

기계학습 개론

- Probabilistic Classification (2/2)

정보통신공학과

Prof. Jinkyu Kang



MYONGJI
UNIVERSITY

이번주 수업의 목차

- 확률적 분류 방법
 - 베이시언 결정 이론
 - 확률과 통계
 - 베이시언 분류기
 - 분별 함수
 - 정규 분포에서 베이시언 분류기
 - 베이시언 분류의 특성
 - 확률 분포 추정
 - 히스토그램 추정
 - 최대 우도
 - 비모수적 방법
 - 혼합 모델



정규분포에서 베이시언 분류기 I

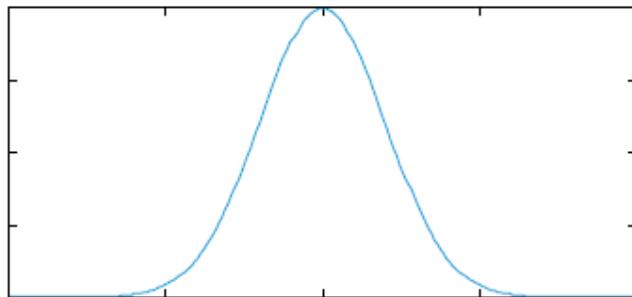
- 정규 분포 (가우시언 분포)
 - 현실 세계에 맞는 경우 있음
 - 평균과 분산이라는 두 종류의 매개 변수만으로 표현 가능
 - 수학적인 매력
- 우도가 정규 분포를 따른다는 가정 하에 베이시언 분류기의 특성을 해석해 보자.

정규분포에서 베이시언 분류기 I

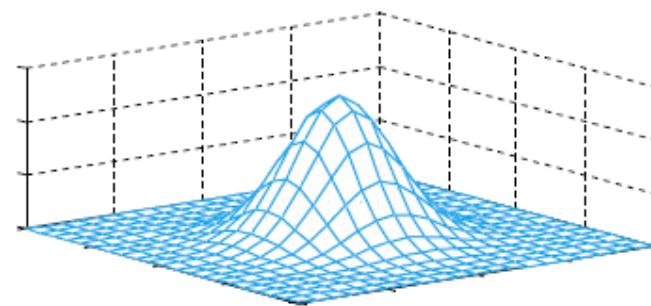
정규분포와 분별 함수

- 정규 분포 (가우시언 분포)

$$\left. \begin{aligned} N(\mu, \sigma^2) &= \frac{1}{(2\pi)^{1/2} \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ N(\mu, \Sigma) &= \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right) \end{aligned} \right\} \quad (2.32)$$



(a) 1 차원 정규 분포



(b) 2 차원 정규 분포

그림 2.9 정규 분포의 예



정규분포에서 베이시언 분류기 I

정규분포와 분별 함수

- 우도를 다시 쓰면,

$$p(\mathbf{x} | \omega_i) = N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\right) \quad (2.33)$$

- 로그를 취하여 분별 함수를 만들어 보면,

$$\begin{aligned} g_i(\mathbf{x}) &= \ln(p(\mathbf{x} | \omega_i) P(\omega_i)) \\ &= \ln N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + \ln P(\omega_i) \\ &= -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i) \\ &= -\frac{1}{2} \left(\mathbf{x}^T \boldsymbol{\Sigma}_i^{-1} \mathbf{x} - \mathbf{x}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i - \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{x} + \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i \right) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i) \\ &= -\frac{1}{2} \left(\mathbf{x}^T \boldsymbol{\Sigma}_i^{-1} \mathbf{x} - 2\boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{x} + \boldsymbol{\mu}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\mu}_i \right) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i) \end{aligned} \quad \left. \right\} (2.34)$$

- $g_i(\mathbf{x})$ 는 변수 \mathbf{x} 에 대한 2차식

정규분포에서 베이시언 분류기 I

정규분포와 분별 함수

- 예제 2.4
 - $\sigma^2=2$ 이고 아래와 같다고 가정

부류 ω_i 의 정규 분포가 $\mu_i = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ 와 $\Sigma_i = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

- 분별 함수를 유도해 보면,

$$\begin{aligned}g_i(\mathbf{x}) &= -\frac{1}{2} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right)^T \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right) - \frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \left| \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \right| + \ln P(\omega_i) \\&= -\frac{1}{2} (x_1 - 3 \quad x_2 - 1) \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 - 3 \\ x_2 - 1 \end{pmatrix} - \ln 2\pi - \frac{1}{2} \ln 4 + \ln P(\omega_i) \\&= -\frac{1}{4} \underbrace{(x_1^2 + x_2^2)}_{2\text{차식}} + \frac{1}{2} \underbrace{(3x_1 + x_2)}_{1\text{차식}} - \frac{1}{2} \underbrace{(5 + 2\ln 2\pi + \ln 4 - 2\ln P(\omega_i))}_{상수}\end{aligned}$$

