기계학습 개론

- Probabilistic Classification (1/2)

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

베이시언 결정 이론 I

- 보편적인 인식 법칙
  - ‘가장 그럴듯한’ 부류로 분류
    - TV에서 유튜버 이사배인지 가수 선미인지 헛갈리는데 선미 같다.
    - 도로 표지판이 전라북도 전주인지 경상남도 진주인지 확실치 않은데 전주인것 같다.
- 기계 (컴퓨터)의 인식
  - 수학 틀에 넣어야 프로그래밍이 가능해짐

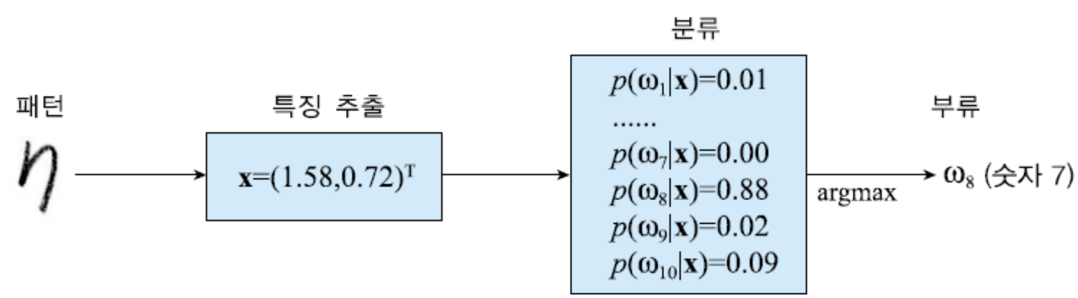


그림 2.1 ‘가장 그럴듯한’ 이라는 철학을 수학 틀에 넣음

  - $P(\omega_i|\mathbf{x})$: **x**가 주어졌을때 그것이 부류 ω_i에서 발생했을 확률 (사후 확률)

베이시언 결정 이론 I

- 사후 확률 $P(\omega_i|\mathbf{x})$의 추정
  - 어려운가? (그림 1.6을 가지고 생각해 보자.)
  - 왜?
  - 어떻게 추정하나?
  - 확률적 분류 방법의 핵심 주제

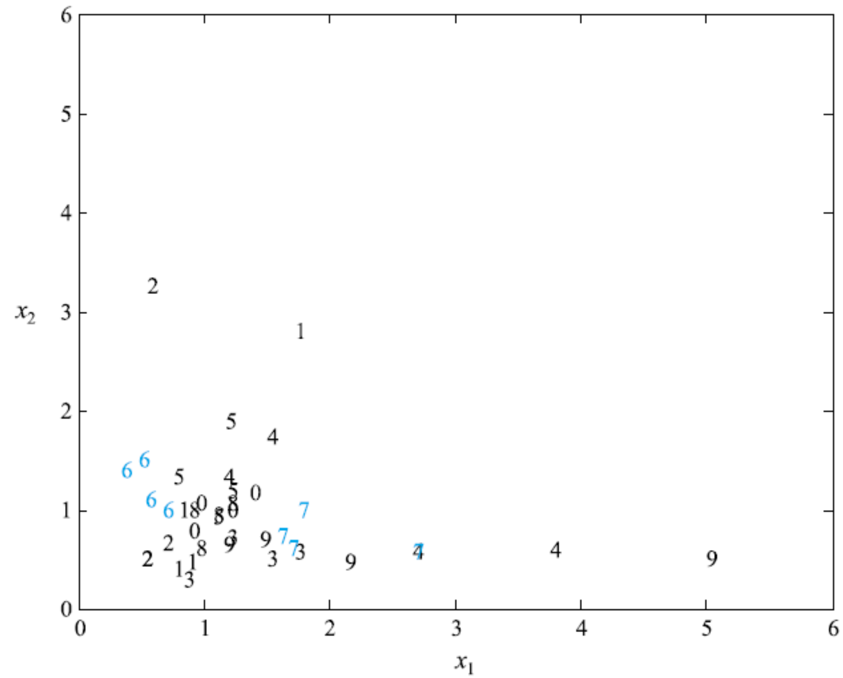


그림 1.6 특징 공간에서 샘플의 분포

확률과 통계 | 확률 기초

- 주사위
  - 주사위 던졌을 때 3이 나올 확률 $P(X=3)=1/6$
  - X를 랜덤 변수라 부름
  - 이 경우 X는 이산 값을 가짐
- 사람 키
  - 연속 값
  - 확률 밀도 함수 $p(X)$

