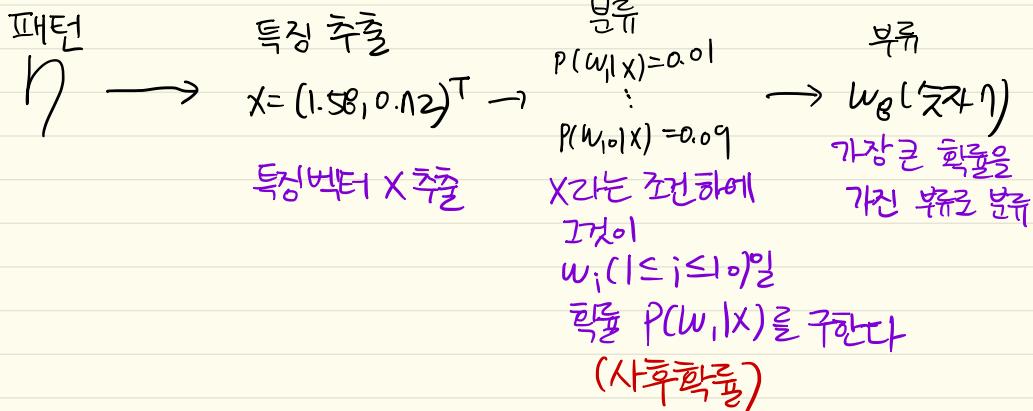


2. 베이시언 결정 이론

보편적인 인식법칙 : '가장 그럴듯한' 쪽으로 인식

↓ 컴퓨터는 이를 수학들어 넣어야 함



2.1 학률과 통계

'그럴듯한' 을 표현하는 수학적인 도구는 학률과 통계

2.1.1 학률 기초

• 주사위

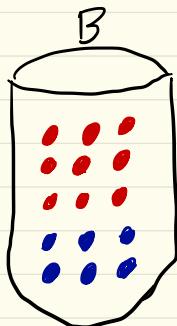
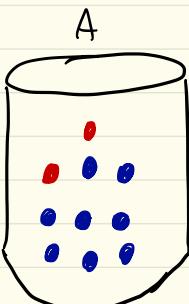
던졌을 때 3이 나올 학률 $P(x=3) = \frac{1}{6}$

랜덤변수 학률
이산값을 가짐

• 사람기

연속값 \rightarrow 학률밀도함수

패턴인식에서는 개별 특징 x_i 가 랜덤변수이다.



$$X \in \{A, B\} \quad Y \in \{\text{빨강}, \text{파랑}\}$$

1. 상자 A가 선택될 확률?

$$P(X=A) = P(A) = \frac{7}{10}$$

conditional probability

2. 상자 A에서 빨간 공이 뽑힐 확률은? 조건부 확률

$$P(Y=\text{빨강} | X=A) = P(\text{빨강}|A) = \frac{2}{10}$$

3. 상자는 A이고, 공은 빨강이 뽑힐 확률은? joint probability
결합확률

$$P(A, \text{빨강}) = P(\text{빨강}|A) \cdot P(A) = \frac{2}{10} \cdot \frac{7}{10} = \frac{7}{50}$$

4. 빨간 공이 나올 확률은? marginal Probability
주변확률

$$P(\text{빨강}) = P(\text{빨강}|A) \cdot P(A) + P(\text{빨강}|B) \cdot P(B)$$

$$= \frac{2}{10} \cdot \frac{7}{10} + \frac{9}{15} \cdot \frac{3}{10} = \frac{7}{50} + \frac{9}{50} = \frac{16}{50} = \frac{8}{25}$$

- $P(X, Y) = P(X) \cdot P(Y)$ 이면 X와 Y는 독립
- $P(A)$ 와 $P(B)$ 를 사전 확률 (Prior Probability)라고 한다.

* 빨간 공이 뽑혔는데 어느 상자에서 나왔는지 맞추어라.
 → 상자 A와 B에서 나왔을 가능성 각각을 구하고 큰 가능성을 보인 상자를 답으로 취한다.

생각 1.

