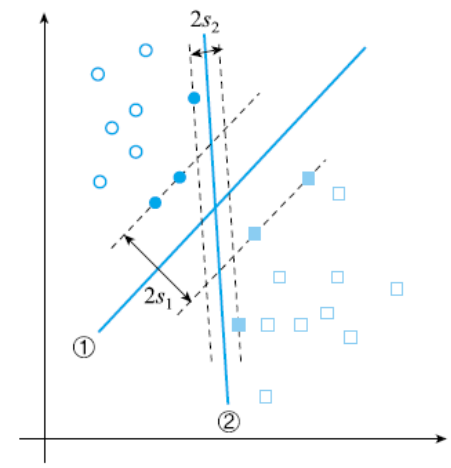
  - 여백을 가장 크게 하는 결정 초평면의 방향, 즉 **w**를 찾아라. (5.3)[2]

  - 그림 5.3에서 ①과 ②는 어느 것이 최적에 가까운가?
  - ①보다 나은 것이 있나?
  - 전형적인 최적화 문제임

5.2.1 선형 분리 가능한 상황

- 여백은 아래와 같이 공식화

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



- 이제 문제를 **조건부 최적화 문제**로 공식화
  - 조건부 최적화 문제

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

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

5.2.1 선형 분리 가능한 상황

- (5.5)를 간단한 형태로 변형하면,
  - 조건부 최적화 문제

$$\left.\begin{array}{l} \text{아래 조건하에,} \\ \quad t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1 \geq 0,\ i = 1, \cdots, N \\ J(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \text{ 을 최소화하라.} \end{array}\right\} \tag{5.6}$$

- 문제의 특성
  - 해의 유일성
    - (5.6)은 볼록이므로 해는 유일하다.
    - 따라서 구한 해는 전역 최적 점을 보장한다.
  - 문제의 난이도
    - N개의 선형 부등식을 조건으로 가진 2차 함수의 최적화 문제
  - 조건부 최적화 문제는 라그랑제 승수로 푼다. (11.2.3 절)

11.2.3 라그랑제 승수

- 등식 조건부 최적화 문제equality constrained optimization problem
  - 등식 조건부 최소화 문제

$f_i(\theta)$ 조건하에서 목적함수 $J(\theta)$를 구함

$$\left.\begin{array}{l}\text{아래 조건하에,}\\ \qquad f_i(\boldsymbol{\theta}) = 0,\ i = 1, \cdots, n \\ J(\boldsymbol{\theta})\text{를 최소화하는 해 } \hat{\boldsymbol{\theta}}\text{를 구하라.}\end{array}\right\} \tag{11.5}$$

- 라그랑제 함수와 라그랑제 승수
  - (11.6)은 라그랑제 함수
  - 조건식마다 라그랑제 승수 λ_i를 할당

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = J(\boldsymbol{\theta}) - \sum_{i=1}^{n} \lambda_i f_i(\boldsymbol{\theta}) \tag{11.6}$$

라그랑제 함수를 도입하는 이유:
기존 목적함수(11.5)는 n개의 조건을 가지는 복잡한 최적화 문제였음
⇒이를 하나의 식으로 변경하여 문제를 품(11.6)

(11.6)을 θ와 λ로미분한 식을 0으로하여(최소화문제이므로), θ구할 수 있음

$$\frac{\partial L}{\partial \theta_i} = 0, i = 1..., k \qquad \frac{\partial L}{\partial \lambda_i} = 0, i = 1..., n$$

11.2.3 라그랑제 승수

- 부등식 조건부 최적화 문제inequality constrained optimization problem
  - 부등식 조건부 최소화 문제

$$\left.\begin{array}{l} \text{아래 조건하에,} \\ \qquad f_i(\boldsymbol{\theta}) \geq 0,\ i=1,\cdots,n \\ J(\boldsymbol{\theta})\text{를 최소화하는 해 }\hat{\boldsymbol{\theta}}\text{를 구하라.} \end{array}\right\} \qquad (11.8)$$

이 문제는 Karush-Kuhn-Tucker (KKT) 조건을 사용하여 푼다. KKT 조건은 (11.9)와 같이 세 종류의 조건으로 구성된다.