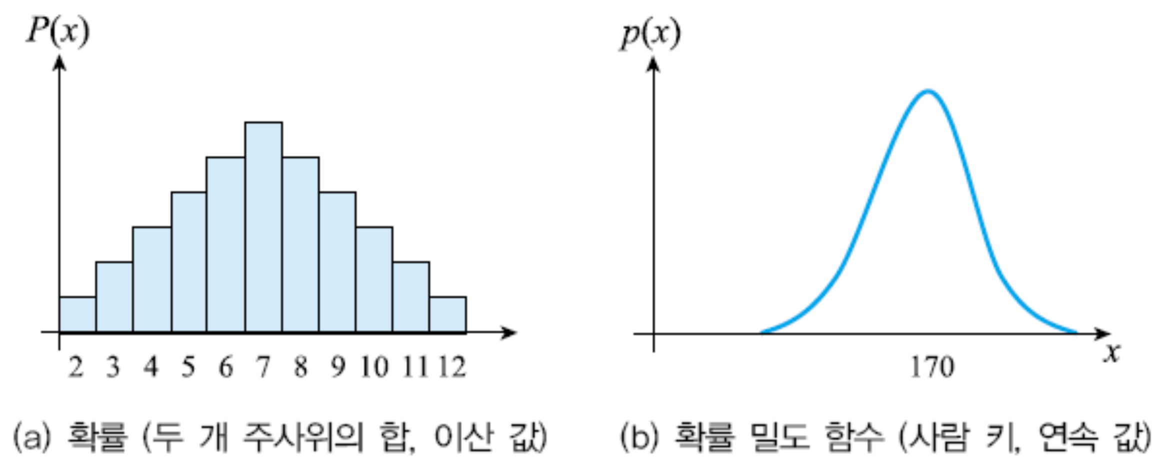


그림 2.2 확률 분포 함수

- 기계학습에서 특징 각각이 랜덤 변수에 해당

확률과 통계 | 확률 기초

- 확률 실험 (사전 확률, 우도, 사후 확률을 설명할 목적의 시나리오)

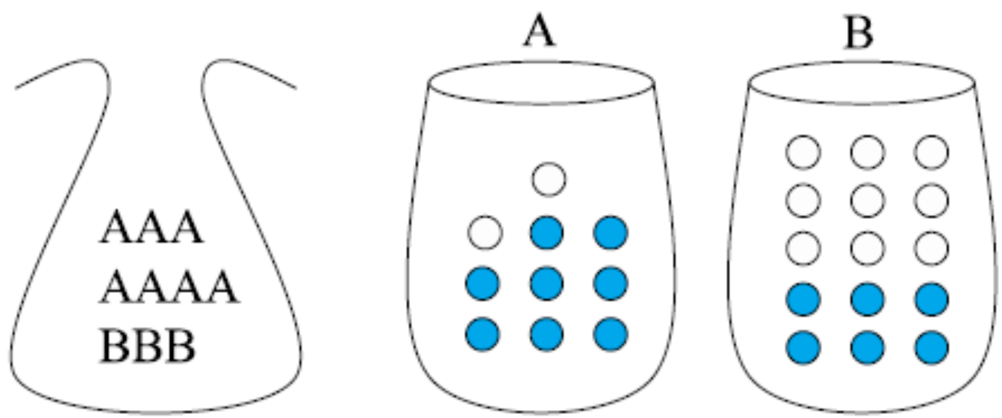


그림 2.3 확률 실험 장치

- 주머니에서 카드를 뽑아 상자를 선택하고 선택된 상자에서 공을 뽑아 관찰
- 랜덤 변수 $X \in \{A, B\}$, $Y = \{$파랑, 하양$\}$

확률과 통계 | 확률 기초

- 베이스 정리의 유도

$$P(X,Y) = P(Y,X)$$

$$P(X)P(Y \mid X) = P(Y)P(X \mid Y)$$

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)} = \frac{\text{우도} * \text{사전확률}}{P(Y)} \qquad (2.1)$$

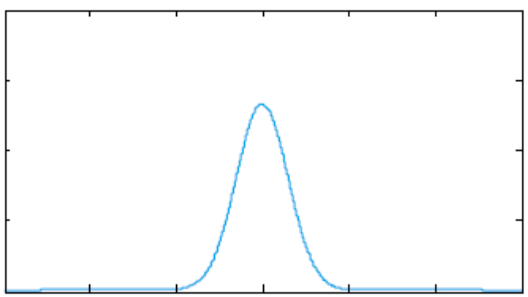
- 베이스 정리를 이용한 사후 확률 계산

$$P(\mathrm{A} \mid \text{하양}) = \frac{P(\text{하양} \mid \mathrm{A})P(\mathrm{A})}{P(\text{하양})} = \frac{(2/10)(7/10)}{(8/25)} = 0.4375$$

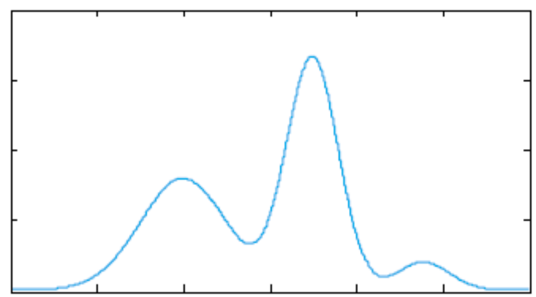
$$P(\mathrm{B} \mid \text{하양}) = \frac{P(\text{하양} \mid \mathrm{B})P(\mathrm{B})}{P(\text{하양})} = \frac{(9/15)(3/10)}{(8/25)} = 0.5625$$