- 상자 A의 빨간 공 확률과 상자 B의 빨간 공 확률을 비교하여 큰 쪽을 취한다.
- $P(\text{빨강} | B) = 9/15 > P(\text{빨강} | A) = 2/10$ 이므로
 '상자 B에서 나왔다'고 말함
- 조건부 확률 $P(Y|X)$ 를 사용하는 심이다. 타당한가?
- 이 조건부 확률을 우도 (likelihood)라고 부름

생각 2.

상자 A와 상자 B의 선택 가능성을 비교하여 큰 쪽을 취한다.
 $P(A) = 7/10 > P(B) = 3/10$ 이므로 '상자 A에서 나왔다'고 말함

올바른 생각

생각 1과 생각 2의 한계

극단적으로 $P(A) = 0.999$ 라면 생각 1이 틀린 것이 확실하다.

극단적으로 $P(\text{빨강}|A) = 0.999$ 라면 생각 2가 틀린 것이 확실하다.

우도와 사전 확률을 모두 고려함이 타당해 보임

→ 조건부 확률 $P(A|\text{빨강})$ 과 $P(B|\text{빨강})$ 을 비교하여 큰 쪽을 취한다. 즉, $P(X|Y)$ 를 사용하겠다는 생각이 타당하다.

$P(X|Y)$ 를 **사후 확률 (Posterior Probability)** 이라 함.

$P(X|Y)$ 는 어떻게 구할까?

$$P(X|Y) = P(Y|X)$$

$$P(Y|X) \cdot P(X) = P(X|Y) \cdot P(Y) \quad \text{베이스 정리}$$

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} = \frac{\text{우도} \times \text{사전확률}}{P(Y)}$$

$$P(A|\text{빨강}) = \frac{P(\text{빨강}|A) \cdot P(A)}{P(\text{빨강})} = \frac{(2/10)(7/10)}{8/25} = 0.4375$$

$$P(B|\text{빨강}) = \frac{P(\text{빨강}|B) \cdot P(B)}{P(\text{빨강})} = \frac{(9/15)(3/10)}{8/25} = 0.5625$$

2.1.2 평균과 분산

① 이산 확률분포가 주어진 경우

이산 확률 분포 $\begin{aligned} \text{평균} &= \sum_{x_i} x_i \cdot P(x_i) \\ \text{분산} (\sigma^2) &= \sum_{x_i} (x_i - \text{평균})^2 \cdot P(x_i) \end{aligned}$

가장 평균 로부터 얼마나 떨어져 있는지

연속 확률 분포 $\text{평균} = \int_{-\infty}^{\infty} x \cdot p(x) dx$

분산 = $\int_{-\infty}^{\infty} (x - \text{평균})^2 \cdot p(x) dx$

② 샘플 집합만 주어진 경우

$$\text{평균} = \frac{1}{N} \cdot \sum_{i=1}^N x_i$$

$$\text{분산} (\sigma^2) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \text{평균})^2$$

바이어스 때문에

현실에서는 여러 랜덤 변수가 랜덤 벡터를 구성하는 경우가 많다.

랜덤 벡터를 d차원 벡터 $x = (x_1, \dots, x_d)^T$ 로 표기하자.

분산에 대해 생각해보면, 랜덤 벡터 i번째 요소 x_i 의 분산 s_i^2 로 필요하나, x_i 와 x_j 사이의 공분산 (covariance) s_{ij} 도 중요한 통계적 특성이 된다. 확률변수들의 관계를 보여주는 값

$$\Sigma = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1d} \\ s_{21} & s_{22} & & s_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ s_{d1} & s_{d2} & & s_{dd} \end{pmatrix}$$

이산 확률 분포 $\Sigma = \sum_x (x - \mu)(x - \mu)^T \cdot p(x)$

연속 " " $\Sigma = \int_{\mathbb{R}^d} (x - \mu)(x - \mu)^T \cdot p(x) dx$

샘플 잡합 $\Sigma = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T$

두 랜덤 변수 x_i 와 x_j 가 크면 x_j 도 커지고,

x_i 가 작아지면 x_j 도 작아지는 경향을 가졌다면,

둘의 공분산 s_{ij} 는 양수를 갖는다. 그런 경향이

강할수록 같은 더 커진다. 공분산이 0에 가까우면 상관관계가 적은 것. 0이면 서로 독립!