$$\left.\begin{array}{l} \dfrac{\partial L(\boldsymbol{\theta},\boldsymbol{\lambda})}{\partial \boldsymbol{\theta}} = 0 \\ \lambda_i \geq 0,\ i=1,\cdots,n \\ \lambda_i f_i(\boldsymbol{\theta}) = 0,\ i=1,\cdots,n \end{array}\right\}$$

5.2.1 선형 분리 가능한 상황

- 라그랑제 승수 방법
  - 목적 함수와 조건을 하나의 식 (즉 라그랑제 함수 L)으로 만들고, KKT조건을 이용하여 라그랑제 함수를을 최적화하는 해를 구한다. (α_i는 라그랑제 승수)

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1) \qquad (5.7)$$

  - KKT 조건이란?

라그랑제 함수 $L(\boldsymbol{\theta}, \boldsymbol{\lambda}) = J(\boldsymbol{\theta}) - \sum_{i=1,N} \lambda_i f_i(\boldsymbol{\theta})$ 에 대한 KKT 조건은

① $\partial L(\boldsymbol{\theta}, \boldsymbol{\lambda}) / \partial \boldsymbol{\theta} = \mathbf{0}$

② $\lambda_i \geq 0, i = 1, \cdots, N$

③ $\lambda_i f_i(\boldsymbol{\theta}) = 0, i = 1, \cdots, N$

5.2.1 선형 분리 가능한 상황

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1) \qquad (5.7)$$

- (5.7)의 KKT 조건

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = \mathbf{0} \quad \rightarrow \quad \mathbf{w} = \sum_{i=1}^{N} \alpha_i t_i \mathbf{x}_i \qquad (5.8)$$

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0 \quad \rightarrow \quad \sum_{i=1}^{N} \alpha_i t_i = 0 \qquad (5.9)$$

$$\alpha_i \geq 0, i = 1, \cdots, N \qquad (5.10)$$

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

즉 support vector가 아닌 point에서는
lagrange multiplier(라그랑 승수)는 모두 0

- (5.11)에 의하면 모든 샘플이 α_i=0 또는 $t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i+b)-1=0$이어야 함. $\alpha_i \neq 0$인 샘플이 서포트 벡터임

Support vector

- (5.8)에 의하면, 라그랑제 승수 α_i 알면 **w** 구할 수 있음 (결정 초평면을 구한 셈)
  - 이제부터 '**w** 구하는 대신 라그랑제 승수 구하는' 문제로 관심 전환
- (5.11)로 b 구할 수 있음

5.2.1 선형 분리 가능한 상황

- 문제의 볼록 성질을 이용하여 풀기 쉬운 형태로 변환
  - 볼록 성질을 만족하는 조건부 최적화 문제는 Wolfe 듀얼로 변형할 수 있다.

볼록 성질을 만족하는 조건부 최적화 문제는 Wolfe 듀얼 문제로 변형할 수 있다. 원래 문제가 $f_i(\boldsymbol{\theta}) \geq 0,\ i = 1, \cdots, N$이라는 조건 하에 $J(\boldsymbol{\theta})$를 최소화하는 것이라 하자. 이때 Wolfe 듀얼 문제는 $\partial L(\boldsymbol{\theta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\theta} = \mathbf{0}$과 $\alpha_i \geq 0,\ i = 1 \cdots, N$이라는 두 가지 조건 하에 $L(\boldsymbol{\theta}, \boldsymbol{\alpha}) = J(\boldsymbol{\theta}) - \sum_{i=1,N} \alpha_i f_i(\boldsymbol{\theta})$를 최대화하는 것이다. 부등식 조건이 등식 조건으로 바뀌었고 최소화 문제가 최대화 문제로 바뀌었다.

  - (5.6)을 Wolfe 듀얼로 바꾸어 쓰면,
    - 조건부 최적화 문제

$$\left.\begin{array}{l}
\text{아래 조건 하에,} \\
\quad \mathbf{w} = \sum_{i=1}^{N} \alpha_i t_i \mathbf{x}_i \\
\quad \sum_{i=1}^{N} \alpha_i t_i = 0 \\
\quad \alpha_i \geq 0,\ i = 1, 2, \ldots, N \\
L(\mathbf{w}, b, \boldsymbol{\alpha}) \text{ 를 최대화하라.}
\end{array}\right\} \qquad (5.12)$$