확률과 통계 | 확률 분포의 표현과 추정

- 이산인 경우
  - 차원의 저주
  - 변수의 수가 d이고 각 변수가 q개의 구간을 가진다면 q^d에 비례하는 메모리 필요
- 연속인 경우
  - 일정한 형태를 가는 상황
  - 그렇지 않은 상황



(a) 정규 분포



(b) 세 개의 모드를 가진 분포

그림 2.4 다양한 확률 분포

MYONGJI UNIVERSITY

확률과 통계 | 베이시언 분류기

- 분류기 학습 (훈련)에 사용하는 정보는 ‘훈련 집합’
  - 훈련 집합 $X=\{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), \cdots, (\mathbf{x}_N, t_N)\}$
    - $\mathbf{x}_i=(x_1, x_2, \cdots, x_d)$는 특징 벡터
    - $t_i \in \{\omega_1, \omega_2, \cdots, \omega_M\}$ 부류 표지 (이진 분류기인 경우 $M=2$)

- 예) 필기 숫자

$\mathbf{x}_1=(13/11, 12/12)^T$, $t_1=\omega_1$ (숫자 0)

$\mathbf{x}_2=(12/7, 14/5)^T$, $t_2=\omega_2$ (숫자 1)

$\mathbf{x}_3=(6/11, 13/4)^T$, $t_3=\omega_3$ (숫자 2)

$\mathbf{x}_4=(13/11, 10/14)^T$, $t_4=\omega_4$ (숫자 3)

……

$\mathbf{x}_{40}=(15/13, 11/17)^T$, $t_{40}=\omega_{10}$ (숫자 9)

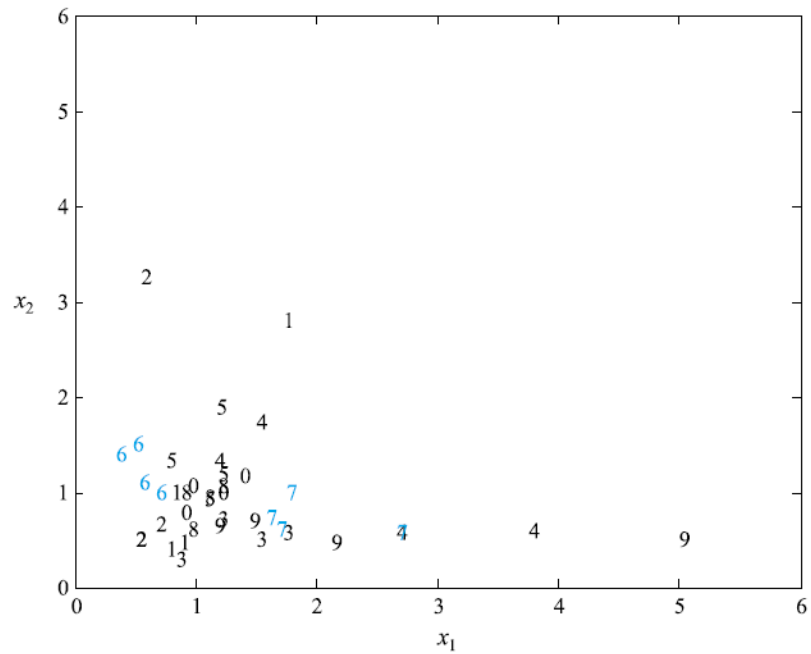


그림 1.6 특징 공간에서 샘플의 분포

베이시언 분류기 | 최소 오류 베이시언 분류기

- 주어진 특징 벡터 **x**에 대해 '가장 그럴듯한' 부류로 분류

$$\left.\begin{array}{l} P(\omega_1 \mid \mathbf{x}) > P(\omega_2 \mid \mathbf{x})\text{이면, }\mathbf{x}\text{를 }\omega_1\text{로 분류하고} \\ P(\omega_1 \mid \mathbf{x}) < P(\omega_2 \mid \mathbf{x})\text{이면, }\mathbf{x}\text{를 }\omega_2\text{로 분류하라.} \end{array}\right\} \tag{2.16}$$

- (2.16)에서 사후 확률은 직접 구할 수 없음. 왜?
- 베이스 정리를 이용하여 사후 확률 계산을 사전 확률과 우도로 대치

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{p(\mathbf{x})} = \frac{\text{우도}*\text{사전확률}}{p(\mathbf{x})} \tag{2.17}$$

- 분모는 무시해도 됨. 왜?
- 우도와 사전 확률은 어떻게 계산?