■



정규분포에서 베이시언 분류기 I

정규분포와 분별 함수

- 결정 경계
 - 두 부류가 차지하는 영역의 경계
 - $g_i(\mathbf{x}) = g_j(\mathbf{x})$ 인 점
 - 즉 $g_{ij}(\mathbf{x}) = 0$ 인 점

$$g_{ij}(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) \quad (2.35)$$

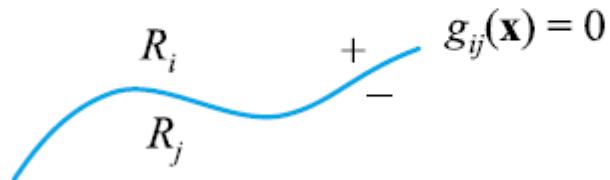


그림 2.10 두 부류 ω_i 와 ω_j 의 결정 경계



정규분포에서 베이시언 분류기 I

선형 분별

- 모든 부류의 공분산 행렬이 같은 상황, $\Sigma_i = \Sigma$ 로 표기
 - 분별 함수를 다시 쓰면

$$g_i(\mathbf{x}) = \frac{1}{2} \underbrace{\left(2\boldsymbol{\mu}_i^T \Sigma^{-1} \mathbf{x} - \boldsymbol{\mu}_i^T \Sigma^{-1} \boldsymbol{\mu}_i + 2 \ln P(\omega_i) \right)}_{i에 따라 다름} - \frac{1}{2} \underbrace{\left(\mathbf{x}^T \Sigma^{-1} \mathbf{x} + d \ln 2\pi + \ln |\Sigma| \right)}_{i에 무관함} \quad (2.36)$$

- 에 무관한 항은 제거해도 됨. 따라서 2차 항 $\mathbf{x}^T \Sigma^{-1} \mathbf{x}$ 없어짐
- 선형식이 됨

$$\begin{aligned} g_i(\mathbf{x}) &= \left(\Sigma^{-1} \boldsymbol{\mu}_i \right)^T \mathbf{x} + \left(\ln P(\omega_i) - \frac{1}{2} \boldsymbol{\mu}_i^T \Sigma^{-1} \boldsymbol{\mu}_i \right) \\ &= \mathbf{w}_i^T \mathbf{x} + b_i \end{aligned} \quad (2.37)$$



정규분포에서 베이시언 분류기 I

선형 분별

- 결정 경계

$$\begin{aligned}
 g_{ij}(\mathbf{x}) &= g_i(\mathbf{x}) - g_j(\mathbf{x}) \\
 &= (\Sigma^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j))^T \mathbf{x} + \left(\ln P(\omega_i) - \ln P(\omega_j) - \frac{1}{2} \boldsymbol{\mu}_i^T \Sigma^{-1} \boldsymbol{\mu}_i + \frac{1}{2} \boldsymbol{\mu}_j^T \Sigma^{-1} \boldsymbol{\mu}_j \right) \\
 &= \underbrace{(\Sigma^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j))^T}_{\mathbf{w}} \left(\mathbf{x} - \underbrace{\left(\frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\boldsymbol{\mu}_i - \boldsymbol{\mu}_j}{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \Sigma^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)} \ln \frac{P(\omega_i)}{P(\omega_j)} \right)}_{\mathbf{x}_0} \right) \\
 &= \mathbf{w}^T (\mathbf{x} - \mathbf{x}_0)
 \end{aligned} \tag{2.38}$$

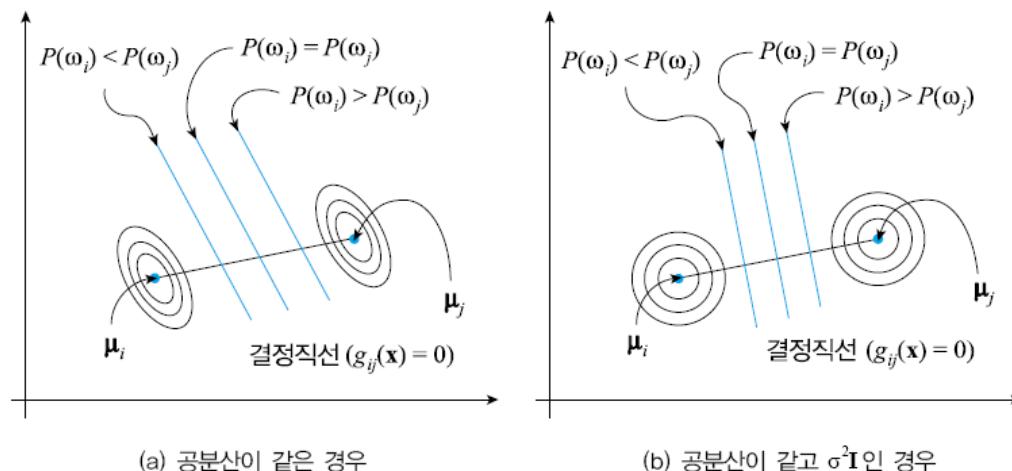


그림 2.11 선형 분류기

정규분포에서 베이시언 분류기 |

선형 분별

- 예제 2.5 $\omega_1: (1,2)^T, (3,1)^T, (5,2)^T, (3,3)^T$
 $\omega_2: (6,6)^T, (8,5)^T, (10,6)^T, (8,7)^T$

$$\mu_1 = \frac{1}{4} \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 5 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} \right) = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$\mu_2 = \frac{1}{4} \left(\begin{pmatrix} 6 \\ 6 \end{pmatrix} + \begin{pmatrix} 8 \\ 5 \end{pmatrix} + \begin{pmatrix} 10 \\ 6 \end{pmatrix} + \begin{pmatrix} 8 \\ 7 \end{pmatrix} \right) = \begin{pmatrix} 8 \\ 6 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

$$\begin{aligned} g_{12}(x) &= \left(\begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix} \begin{pmatrix} 3-8 \\ 2-6 \end{pmatrix} \right)^T x + \left(\ln P(\omega_1) - \ln P(\omega_{12}) - \frac{1}{2} (3-2) \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} + \frac{1}{2} (8-6) \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 2 \end{pmatrix} \begin{pmatrix} 8 \\ 6 \end{pmatrix} \right) \\ &= (-5/2 \quad -8) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (\ln P(\omega_1) - \ln P(\omega_2) + 45.75) \\ &= -\frac{5}{2}x_1 - 8x_2 + (\ln P(\omega_1) - \ln P(\omega_2) + 45.75) \end{aligned}$$