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1) \qquad (5.7)$$

- 부등식 조건이 등식 조건이 되어 풀기에 유리함

5.2.1 선형 분리 가능한 상황

- 간단한 수식 정리를 하면,
  - 조건부 최적화 문제

$$\left.\begin{aligned}&\text{아래 조건하에,}\\&\quad\sum_{i=1}^{N}\alpha_i t_i=0\\&\quad\alpha_i\geq 0,\ i=1,\cdots,N\\&\tilde{L}(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j t_i t_j \mathbf{x}_i{}^{\mathrm{T}}\mathbf{x}_j \text{ 를 최대화하라.}\end{aligned}\right\}\qquad(5.13)$$

- 흥미로운 특성들
  - 2차 함수의 최대화 문제임
  - $\mathbf{w}$와 b가 사라졌다. (α를 찾는 문제가 되었다.)
  - 특징 벡터 $\mathbf{x}_i$가 내적 형태로 나타난다. (비선형으로 확장하는 발판)
  - 목적 함수의 두번째 $\sum$항은 N^2개의 항을 갖는다. (여전히 풀기 어려운 문제)

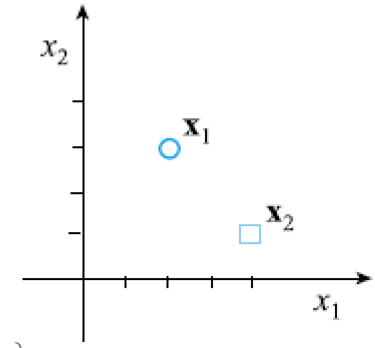
5.2.1 선형 분리 가능한 상황

- 지금까지 한 일을 정리하면,

표 5.1 조건부 최적화 문제의 변형 과정

<table>
<tr><th colspan="3">문제</th><th>특성</th></tr>
<tr><td colspan="3">SVM 철학 (그림 5.1)</td><td>여백을 최대로 하여 일반화 능력 극대화</td></tr>
<tr><td rowspan="6">선
형</td><td rowspan="5">선형
분리
가능</td><td>문제 (5.3)</td><td>SVM 목적과 매개 변수의 관계 설정</td></tr>
<tr><td>조건부 최적화
문제 (5.5)</td><td>조건이 등장함
여백을 공식화하여 목적 함수로 삼음
최대화 문제</td></tr>
<tr><td>조건부 최적화
문제 (5.6)</td><td>t_i를 이용하여 조건을 하나로 씀
목적 함수 $\frac{2}{\|\mathbf{w}\|}$를 $\frac{1}{2}\|\mathbf{w}\|^2$로 바꾸어 최소화 문제로 변형</td></tr>
<tr><td>조건부 최적화
문제 (5.12)</td><td>라그랑제 승수를 도입하여 라그랑제 함수 유도함
문제의 볼록 성질을 이용하여 Wolfe 듀얼 문제로 변형
라그랑제 함수를 최대화하는 문제가 됨
부등식 조건이 등식 조건으로 바뀜</td></tr>
<tr><td>조건부 최적화
문제 (5.13)</td><td>라그랑제 함수를 정리하여 $\mathbf{w}$와 b가 없어지고 α만 남음
$\mathbf{w}$와 b를 구하는 문제가 α를 구하는 문제로 바뀜
특징 벡터가 내적 형태로 나타나 비선형 SVM으로 확장하는 토대가 됨</td></tr>
<tr><td>선형
분리
불가능</td><td>조건부 최적화
문제 (5.27)</td><td>(5.13)을 선형 분리 불가능한 경우로 확장함
$0 \le \alpha$ 조건이 $0 \le \alpha \le C$로 변경된 것 빼고는 (5.13)과 같음</td></tr>
<tr><td>비
선
형</td><td></td><td>조건부 최적화
문제 (5.35)</td><td>(5.27)을 비선형 SVM으로 확장
벡터 내적을 커널 함수 계산으로 대치 (커널 대치 기법)</td></tr>
</table>

5.2.1 선형 분리 가능한 상황



(a) 문제

- 예제 5.2: 두 개 샘플을 가진 경우의 문제 (5.13)의 풀이
  - 훈련집합 $\mathbf{x}_1 = (2,3)^{\mathrm{T}},\ t_1 = 1$