■ 예제 2.3

- ▣ 8개 샘플이 주어진 상황에서 평균 벡터와 공분산 행렬 구함

학생	$\mathbf{x} = (x_1, x_2, x_3)^T$ (x_1 : 나이, x_2 : 몸무게, x_3 : 학점)
1	$\mathbf{x}_1 = (170, 60, 4.1)^T$
2	$\mathbf{x}_2 = (165, 55, 3.0)^T$
3	$\mathbf{x}_3 = (174, 75, 2.8)^T$
4	$\mathbf{x}_4 = (169, 67, 2.9)^T$
5	$\mathbf{x}_5 = (155, 49, 3.1)^T$
6	$\mathbf{x}_6 = (172, 63, 3.6)^T$
7	$\mathbf{x}_7 = (166, 58, 3.7)^T$
8	$\mathbf{x}_8 = (168, 61, 4.0)^T$

$$\mu = (167.375, 61.0, 3.4)^T$$

$$\Sigma = \begin{pmatrix} 33.696 & 39.429 & 0.371 \\ 39.429 & 60.857 & -0.943 \\ 0.371 & -0.943 & 0.263 \end{pmatrix}$$

2.1.3 확률 분포의 표현과 추정

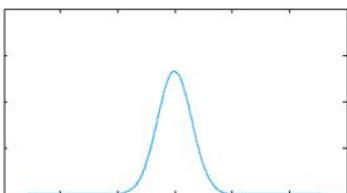
- 이산인 경우

변수의 수가 d 개로 늘어나고 각 변수의 구간이 q 개라면 d 차원 배열을 사용해야 하고, q^d 에 비례하는 메모리가 필요 (차원의 저주)

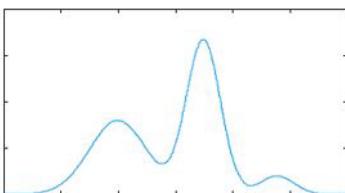
- 연속인 경우

정규 분포 (normal distribution)로 근사 가능

함들 → 3장에서



(a) 정규 분포



(b) 세 개의 모드를 가진 분포

그림 2.4 다양한 확률 분포

2.2 베이시언 분류기

이진 분류기와 M진 분류기로 나눠서 살펴보겠습니다

$$\text{훈련집합 } X = \{(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n)\}$$

x_i 는 i번째 샘플의 특징 벡터, t_i 는 x_i 가 속하는 class label
(부류표지)

이진 분류에서는 $t_i \in \{W_1, W_2\} \rightarrow$ 주로 -1, +1 or +1, 0
M 부류 $t_i \in \{W_1, W_2, \dots, W_M\}$

2.2.1 최소 오류 베이시언 분류기

분류 문제의 해결 \rightarrow 주어진 특징 벡터 x 에 대해 가장 그럴듯한
부류를 찾아내면 됨 \rightarrow 분류기가 이 문제를 해결해 줌

$(P(W_1|x) > P(W_2|x) \text{ 이면, } x \text{를 } W_1 \text{로 분류하고, })$
 $(P(W_1|x) < P(W_2|x) \text{ 이면, } x \text{를 } W_2 \text{로 분류하라 })$

but, x 가 만드는 공간, 특징 공간이 무수히 많은 점을 가지므로,
이들 모든 점에 대해 확률을 표현할 수 없다.

따라서 베이스 정리를 이용한다

예) 필기 숫자

$$x_1 = (13/11, 12/12)^T, t_1 = \omega_1 \text{ (숫자 0)}$$

$$x_2 = (12/7, 14/5)^T, t_2 = \omega_2 \text{ (숫자 1)}$$

$$x_3 = (6/11, 13/4)^T, t_3 = \omega_3 \text{ (숫자 2)}$$

$$x_4 = (13/11, 10/14)^T, t_4 = \omega_4 \text{ (숫자 3)}$$

.....