MYONGJI UNIVERSITY

베이시언 분류기 | 최소 오류 베이시언 분류기

- 사전 확률 계산
  - $P(\omega_1)=n_1/N$, $P(\omega_2)=n_2/N$
  - 정확한 값이 아니라 추정 (N이 커짐에 따라 실제 값에 가까워짐)

- 우도 계산
  - 훈련 집합에서 ω_i에 속하는 샘들들을 가지고 $P(\mathbf{x}|\omega_i)$ 추정
  - 부류 조건부 확률 이라고도 class-conditional probabilty 함

베이시언 분류기 | 최소 오류 베이시언 분류기

- 최소 오류 베이시언 분류기
  - 결정 규칙

$$\left.\begin{array}{l} p(\mathbf{x}|\omega_1)P(\omega_1) > p(\mathbf{x}|\omega_2)P(\omega_2)\text{이면, } \mathbf{x}\text{를 } \omega_1\text{로 분류하고} \\ p(\mathbf{x}|\omega_1)P(\omega_1) < p(\mathbf{x}|\omega_2)P(\omega_2)\text{이면, } \mathbf{x}\text{를 } \omega_2\text{로 분류하라.} \end{array}\right\} \quad (2.18)$$

  - 특수한 경우로 (2.18)의 의미 해석하면,
    - 사전 확률이 0.5인 경우 우도만으로 분류
    - $P(\omega_1) >> P(\omega_2)$인 경우 사전 확률이 의사 결정 주도

베이시언 분류기 | 최소 오류 베이시언 분류기

- 최소 오류 베이시언 분류기
  - 오류 확률

$$E = \frac{1}{2}\left(\int_{-\infty}^{t} p(x \mid \omega_2)dx + \int_{t}^{\infty} p(x \mid \omega_1)dx\right)$$

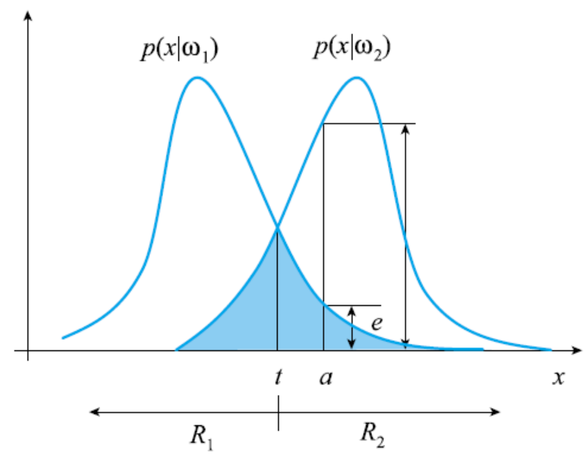


그림 2.5 베이시언 분류기의 오류 확률 (사전 확률은 같다고 가정)

  - 최적성

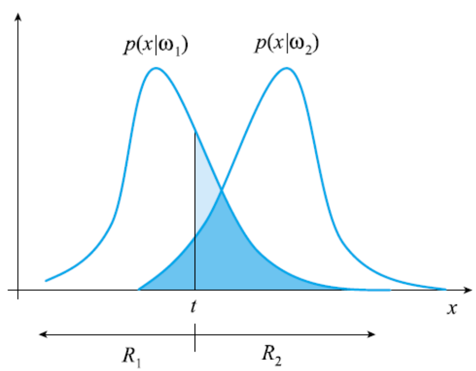


그림 2.6 베이시언 분류기를 사용하지 않을 때 증가하는 오류

베이시언 분류기 | 최소 위험 베이시언 분류기

- 성능 기준으로 오류가 적절하지 못한 상황
  - 정상인과 암 환자 분류
  - 과일을 상품과 하품으로 분류

- 손실 행렬

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

베이시안 분류기 | 최소 위험 베이시안 분류기

- 최소 위험 베이시안 분류기

$$\left.\begin{array}{l} \mathbf{x}\text{를 } q_2 > q_1\text{이면 } \omega_1\text{로 분류하고, } q_1 > q_2\text{이면 } \omega_2\text{로 분류하라.} \\ \text{이때 } q_1 = c_{11}p(\mathbf{x}|\omega_1)P(\omega_1) + c_{21}p(\mathbf{x}|\omega_2)P(\omega_2) \\ \quad\quad q_2 = c_{12}p(\mathbf{x}|\omega_1)P(\omega_1) + c_{22}p(\mathbf{x}|\omega_2)P(\omega_2) \end{array}\right\} \quad (2.23)$$

- 우도비로 다시 쓰면
  - 우도비 결정 규칙

$$\left.\begin{array}{l} \mathbf{x}\text{를 } \dfrac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > T \text{ 이면 } \omega_1\text{로 분류하고, } \dfrac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} < T \text{ 이면 } \omega_2\text{로 분류하라.} \\ \text{이때 } T = \dfrac{(c_{21}-c_{22})P(\omega_2)}{(c_{12}-c_{11})P(\omega_1)} \end{array}\right\} \quad (2.24)$$