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

아래 조건하에,

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

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

$$\tilde{L}(\boldsymbol{\alpha}) = (\alpha_1 + \alpha_2)$$

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

- 조건부 최적화 문제

아래 조건하에,

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

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

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

  - 실제 값을 대입하면,

아래 조건 하에,

$$\alpha_1 - \alpha_2 = 0$$

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

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

5.2.1 선형 분리 가능한 상황

- 예제 5.2: 두 개 샘플을 가진 경우의 문제 (5.13)의 풀이
  - 정리하여 풀면,

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

$(1/4,1/4)^{\mathrm{T}}$에서 최대값을 가짐

  - (5.8)로 **w**, (5.11)로 b를 구하면

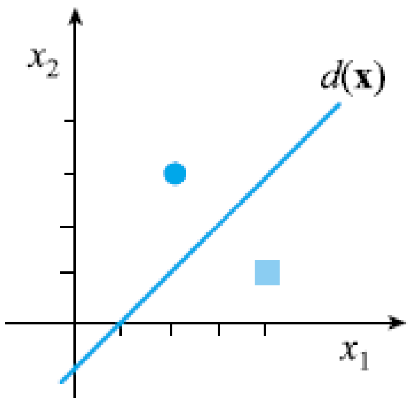
$$\alpha_i(t_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) - 1) = 0,\ i = 1, \cdots, N \qquad (5.11)$$

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

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

  - 결국 결정 직선은

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

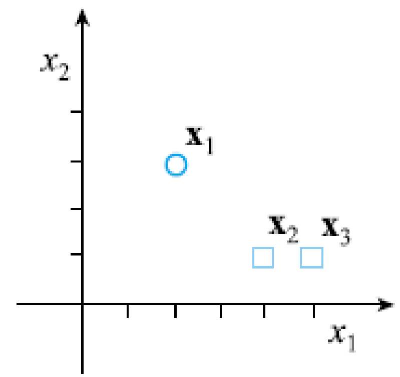


(b) SVM (속이 찬 샘플이 서포트 벡터)

5.2.1 선형 분리 가능한 상황

- 예제 5.3: 세 개 샘플을 가진 경우의 문제 (5.13)의 풀이
  - 훈련집합

$$\mathbf{x}_1=(2,3)^{\mathrm{T}},\ t_1=1$$
$$\mathbf{x}_2=(4,1)^{\mathrm{T}},\ t_2=-1$$
$$\mathbf{x}_3=(5,1)^{\mathrm{T}},\ t_3=-1$$



(a) 문제

  - (5.13)에 실제 값을 대입하면,

아래 조건 하에,

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

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

최대화하는 $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3)^{\mathrm{T}}$을 찾아라.

- 이 문제를 어떻게 풀 것인가?

• 조건부 최적화 문제

아래 조건하에,

$$\sum_{i=1}^{N}\alpha_i t_i=0$$
$$\alpha_i\geq 0, i=1,\cdots,N$$
$$\tilde{L}(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j t_i t_j \mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j \text{ 를 최대화하라.}$$

5.2.1 선형 분리 가능한 상황

- 예제 5.3: 세 개 샘플을 가진 경우의 문제 (5.13)의 풀이
  - 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

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

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

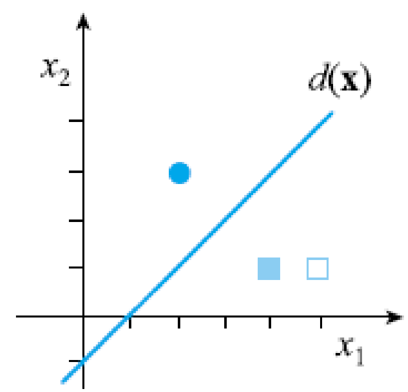
    - (5.8)과 (5.11)로 $\mathbf{w}$와 b를 구하면

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

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

    - 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)