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

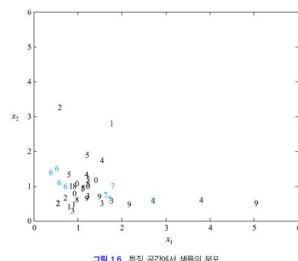


그림 1.6 특징 공간에서 샘플의 분포

$$P(W_i | x) = \frac{P(x|W_i) \cdot P(W_i)}{P(x)} = \frac{\text{우도} * \text{사전확률}}{P(x)}$$

사후 확률 $P(W_i | x)$ 와 사전확률 $P(W_i)$ 는 매개변수가 부류를 나타내는 W_i , 보면 우도 $P(x|W_i)$ 는 부류가 주어진 조건부 확률이고, 매개변수가 특징 벡터 x .

* 사전 확률 $P(W_i)$ 불이익 확률

훈련 집합에 있는 샘플 중에 W_i 에 속하는 샘플의 개수를 N_i 라 하면, $P(W_i) = N_i / N$, $P(W_2) = N_2 / N$ 로 추정
(N 이 충분히 크면 이 같은 실제 확률 값에 아주 가깝게 됨)

* 우도 $P(x|W_i)$ 불이익는데 연기인 확률

$P(x|W_i)$ 는 훈련 집합에서 W_i 에 속하는 샘플들을 이용하여 추정
우도는 부류가 주어진 조건하의 조건부 확률이므로 부류조건부 확률
(class conditional probability)라고 부름. → 이미 구해 놓았다
가정하자.

* $P(x)$

사후 확률 값 자체가 필요한 게 아니라 두 사후 확률 중 어느 것이 더 큰지 판별하면 되므로 생각 가능

결정 규칙 (decision rule) : 특징 벡터를 여러 개 부류 중의 하나로 분류하는 규칙

Minimum error Bayesian classifier(최소 오류 베이시언 분류기): 결정규칙을 사용하는 분류기

- 결정규칙

$P(x|w_1)P(w_1) > P(x|w_2)P(w_2)$ 이면, x 를 w_1 로 분류하고
 $P(x|w_1)P(w_1) < P(x|w_2)P(w_2)$ 이면, x 를 w_2 로 분류하라

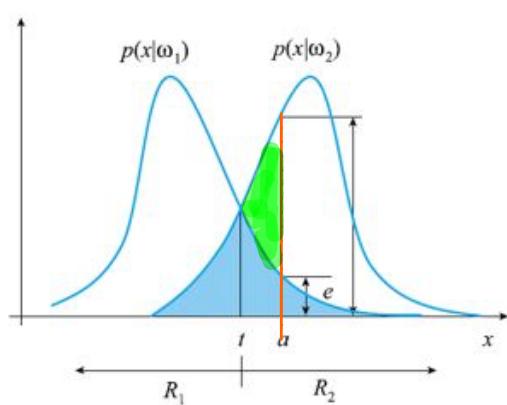
사전 확률이 0.5인 경우 우도만으로 분류

$P(w_1) > P(w_2)$ 인 경우 사전 확률이 의사결정 주도

- 오류확률

$$E = \frac{1}{2} \left(\int_{-\infty}^t P(x|w_2) dx + \int_t^{\infty} P(x|w_1) dx \right)$$

w_1, w_2 의
사전 확률



t 를 오나원으로 옮기면, 오류율 증가, 그러므로 베이시언 분류기는 '오류율 기준으로 최적'

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

2.2.2 최소 위험 베이시언 분류기

오류 확률이라는 기준은 적절하지 않은 상황도 있다.

ex) 암환자와 정상인 분류

손실 행렬 (loss matrix)

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

$$d_1 = C_{11} \int_{R_1} P(x|w_1) dx + C_{12} \int_{R_2} P(x|w_1) dx$$

$$d_2 = C_{21} \int_{R_1} P(x|w_2) dx + C_{22} \int_{R_2} P(x|w_2) dx$$

두 부류를 모두 고려한 평균 손실 D는 d_i 에 w_i 의 사건 확률을 곱한 값을 더하면 된다.