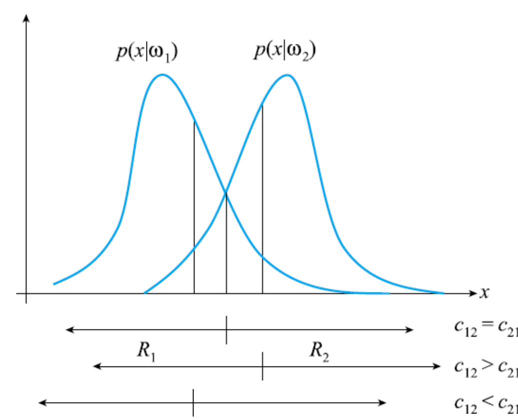


그림 2.7 최소 위험 베이시안 분류기 (사전 확률은 같다고 가정)

베이시언 분류기 | *M* 부류로 확장

- *M* 부류 최소 오류 베이시언 분류기
  - 사후 확률로 쓰면

$$\mathbf{x}\text{를 } k = \arg\max_i P(\omega_i \mid \mathbf{x}) \text{ 일 때 } \omega_k\text{로 분류하라.} \qquad (2.26)$$

  - 사전 확률과 우도로 쓰면

$$\mathbf{x}\text{를 } k = \arg\max_i p(\mathbf{x} \mid \omega_i)P(\omega_i) \text{ 일 때 } \omega_k\text{로 분류하라.} \qquad (2.27)$$

- *M* 부류 최소 위험 베이시언 분류기

$$\left.\begin{array}{l} \mathbf{x}\text{를 } k = \arg\min_i q_i \text{ 일 때 } \omega_k\text{로 분류하라.} \\ \text{이때 } q_i = \sum_{j=1}^{M} c_{ji}\, p(\mathbf{x} \mid \omega_j)P(\omega_j) \end{array}\right\} \qquad (2.29)$$

MYONGJI UNIVERSITY

베이시언 분류기 | 분별 함수

- 지금까지 분류기를 분별 함수로 다시 작성하면

$$
\left.
\begin{array}{l}
\mathbf{x}\text{를 } k = \underset{i}{\arg\max}\ g_i(\mathbf{x}) \text{ 일때 } \omega_k \text{로 분류하라.} \\
\text{이때 } g_i(\mathbf{x}) = \begin{cases} p(\mathbf{x} \mid \omega_i) P(\omega_i) & \text{(최소 오류 베이시언 분류기)} \\ \dfrac{1}{\sum_{j=1}^{M} c_{ji}\, p(\mathbf{x} \mid \omega_j) P(\omega_j)} & \text{(최소 위험 베이시언 분류기)} \end{cases}
\end{array}
\right\} \quad (2.30)
$$

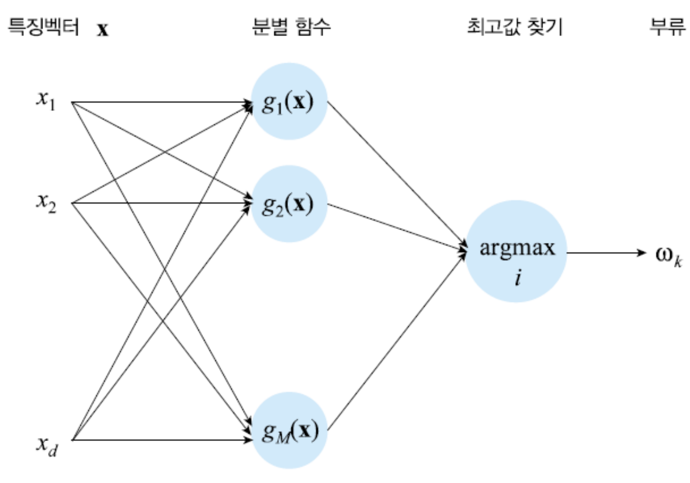


그림 2.8 분류기의 일반적인 틀

알고리즘 [2.1] *M* 부류 베이시언 분류기

입력: 특징 벡터 $\mathbf{x}$

출력: 부류

알고리즘:

1. **for** ($i = 1$ **to** M)
2. **if** (최소 오류) $g_i(\mathbf{x}) = p(\mathbf{x} \mid \omega_i) P(\omega_i)$;
3. **else if** (최소 위험) $g_i(\mathbf{x}) = 1 \Big/ \sum_{j=1}^{M} c_{ji}\, p(\mathbf{x} \mid \omega_j) P(\omega_j)$;
4. $g_i(\mathbf{x})$, $1 \le i \le M$ 중 가장 큰 것을 찾아 그것의 첨자를 k라 하자.
5. **return** ω_k;

분별 함수 I

- 분별 함수 표현의 장점
  - 여러 분류기를 하나의 틀로 표현
  - $f(.)$가 단조 증가라면 $p(\mathbf{x}|\omega_i)\ P(\omega_i)$대신 $g_i(x)=f(p(\mathbf{x}|\omega_i)\ P(\omega_i))$ 사용하여도 같은 결과
    - $f(.)$로 log 함수를 주로 사용
    - log는 곱셈을 덧셈으로 바꾸어 주므로 수식 전개에 유리하고 log 취하면 값의 규모가 커져 수치 오류에 둔감한 이점