5.2.2 선형 분리 불가능한 상황

- 선형 분리 불가능한 상
  - 샘플 $(\mathbf{x},t)$의 세 가지 상

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

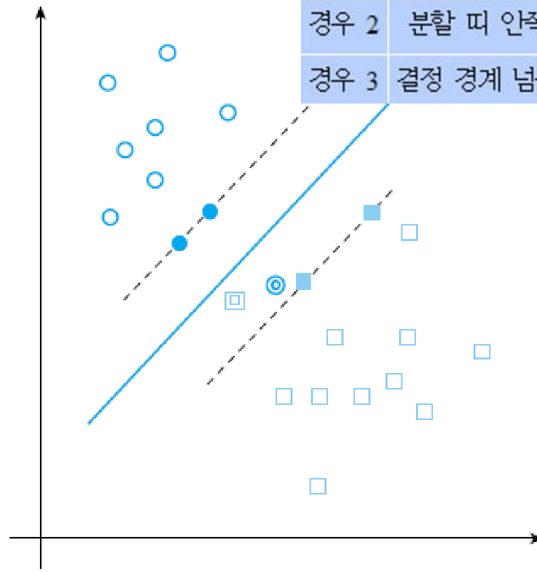


그림 5.6 선형 분리 불가능한 상황의 SVM

- 경우 1: 분할 띠의 바깥에 있다. $1 \le t(\mathbf{w}^\mathrm{T}\mathbf{x}+b)$ 를 만족한다.
- 경우 2: 분할 띠의 안쪽에 있는데 자기가 속한 부류의 영역에 있다.

  $0 \le t(\mathbf{w}^\mathrm{T}\mathbf{x}+b) < 1$ 을 만족한다.
- 경우 3: 결정 경계를 넘어 자신이 속하지 않은 부류의 영역에 놓여 있다.

  $t(\mathbf{w}^\mathrm{T}\mathbf{x}+b) < 0$ 을 만족한다.

5.2.2 선형 분리 불가능한 상황

- 슬랙 변수 ξ를 도입하여 하나의 식으로 쓰면,

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

표 5.2 선형 분리 불가능한 상황에서 샘플의 세 가지 경우

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

5.2.2 선형 분리 불가능한 상황

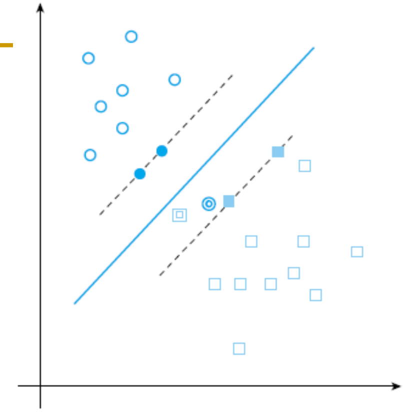


그림 5.6 선형 분리 불가능한 상황의 SVM

- 문제 공식화
  - 길항tradeoff 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 **w**를 찾아라. (5.15)
  - 목적 함수를 다시 쓰면

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

- 첫번째 항은 목적 1, 두번째 항은 목적 2

5.2.2 선형 분리 불가능한 상황

- 문제 공식화
  - 조건부 최적화 문제

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

아래 조건 하에,

$$t_i(\mathbf{w}^T\mathbf{x}_i + b) \ge 1 - \xi_i,\ i = 1, \cdots, N$$
$$\xi_i \ge 0,\ i = 1, \cdots, N$$

$J(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$ 를 최소화하라.

5.2.2 선형 분리 불가능한 상황

- 라그랑제 승수로 풀어보면,
  - 라그랑제 함수

$$L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \left(\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i\right) - \left(\sum_{i=1}^{N}\alpha_i(t_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i) + \sum_{i=1}^{N}\beta_i\xi_i\right) \quad (5.18)$$

  - KKT 조건

$$\mathbf{w} = \sum_{i=1}^{N}\alpha_i t_i \mathbf{x}_i \quad (5.19)$$

$$\sum_{i=1}^{N}\alpha_i t_i = 0 \quad (5.20)$$

$$C = \alpha_i + \beta_i \quad (5.21)$$

$$\alpha_i \ge 0,\ i = 1, \cdots, N \quad (5.22)$$

$$\beta_i \ge 0,\ i = 1, \cdots, N \quad (5.23)$$

$$\alpha_i(t_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i) = 0,\ i = 1, \cdots, N \quad (5.24)$$

$$\beta_i\xi_i = 0,\ i = 1, \cdots, N \quad (5.25)$$

5.2.2 선형 분리 불가능한 상황

- 라그랑제 승수로 풀어보면,
  - Wolfe 듀얼로 변형
    - 조건부 최적화 문제

$$\left.\begin{aligned}
&\text{아래 조건 하에,}\\
&\qquad \mathbf{w} = \sum_{i=1}^{N} \alpha_i t_i \mathbf{x}_i\\
&\qquad \sum_{i=1}^{N} \alpha_i t_i = 0\\
&\qquad C = \alpha_i + \beta_i\\
&\qquad \alpha_i \geq 0,\ i = 1, 2, \ldots, N\\
&\qquad \beta_i \geq 0,\ i = 1, 2, \ldots, N\\
&L(\mathbf{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \text{ 을 최대화하라.}
\end{aligned}\right\} \qquad (5.26)$$