$$\begin{aligned} D &= d_1 P(w_1) + d_2 P(w_2) = \sum_{i=1}^2 d_i P(w_i) \\ &= \left(C_{11} \int_{R_1} P(x|w_1) P(w_1) dx + C_{12} \int_{R_2} P(x|w_1) P(w_1) dx \right) \\ &\quad + \left(C_{21} \int_{R_1} P(x|w_2) P(w_2) dx + C_{22} \int_{R_2} P(x|w_2) P(w_2) dx \right) \\ &= \int_{R_1} \left(\underbrace{C_{11} P(x|w_1) P(w_1)}_{q_1} + \underbrace{C_{21} P(x|w_2) P(w_2)}_{q_2} \right) dx \\ &\quad + \int_{R_2} \left(\underbrace{C_{12} P(x|w_1) P(w_1)}_{q_1} + \underbrace{C_{22} P(x|w_2) P(w_2)}_{q_2} \right) dx \\ &= \sum_{i=1}^2 \int_{R_i} \left(\sum_{j=1}^2 C_{ji} P(x|w_j) P(w_j) \right) dx \end{aligned}$$

χ 를 R_1 에 소속시키면 9, 만큼의 손실이 발생하고,

R_2 에 소속시키면 Q_2 만큼의 손실이 발생한다.

D를 최소화시키는 결정 규칙은 다음과 같다.

x 를 $q_2 > q_1$ 이면 w_1 로 분류하고, $q_1 > q_2$ 이면 w_2 로 분류하라.

$$\text{이때, } Q_1 = C_{11} P(x|W_1) P(W_1) + C_{21} P(x|W_2) P(W_2) \quad] \text{최소 위험}$$

$$Q_2 = C_{12} P(x|W_1) P(W_1) + C_{22} P(x|W_2) P(W_2) \quad] \text{베이시언 분류기}$$

정상인이라고 분류했을 때의 손실과 암환자로 분류했을 때의 손실 비교

위의 식은 x 와 무관하여 미리 계산이 가능한 항들이 있다.

미리 알아보자

$$q_2 > q_1 \quad (c_{ij} > c_{ii}, \quad i \neq j)$$

$$(C_{12} - C_{11}) P(X|W_1) P(W_1) > (C_{21} - C_{22}) P(X|W_2) P(W_2)$$

$$\frac{P(x_1|w_1)}{P(x_1|w_2)} > \frac{(c_{21} - c_{22}) P(w_2)}{(c_{12} - c_1) P(w_1)}$$

우도비
(두 우도의 비율) X와 무관함
임계값

x_2 를 $\frac{P(x|w_1)}{P(x|w_2)}$ > T이면 w_1 로 분류하고,

$\frac{P(x|w_1)}{P(x|w_2)} < T$ 이면 w_2 로 분류하라.

$$T = \frac{(C_{21} - C_{22}) P(W_2)}{(C_{12} - C_{11}) P(W_1)}$$

우도비 결정 규칙

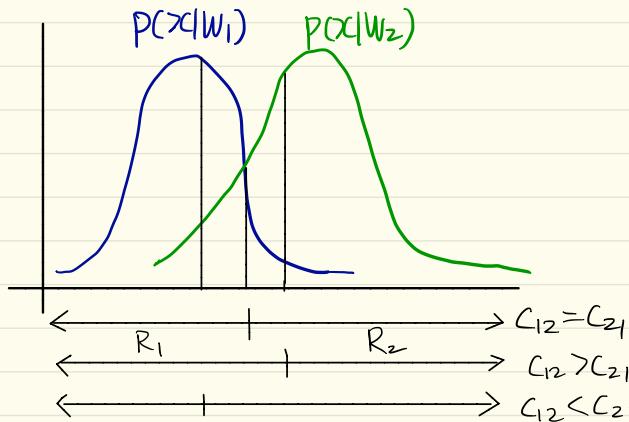
만약, $C_{ii} = 0$ 으로 놓고, 사전 확률도 같다고 하면,

x 를 $P(x|w_1) > P(x|w_2) \frac{C_{21}}{C_{12}}$ 이면 w_1 로 분류하고,

$P(x|w_1) < P(x|w_2) \frac{C_{12}}{C_{21}}$ 이면 w_2 로 분류하라.