정규분포에서 베이시언 분류기 I

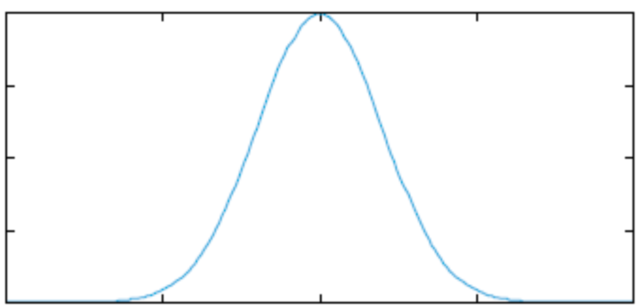
- 정규 분포 (가우시언 분포)
  - 현실 세계에 맞는 경우 있음
  - 평균과 분산이라는 두 종류의 매개 변수만으로 표현 가능
  - 수학적인 매력

- 우도가 정규 분포를 따른다는 가정 하에 베이시언 분류기의 특성을 해석해 보자.

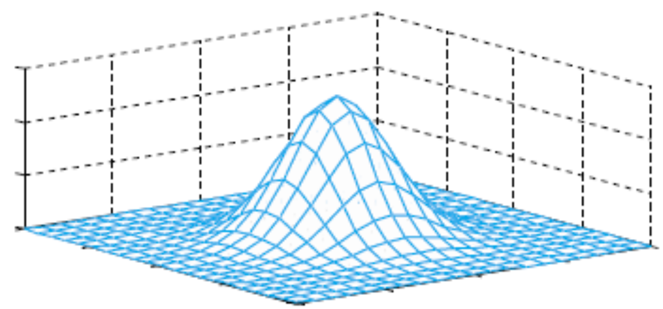
정규분포에서 베이시언 분류기 I
정규분포와 분별 함수

- 정규 분포 (가우시언 분포)

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



(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,

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

- 로그를 취하여 분별 함수를 만들어 보면,

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

  - $g_i(\mathbf{x})$는 변수 **x**에 대한 2차 식

MYONGJI UNIVERSITY

정규분포에서 베이시언 분류기 I
정규분포와 분별 함수

- 예제 2.4
  - d=2이고 아래와 같다고 가정

부류 ω_i의 정규 분포가 $\boldsymbol{\mu}_i = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$와 $\boldsymbol{\Sigma}_i = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

  - 분별 함수를 유도해 보면,

$$g_i(\mathbf{x}) = -\frac{1}{2}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right)^{\mathrm{T}} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix}\right) - \frac{2}{2}\ln 2\pi - \frac{1}{2}\ln\left|\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\right| + \ln P(\omega_i)$$

$$= -\frac{1}{2}\begin{pmatrix} x_1 - 3 & x_2 - 1 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 - 3 \\ x_2 - 1 \end{pmatrix} - \ln 2\pi - \frac{1}{2}\ln 4 + \ln P(\omega_i)$$

$$= -\frac{1}{4}\underbrace{(x_1^{\,2} + x_2^{\,2})}_{2\text{차 식}} + \frac{1}{2}\underbrace{(3x_1 + x_2)}_{1\text{차 식}} - \frac{1}{2}\underbrace{(5 + 2\ln 2\pi + \ln 4 - 2\ln P(\omega_i))}_{\text{상수}}$$

MYONGJI UNIVERSITY

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

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

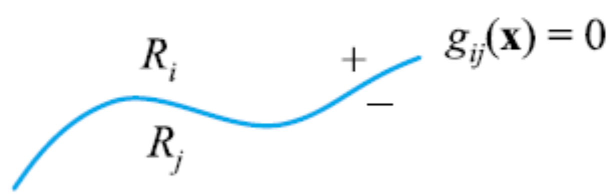


그림 2.10 두 부류 ω_i와 ω_j의 결정 경계

MYONGJI UNIVERSITY

정규분포에서 베이시언 분류기 I
선형 분별

- 모든 부류의 공분산 행렬이 같은 상황, $\mathbf{\Sigma}_i = \mathbf{\Sigma}$로 표기
  - 분별 함수를 다시 쓰면

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

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

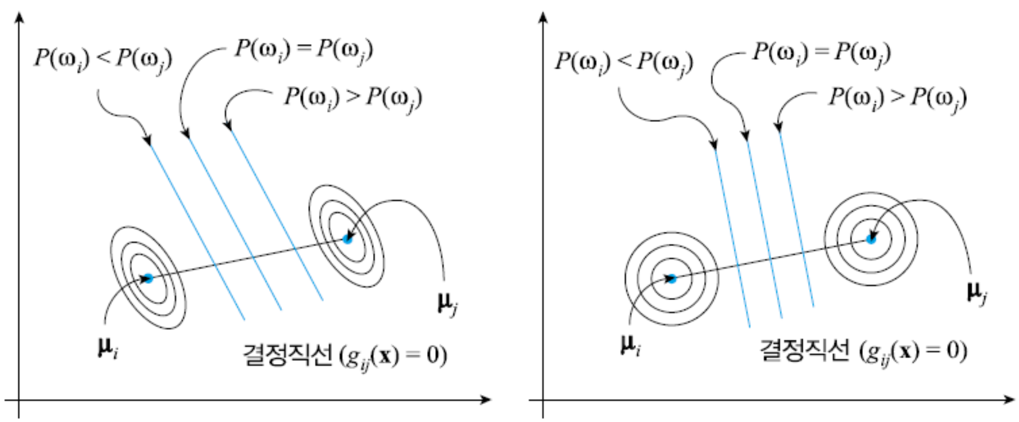
$$\left.\begin{aligned} g_i(\mathbf{x}) &= \left(\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i\right)^{\mathrm{T}}\mathbf{x} + \left(\ln P(\omega_i) - \frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_i\right) \\ &= \mathbf{w}_i^{\mathrm{T}}\mathbf{x} + b_i \end{aligned}\right\}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