5.2.2 선형 분리 불가능한 상황

- 라그랑제 승수로 풀어보면,
  - (5.26)을 정리하면

    - 조건부 최적화 문제 (선형 SVM)

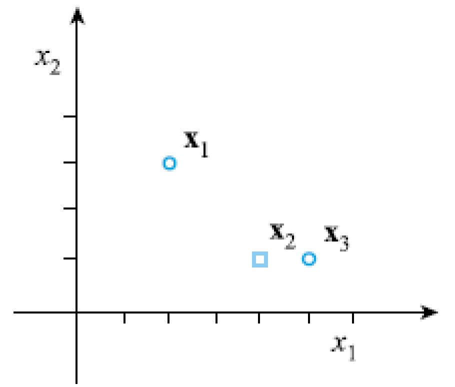
$$
\left.\begin{array}{l}
\text{아래 조건하에,} \\
\qquad \sum_{i=1}^{N} \alpha_i t_i = 0 \\
\qquad 0 \le \alpha_i \le C,\ i = 1, \cdots, N \\
\tilde{L}(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j t_i t_j \mathbf{x}_i{}^{\mathrm{T}} \mathbf{x}_j \text{ 를 최대화하라.}
\end{array}\right\} \qquad (5.27)
$$

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

5.2.2 선형 분리 불가능한 상황

- 예제 5.4: 세 개 샘플을 가진 경우의 문제 (5.27)의 풀이
  - 훈련집합 (선형 분리 가능한가?)

$$\mathbf{x}_1 = (2,3)^{\mathrm{T}},\ \ t_1 = 1$$
$$\mathbf{x}_2 = (4,1)^{\mathrm{T}},\ \ t_2 = -1$$
$$\mathbf{x}_3 = (5,1)^{\mathrm{T}},\ \ t_3 = 1$$



(a) 문제

  - (5.27)에 실제 값을 대입하면,

아래 조건 하에,

$$\alpha_1 - \alpha_2 + \alpha_3 = 0$$
$$0 \le \alpha_1 \le C,\ \ 0 \le \alpha_2 \le C, 0 \le \alpha_3 \le C$$

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

최대화하는 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^{\mathrm{T}}$을 찾아라.

- 이 문제를 어떻게 풀 것인가?

5.2.2 선형 분리 불가능한 상황

- 예제 5.4: 세 개 샘플을 가진 경우의 문제 (5.27)의 풀이
  - 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

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

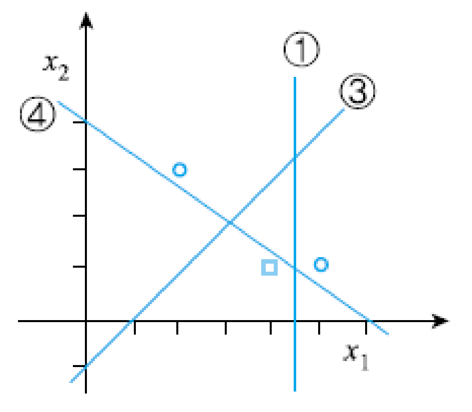
결국 $\alpha_1=0, \alpha_2=2, \alpha_3=2$

    - $\mathbf{w}$와 b를, 그리고 결정 직선

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

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

  - 그림에서 직선 ①
    - $\mathbf{x}_1$은 오분류됨



(b) 경우 ①, ③, ④의 결정 직선

5.2.2 선형 분리 불가능한 상황

- 예제 5.4: 세 개 샘플을 가진 경우의 문제 (5.27)의 풀이
  - 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