$P(\omega_1) = 0.5, P(\omega_2) = 0.5$ 인 경우: $5x_1 + 16x_2 - 91.5 = 0$

$P(\omega_1) = 0.8, P(\omega_2) = 0.2$ 인 경우: $5x_1 + 16x_2 - 94.273 = 0$

$P(\omega_1) = 0.2, P(\omega_2) = 0.8$ 인 경우: $5x_1 + 16x_2 - 88.727 = 0$

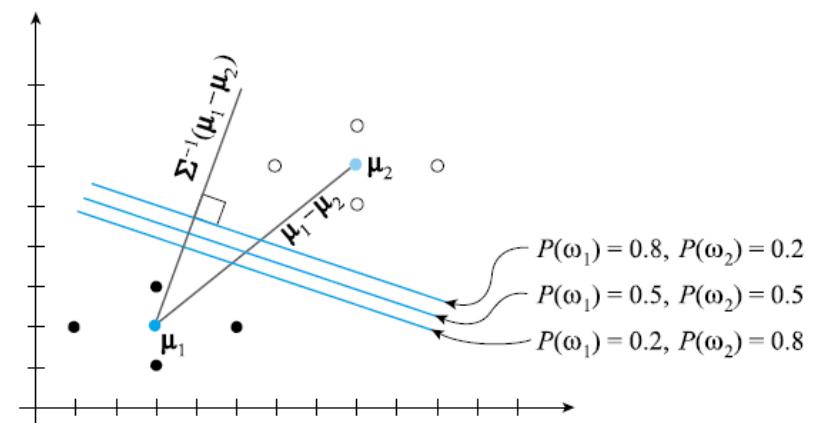


그림 2.12 선형 분별 분석 (LDA) 사례



정규분포에서 베이시언 분류기 I

2차 분별

- 임의의 공분산 행렬

$$g_i(\mathbf{x}) = \underbrace{-\frac{1}{2} \mathbf{x}^T \Sigma_i^{-1} \mathbf{x}}_{2차 항} + \underbrace{\mu_i^T \Sigma_i^{-1} \mathbf{x}}_{1차 항} + \underbrace{\left(-\frac{1}{2} \mu_i^T \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i) \right)}_{상수 항} \quad (2.40)$$

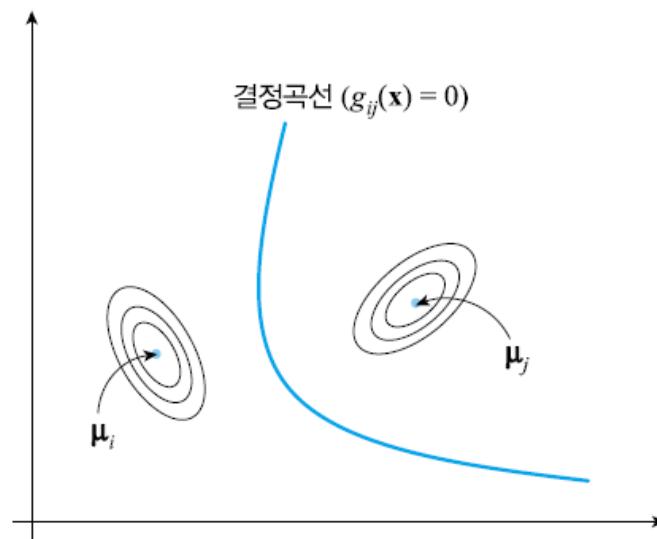


그림 2.13 2차 분별 분석으로 만든 2차 분류기

정규분포에서 베이시언 분류기 I

2차 분별

- 예제 2.6

$$\begin{aligned}\omega_1: (1,2)^T, (3,1)^T, (5,2)^T, (3,3)^T \quad \mu_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \\ \omega_2: (7,6)^T, (8,4)^T, (9,6)^T, (8,8)^T \quad \mu_2 = \begin{pmatrix} 8 \\ 6 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}\end{aligned}$$

$$g_{12}(x) = g_1(x) - g_2(x) = 0.75x_1^2 - 0.75x_2^2 - 14.5x_1 + x_2 + 66.75 + \ln P(\omega_1) - \ln P(\omega_2)$$

$$P(\omega_1) = 0.8, P(\omega_2) = 0.2 \text{ 경우: } 3x_1^2 - 3x_2^2 - 58x_1 + 4x_2 + 274.3936 = 0$$

$$P(\omega_1) = 0.5, P(\omega_2) = 0.5 \text{ 경우: } 3x_1^2 - 3x_2^2 - 58x_1 + 4x_2 - 267 = 0$$

$$P(\omega_1) = 0.2, P(\omega_2) = 0.8 \text{ 경우: } 3x_1^2 - 3x_2^2 - 58x_1 + 4x_2 + 259.6064 = 0$$