MYONGJI UNIVERSITY

정규분포에서 베이시언 분류기 | 선형 분별

- 결정 경계

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



(a) 공분산이 같은 경우 (b) 공분산이 같고 $\sigma^2\mathbf{I}$인 경우

그림 2.11 선형 분류기

정규분포에서 베이시언 분류기 | 선형 분별

- 예제 2.5

ω_1: $(1,2)^T$, $(3,1)^T$, $(5,2)^T$, $(3,3)^T$

ω_2: $(6,6)^T$, $(8,5)^T$, $(10,6)^T$, $(8,7)^T$

$$\boldsymbol{\mu}_1 = \frac{1}{4}\left(\begin{pmatrix}1\\2\end{pmatrix}+\begin{pmatrix}3\\1\end{pmatrix}+\begin{pmatrix}5\\2\end{pmatrix}+\begin{pmatrix}3\\3\end{pmatrix}\right)=\begin{pmatrix}3\\2\end{pmatrix},\quad \Sigma_1 = \begin{pmatrix}2 & 0\\ 0 & 1/2\end{pmatrix}$$

$$\boldsymbol{\mu}_2 = \frac{1}{4}\left(\begin{pmatrix}6\\6\end{pmatrix}+\begin{pmatrix}8\\5\end{pmatrix}+\begin{pmatrix}10\\6\end{pmatrix}+\begin{pmatrix}8\\7\end{pmatrix}\right)=\begin{pmatrix}8\\6\end{pmatrix},\quad \Sigma_2 = \begin{pmatrix}2 & 0\\ 0 & 1/2\end{pmatrix}$$

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

$$= (-5/2 \quad -8)\begin{pmatrix}x_1\\x_2\end{pmatrix} + (\ln P(\omega_1) - \ln P(\omega_2) + 45.75)$$

$$= -\frac{5}{2}x_1 - 8x_2 + (\ln P(\omega_1) - \ln P(\omega_2) + 45.75)$$

$P(\omega_1) = 0.5, P(\omega_2) = 0.5$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 : $5x_1 + 16x_2 - 88.727 = 0$

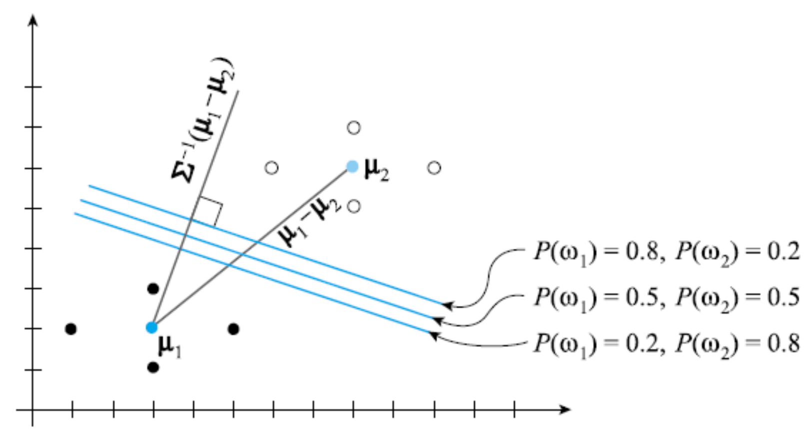


그림 2.12 선형 분별 분석 (LDA) 사례

정규분포에서 베이시언 분류기 I
2차 분별

- 임의의 공분산 행렬

$$g_i(\mathbf{x}) = \underbrace{-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x}}_{\text{2차 항}} + \underbrace{\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{x}}_{\text{1차 항}} + \underbrace{\left(-\frac{1}{2}\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)\right)}_{\text{상수 항}} \qquad (2.40)$$