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

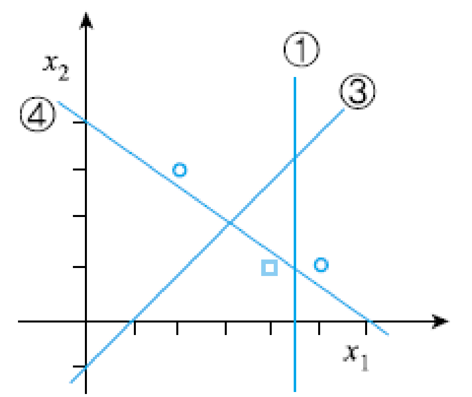
결국 $\alpha_1 = 1/4$, $\alpha_2 = 1/4$, $\alpha_3 = 0$

- $\mathbf{w}$와 b를, 그리고 결정 직선

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

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

  - 그림에서 직선 ③
- $\mathbf{x}_3$은 오분류됨



(b) 경우 ①, ③, ④의 결정 직선

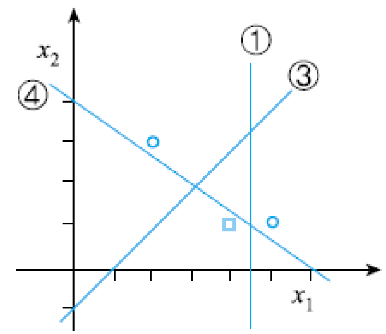
5.2.2 선형 분리 불가능한 상황

- 예제 5.4: 세 개 샘플을 가진 경우의 문제 (5.27)의 풀이
  - 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,
      $\partial \tilde{L}(\boldsymbol{\alpha})/\partial \alpha_2 = 0$과 $\partial \tilde{L}(\boldsymbol{\alpha})/\partial \alpha_3 = 0$ 를 풀면

      결국 $\alpha_1=3/2$, $\alpha_2=13/2$, $\alpha_3=5$ $\quad \mathbf{w} = (2,3)^{\mathrm{T}},\ b = -12$

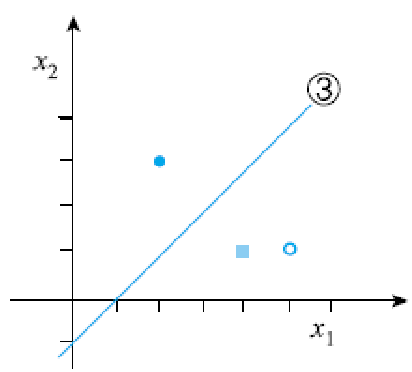
      $$d(x) = 2x_1 + 3x_2 - 12$$

      - 그림에서 직선 ④
    - 세 개의 샘플 모두 서포트 벡터

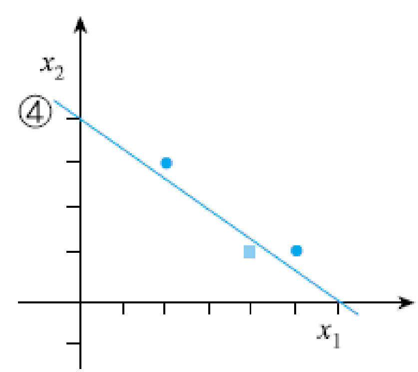


(b) 경우 ①, ③, ④의 결정 직선

  - C에 따른 유효성
    - C<2이면 ③만 유효
    - 2≤C<6.5이면 ①③만 유효
    - 6.5≤C이면 모두 유효
    - 무엇을 의미하는가?



(c) $C<6.5$일 때 SVM



(d) $C \geq 6.5$일 때 SVM

Vladimir Vapnik
러시아

Vapnik은 SVM을 창안한 사람으로 유명하다. 현재 SVM은 일반화 능력이 가장 뛰어난 분류기로 인정 받고 있다. 패턴 인식에서 기술 돌파로 breakthrough 평가되고 있는 것이다. 그는 러시아에서 태어났으며 모스코바에 있는 제어 과학 연구원에서 Institute of Control Sciences 통계학으로 박사 학위를 취득하였다. 그 후 이 연구원에서 1990년까지 근무하였고 이후 미국 AT&T로 이적하였고 주로 미국에서 연구 활동을 하였다. 현재는 Columbia 대학과 London 대학의 교수이다. SVM을 포함하여 통계적 학습 이론을 정리한 책이 그의 대표적인 저서 중의 하나이다 [Vapnik98].

[Vapnik98] Vladimir Vapnik, *Statistical Learning Theory*, John Wiley and Sons, 1998.