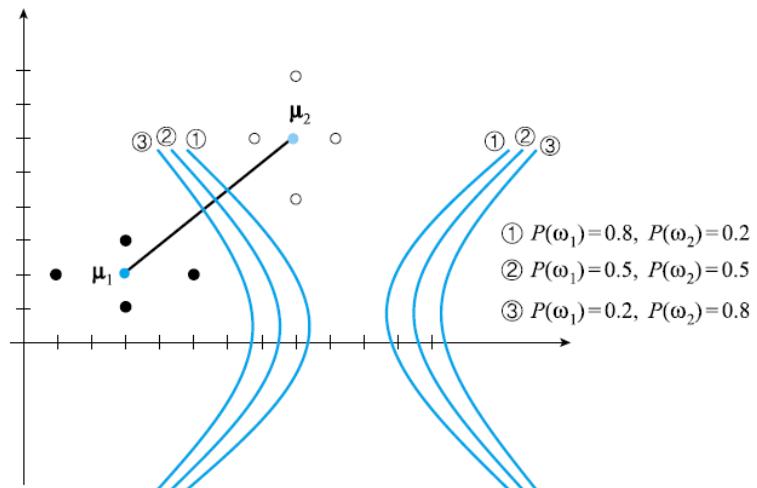


그림 2.14 2차 분별 분석 (QDA) 사례

정규분포에서 베이시언 분류기 I

최소 거리 분류기

- 최소 거리 분류기로 다시 해석해 보자.
 - 수식 유도 편의를 위해 두 부류의 사전 확률과 공분산 행렬 같다고 가정

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \quad (2.41)$$

- 최소 거리 분류기

$$\mathbf{x} \text{를 } k = \arg \min_i \left((\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right)^{1/2} \text{ 일 때 } \omega_k \text{로 분류하라.} \quad (2.44)$$

- 거리 척도

$$\text{마할라노비스 거리: } \left((\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right)^{1/2} \quad (2.42)$$

$$\text{유클리디언 거리: } \left((\mathbf{x} - \boldsymbol{\mu}_i)^T (\mathbf{x} - \boldsymbol{\mu}_i) \right)^{1/2} = \|\mathbf{x} - \boldsymbol{\mu}_i\| \quad (2.43)$$



정규분포에서 베이시언 분류기 I

최소 거리 분류기

- 예제 2.7

$$\omega_1: (1,2)^T, (3,1)^T, (5,2)^T, (3,3)^T$$

$$\omega_2: (6,6)^T, (8,5)^T, (10,6)^T, (8,7)^T$$

$$\mathbf{x} = (8,2)^T$$

$$\mu_1 \text{까지의 마할라노비스 거리} = \left((8-3) \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix} (8-3) \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \right)^{1/2} = 3.536$$

$$\mu_2 \text{까지의 마할라노비스 거리} = \left((8-8) \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix} (8-8) \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \right)^{1/2} = 5.657$$

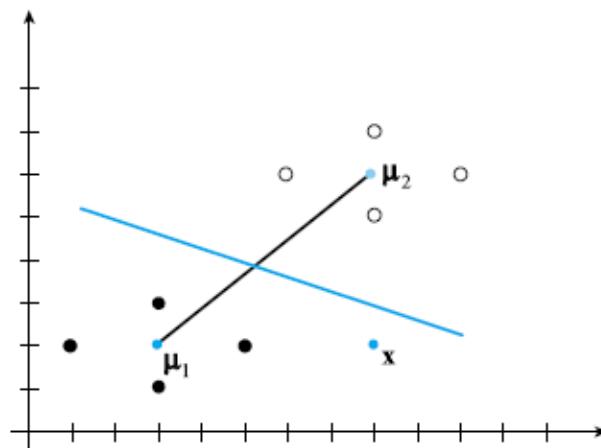


그림 2.15 두 통계 분포까지의 마할라노비스 거리

베이시언 분류의 특성 |

- 베이시언 분류의 특성

1. 비현실성을 내포하고 있다. 이미 언급한 바와 같이 일반적인 확률 분포를 사용 하려 하면 차원의 저주가 발생한다. 정규 분포를 가정하면 실제 확률 분포와 차이가 발생한다.
2. 1의 비현실성을 무시하고 실제 확률 분포를 있다고 가정하면 베이시언 분류기는 오류율 측면에서 최적이다. 이 최적성을 수학적으로 완벽하게 증명할 수 있다.
3. 베이시언 분류기는 M 개 부류 각각에 대해 그에 속할 확률을 출력한다. 즉 (2.30)의 베이시언 분류기에서 $g_j(\mathbf{x})$ 를 확률로 해석할 수 있다. 따라서 이 확률을 신뢰도 값으로 삼아 후처리 등에 유용하게 활용할 수 있다.



베이시언 분류의 특성 I

- 나이브^{naïve} 베이시언 분류기
 - 특징들이 서로 독립이라는 가정

$$p(\mathbf{x} | \omega_i) = \prod_{j=1}^d p(x_j | \omega_i) \quad (2.45)$$

- 우도 계산을 (2.45)로 하는 분류기를 나이브 베이시언 분류기라 함
- 얻은 것: 차원의 저주를 피함
- 잃은 것: 성능 저하



이번주 수업의 목차

- **확률적 분류 방법**

- 베이시언 결정 이론
 - 확률과 통계
 - 베이시언 분류기
 - 분별 함수
 - 정규 분포에서 베이시언 분류기
 - 베이시언 분류의 특성
 - 기각 처리