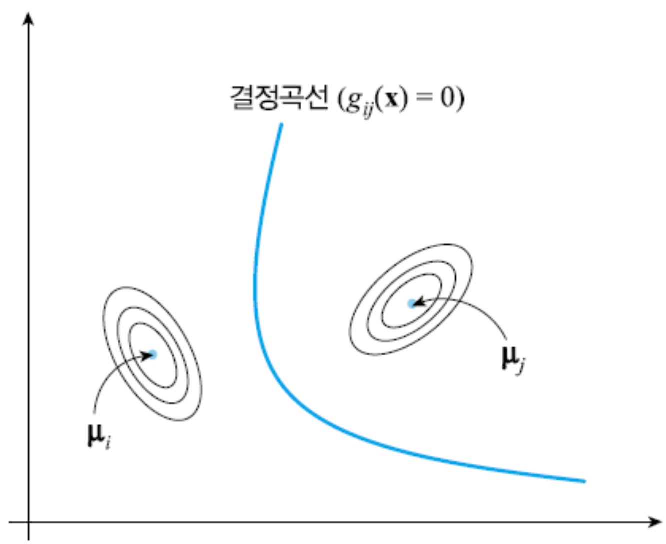


그림 2.13 2차 분별 분석으로 만든 2차 분류기

정규분포에서 베이시언 분류기 | 2차 분별

- 예제 2.6

ω_1: $(1,2)^T$, $(3,1)^T$, $(5,2)^T$, $(3,3)^T$

ω_2: $(7,6)^T$, $(8,4)^T$, $(9,6)^T$, $(8,8)^T$

$\boldsymbol{\mu}_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$, $\Sigma_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}$

$\boldsymbol{\mu}_2 = \begin{pmatrix} 8 \\ 6 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$

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

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

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

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

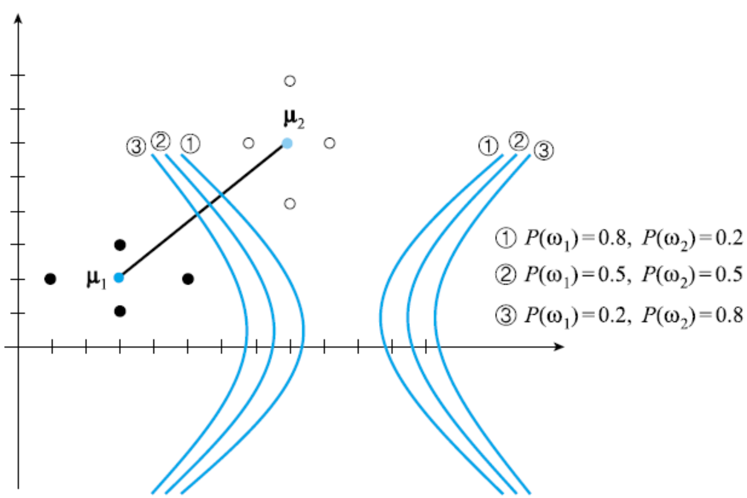


그림 2.14 2차 분별 분석 (QDA) 사례

MYONGJI UNIVERSITY

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

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

  - 최소 거리 분류기

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

- 거리 척도

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

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

MYONGJI UNIVERSITY

정규분포에서 베이시언 분류기 | 최소 거리 분류기

- 예제 2.7

ω_1: $(1,2)^T$, $(3,1)^T$, $(5,2)^T$, $(3,3)^T$

ω_2: $(6,6)^T$, $(8,5)^T$, $(10,6)^T$, $(8,7)^T$

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

$$\boldsymbol{\mu}_1\text{까지의 마할라노비스 거리} = \left((8-3 \quad 2-2) \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 8-3 \\ 2-2 \end{pmatrix} \right)^{1/2} = 3.536$$

$$\boldsymbol{\mu}_2\text{까지의 마할라노비스 거리} = \left((8-8 \quad 2-6) \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 8-8 \\ 2-6 \end{pmatrix} \right)^{1/2} = 5.657$$

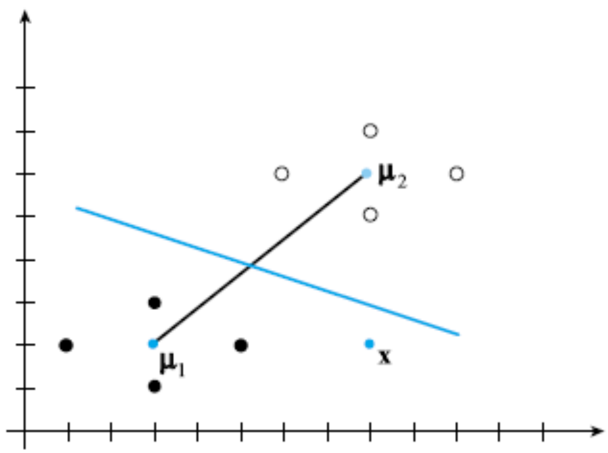


그림 2.15 두 통계 분포까지의 마할라노비스 거리

베이시언 분류의 특성 I

- 베이시언 분류의 특성

1. 비현실성을 내포하고 있다. 이미 언급한 바와 같이 일반적인 확률 분포를 사용하려 하면 차원의 저주가 발생한다. 정규 분포를 가정하면 실제 확률 분포와 차이가 발생한다.
2. 1의 비현실성을 무시하고 실제 확률 분포를 안다고 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

- 나이브naïve 베이시언 분류기
  - 특징들이 서로 독립이라는 가정

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

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

Thank you