- **확률 분포 추정**

- 히스토그램 추정
 - 최대 우도
 - 비모수적 방법
 - 혼합 모델



들어가는 말

- 베이시언 분류에서의 학습은 사전 확률과 우도의 추정

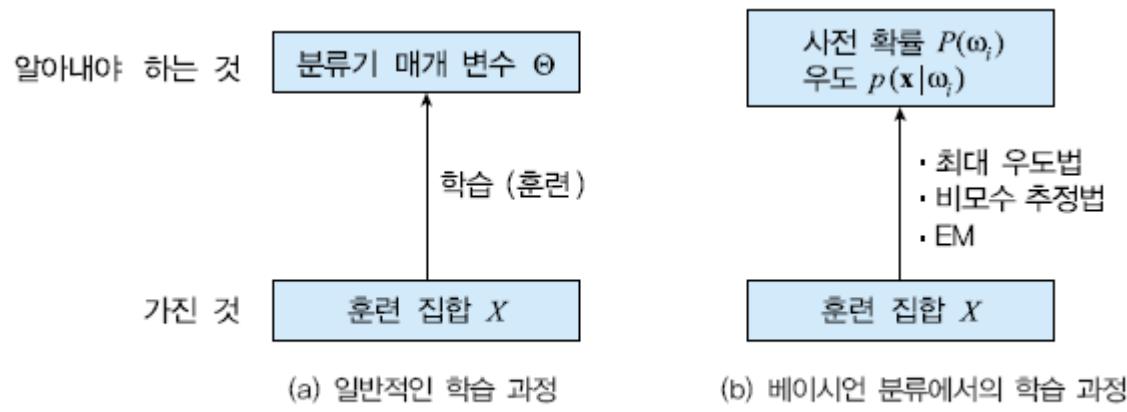


그림 3.1 일반적인 학습 과정과 베이시언 분류에서의 학습

들어가는 말

- 사전 확률 $P(\omega_i)$ 의 추정

$$P(\omega_i) = N_i / N \quad (3.1)$$

- N 은 X 의 크기이고 N_i 는 ω_i 에 속하는 샘플 수
- N 이 충분히 크면 (3.1)은 실제 값에 근접
- 우도 $P(\mathbf{x}|\omega_i)$ 추정

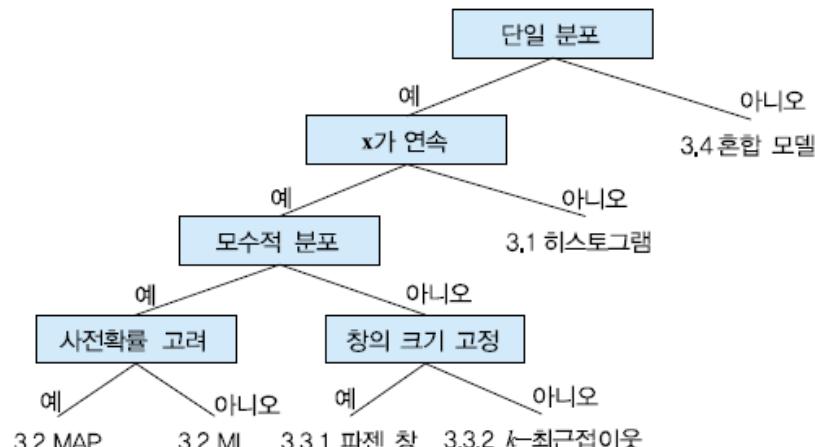


그림 3.2 여러 방법을 구별하는 트리

히스토그램

- 히스토그램

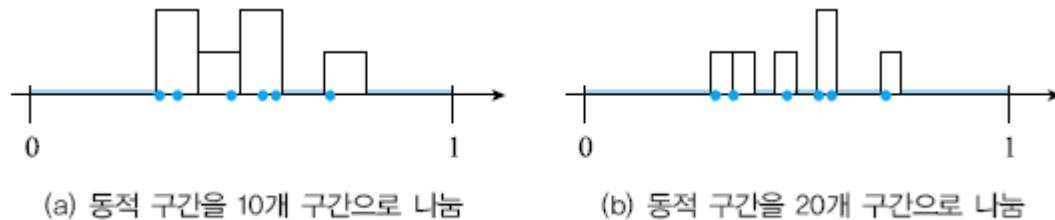


그림 3.3 1 차원에서 히스토그램 추정 사례

- 총 s^d 개의 빈이 발생 (각 차원을 s 개 구간으로 나눈다 했을 때)
 - 전형적인 차원의 저주
 - N 은 충분히 크고 d 는 작아야 함

최대 우도

- 문제 정의
 - “주어진 X 를 발생시켰을 가능성이 가장 높은 매개 변수 Θ 를 찾아라.”
 - 주어진 X 에 대해 가장 큰 우도를 갖는 Θ 를 찾아라.
 - 아래 예에서
 - $P(X|\Theta_1) > P(X|\Theta_2)$
 - 최대 우도를 갖는 Θ 는?

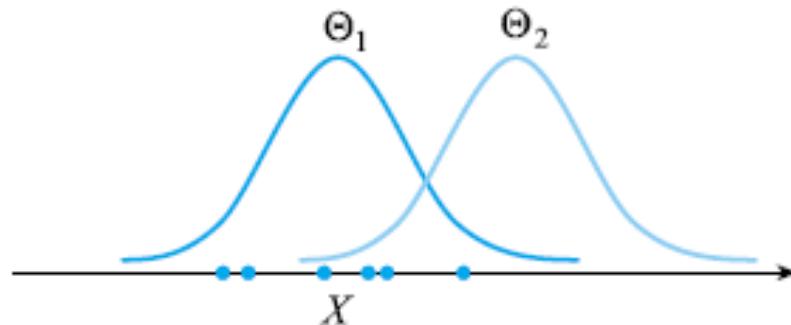


그림 3.4 최대 우도를 갖는 Θ 를 찾는 문제

최대 우도

- 최대 우도 ML 방법
 - 아래 최적화 문제를 풀어 답을 구하는 방법

$$\hat{\Theta} = \arg \max_{\Theta} p(X | \Theta) \quad (3.2)$$

$$p(X | \Theta) = p(\mathbf{x}_1 | \Theta) p(\mathbf{x}_2 | \Theta) \cdots p(\mathbf{x}_N | \Theta) = \prod_{i=1}^N p(\mathbf{x}_i | \Theta) \quad (3.3)$$

- 로그 우도로 바꾸면

$$\hat{\Theta} = \arg \max_{\Theta} \sum_{i=1}^N \ln p(\mathbf{x}_i | \Theta) \quad (3.4)$$

- 미분을 이용한 최적화 문제 풀이
 - $L(\Theta)$ 의 도함수를 0으로 두고 풀어 구한 답이 $\hat{\Theta}$

$$\left. \begin{array}{l} \frac{\partial L(\Theta)}{\partial \Theta} = 0 \\ \text{or 때 } L(\Theta) = \sum_{i=1}^N \ln p(\mathbf{x}_i | \Theta) \end{array} \right\} \quad (3.5)$$

최대 우도

- MAP 방법
 - $P(\Theta)$ 가 균일하지 않은 경우

$$\hat{\Theta} = \arg \max_{\Theta} \ln p(\Theta) + \sum_{i=1}^N \ln p(\mathbf{x}_i | \Theta) \quad (3.7)$$

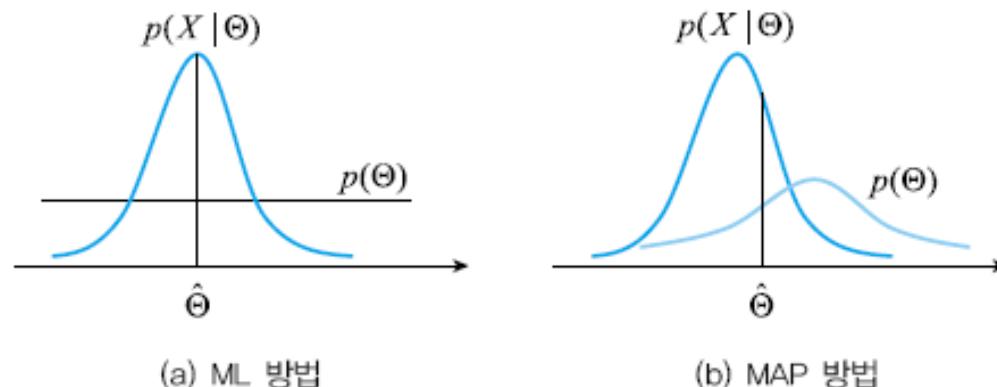


그림 3.5 ML과 MAP의 비교

비모수적 방법

- 확률 분포 추정 방법
 - 모수적 방법
 - 확률 분포가 매개 변수 (모수)로 표현되는 형태
 - ML,MAP 방법등
 - 비모수적 방법
 - 확률 분포가 임의의 형태
 - 파젠 창, k -최근접 이웃 추정 방법 등



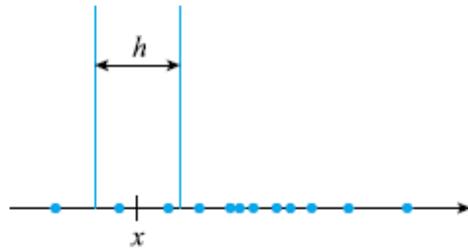
비모수적 방법 | 파젠 창

- 히스토그램 방법을 확장하여 확률 밀도 함수^{pdf} 추정
 - 그림 3.6에서 임의의 점 x 에서 확률 값 추정
 - 크기 h 인 창을 씌우고 그 안의 샘플의 개수를 k 라 하면,

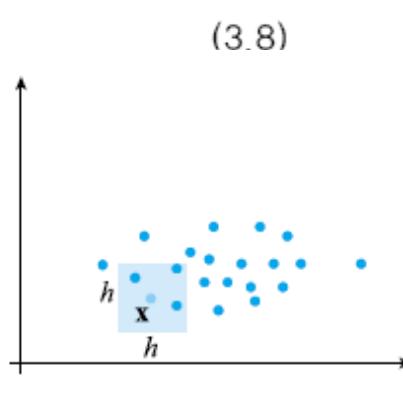
$$p(x) = \frac{1}{h} \frac{k}{N}$$

- d 차원으로 확대하면,

$$p(\mathbf{x}) = \frac{1}{h^d} \frac{k_x}{N}$$



(a) 1 차원 특징 공간



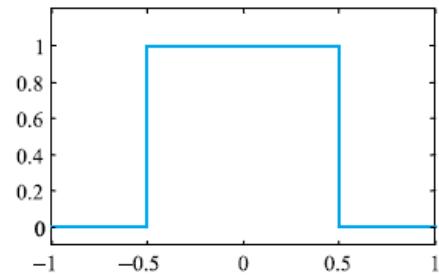
(b) 2 차원 특징 공간

그림 3.6 파젠 창에 의한 확률 밀도 추정

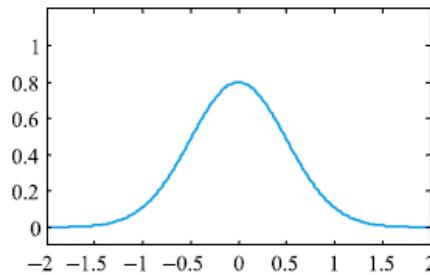
비모수적 방법 | 파젠 창

- 여전히 매끄럽지 않은 함수
 - 예를 들어 그림 3.6(a)에서 x 를 오른쪽 옮기면 계속 두 개이다가 어느 순간에 3으로 바뀜. 따라서 불연속인 pdf
- 매끄러운 pdf
 - 창 안의 샘플에 가중치를 준다. (중앙에 가까운 샘플이 더 높은 가중치)
 - 어떻게 이러한 아이디어를 구현할까?
- 커널 함수

$$\kappa(\mathbf{x}) = \begin{cases} 1, & |x_i| \leq 0.5, 1 \leq i \leq d \\ 0, & \text{그 외} \end{cases} \quad (3.9)$$



(a) 계단 함수



(b) 가우시언 함수

그림 3.7 파젠 창이 사용하는 대표적인 커널 함수

비모수적 방법 | 파젠 창

- 커널 함수를 사용하여 수식을 다시 쓰면,

$$k_{\mathbf{x}} = \sum_{i=1}^N \kappa\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) \quad (3.10)$$

$$p(\mathbf{x}) = \frac{1}{h^d} \frac{1}{N} \sum_{i=1}^N \kappa\left(\frac{\mathbf{x} - \mathbf{x}_i}{h}\right) \quad (3.11)$$

- 커널 함수로 가우시언을 채택하면 매끄러운 pdf를 얻게 된다.
- 파젠 창의 특성
 - 차원의 저주에서 자유로운가?
 - 추정한 pdf가 실제에 가까우려면 N 과 h 는 어떻게 되어야 하나?



비모수적 방법 | k -최근접 이웃 추정

- k -최근접 이웃 추정

- \mathbf{x} 를 중심으로 창을 씌우고 k 개 샘플이 안에 들어올 때까지 확장하고 그 순간의 창의 크기를 h 라 한다. 즉 k 가 고정되고 h 가 가변이다.
- 파젠 창에서는 h 가 고정되고 k 가 가변이다.

$$p(\mathbf{x}) = \frac{1}{h_x^d} \frac{k}{N} \quad (3.12)$$

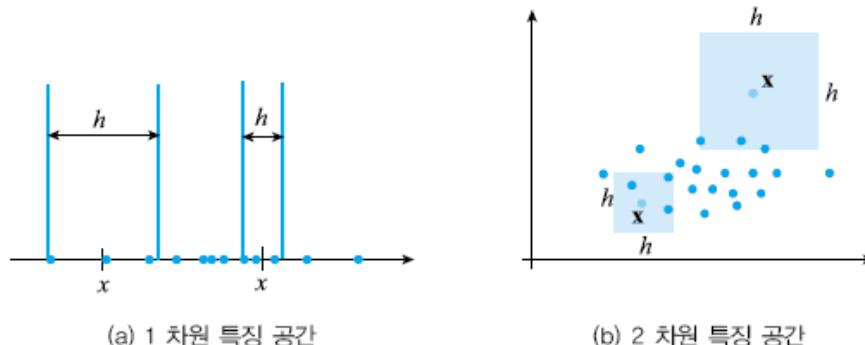


그림 3.8 k -최근접 이웃 방법으로 확률 밀도 함수의 추정 ($k=3$ 일 때)

비모수적 방법 | k -최근접 이웃 추정 예제

- 아래의 6개의 샘플이 있을 때 확률 밀도 함수를 추정하고자 한다.
 $-2.1, -1.3, -0.4, 1.9, 5.1, 6.2$
 - 1) $h=2$ 일때 파젠 창 방법으로 $p(-2), p(0), p(2), p(4), p(6)$ 의 확률 값을 구해라.
 - 2) $k=2$ 일때 k -최근접 이웃 추정 방법으로 $p(-2), p(0), p(2), p(4), p(6)$ 의 확률 값을 구해라.



비모수적 방법 | k -최근접 이웃 추정 예제

- 아래의 6개의 샘플이 있을 때 확률 밀도 함수를 추정하고자 한다.

$-2.1, -1.3, -0.4, 1.9, 5.1, 6.2$

- 1) $h=2$ 일때 파젠 창 방법으로 $p(-2), p(0), p(2), p(4), p(6)$ 의 확률 값을 구해라.

0.1667 0.0833 0.0833 0 0.1667

- 2) $k=2$ 일때 k -최근접 이웃 추정 방법으로 $p(-2), p(0), p(2), p(4), p(6)$ 의 확률 값을 구해라.

0.2381 0.1282 0.0694 0.0794 0.1852



비모수적 방법 | k -최근접 이웃 분류기

- k -NN 분류기

- 확률 분포 추정이 아니라 분류기인데, k -NN 추정과 동작이 흡사하여 여기에서 설명함
- \mathbf{x} 를 중심으로 창을 씌우고, k 개 샘플이 안에 들어올 때까지 확장. 이때의 창의 크기를 $h_{\mathbf{x}}$ 라 하면 창의 부피는 $h_{\mathbf{x}}^d$
- 창 안의 샘플 중에 ω_i 에 속하는 것의 개수를 k_i 라 하면,

$$p(\mathbf{x} | \omega_i) = \frac{k_i}{h_{\mathbf{x}}^d N_i} \quad (3.13)$$

$$p(\omega_i) = \frac{N_i}{N} \quad (3.14)$$

$$p(\mathbf{x}) = \frac{k}{h_{\mathbf{x}}^d N} \quad (3.15)$$



비모수적 방법 | k -최근접 이웃 분류기

- 베이스 정리를 적용하면

$$P(\omega_i \mid \mathbf{x}) = \frac{P(\omega_i)p(\mathbf{x} \mid \omega_i)}{p(\mathbf{x})} = \frac{k_i}{k} \quad (3.16)$$

- k -NN 분류기

$$\left. \begin{array}{l} k\text{-NN 분류기: } \mathbf{x} \text{를 } \omega_q \text{로 분류하라.} \\ \text{이 때 } q = \arg \max_i k_i \end{array} \right\} \quad (3.17)$$

알고리즘 [3.1]

k -NN 분류기

입력: 훈련 집합 $X = \{(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_N, t_N)\}$, 미지의 샘플 \mathbf{x}

출력: 부류 ω_q

알고리즘:

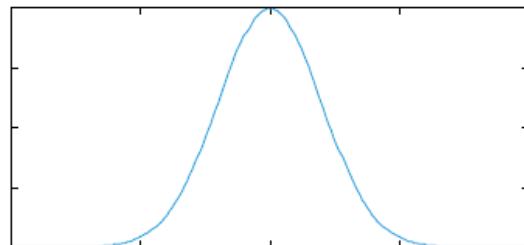
- 훈련 샘플 중에 \mathbf{x} 에 가장 가까운 k 개를 찾는다.
- k 개가 속한 부류를 조사하여 가장 빈도가 높은 부류를 ω_q 라 한다.



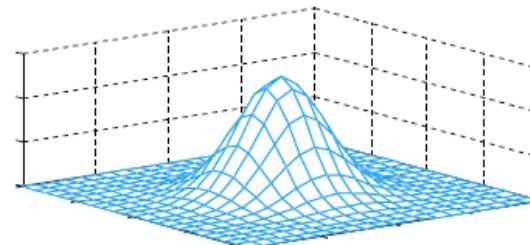
혼합 모델

- 두 개 이상의 서로 다른 확률 분포의 혼합으로 X 를 모델링함
 - 보통 요소 확률 분포로는 가우시언을 사용함

$$\left. \begin{aligned} N(\mu, \sigma^2) &= \frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \end{aligned} \right\} \quad (2.32)$$



(a) 1 차원 정규 분포

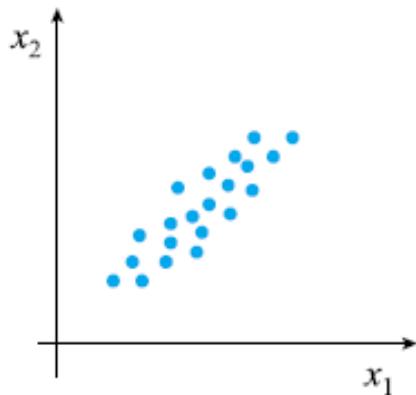


(b) 2 차원 정규 분포

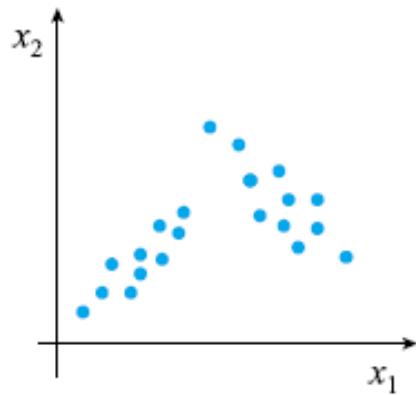
그림 2.9 정규 분포의 예

혼합 모델 | 가우시언 혼합

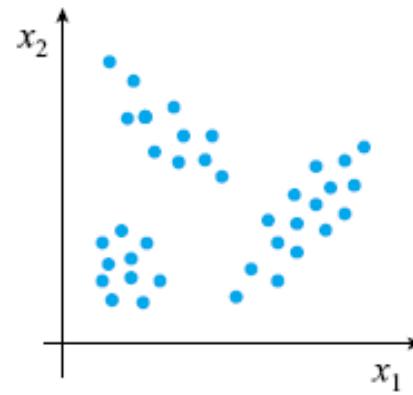
- 다양한 분포들



(a) 한 개의 모드



(b) 두 개의 모드



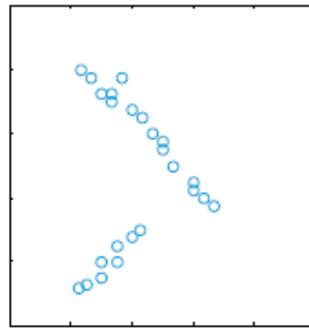
(c) 세 개의 모드

그림 3.10 다양한 분포들

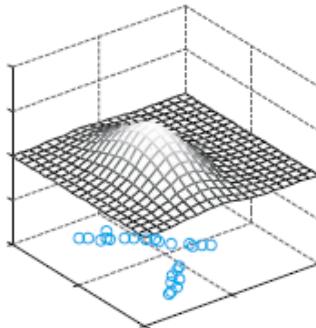
- 어떻게 다중 모드 분포를 정확히 모델링 할 수 있을까?

혼합 모델 | 가우시언 혼합

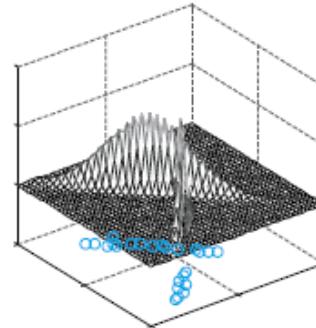
- 가우시언 혼합 Gaussian mixture



(a) 샘플 분포



(b) 한 개의 가우시언 사용



(c) 가우시언 혼합
(두 개의 가우시언) 사용

그림 3.11 가우시언 모델링

가우시언의 개수 K

k 번째 가우시언의 매개 변수 (μ_k, Σ_k) , $k = 1, \dots, K$

k 번째 가우시언의 가중치 π_k , $k = 1, \dots, K$

– 추정해야 할 매개 변수

$$p(\mathbf{x}) = \frac{1}{3} N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + \frac{2}{3} N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$$

Thank you