만약, 암환자(w_1)와 정상인(w_2)을 구별하는 경우라면,

$C_{12} > C_{21}$ 일 터고, 결정경계 t 를 오른쪽으로 옮기라는 것이다.



2.2.3 M 부류로 확장

M부류 최소 오류 베이시언 분류기

x 를 $k = \arg \max p(w_i|x)$ 일 때, w_k 로 분류하라.

$$\arg \max p(x|w_i)p(w_i)$$

$$D = \sum_{i=1}^M \int_{R_i} \left(\sum_{j=1}^M C_{ji} p(x|w_j) p(w_j) \right) dx$$

$$= \int_{R_1} \left(\sum_{j=1}^M C_{j1} p(x|w_j) p(w_j) \right) dx + \cdots + \int_{RM} \left(\sum_{j=1}^M C_{jM} p(x|w_j) p(w_j) \right) dx$$

$q_1 \quad \quad \quad q_M$

평균 손실 D를 최소화하기 위해 x 를 R_1, R_2, \dots, R_M 중 어디에 소속 시킬지를 결정.

M 부류 최소 유형 베이시언 부류기

$$q_i = \sum_{j=1}^M c_{ji} P(x|w_j) P(w_j)$$

2.3 분별함수

x 를 $k = \arg \max g_i(x)$ 일 때 w_k 로 분류하라

$$g_i(x) = \frac{P(x|w_i) P(w_i)}{\sum_{j=1}^M c_{ji} P(x|w_j) P(w_j)}$$

(최소 오류 베이시언 분류기)
(최소 유형 베이시언 분류기)

입력: 특징 벡터 x / 출력: 부류

for($i=1$ to M)

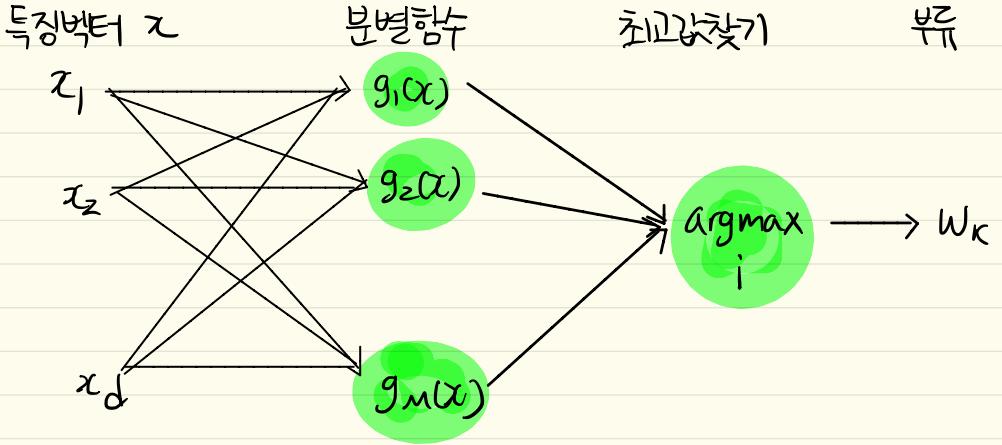
if(최소 오류) $g_i(x) = P(x|w_i) P(w_i);$

else if(최소 유형) $g_i(x) = 1 / \sum_{j=1}^M c_{ji} P(x|w_j) P(w_j);$

$g_i(x), 1 \leq i \leq M$ 중 가장 큰 것을 찾아 그것의

첨자를 k 라 하자.

return w_k



* 분별 함수의 장점

- 여러 분류기를 하나로 표현
- $f(\cdot)$ 가 단조 증가라면 $P(x|w_i) \cdot P(w_i)$ 대신
 $g(x) = f(P(x|w_i)P(w_i))$ 사용해도 같은 결과
- 패턴 인식에서 가장 많이 사용하는 $f(\cdot)$ 함수는 \log 이다.
 $(\log$ 함수는 곱하기 연산을 더하기로 바꾸어 주으로 수식을 간단히 정리하는데 도움, 값의 규모가 커져 수치 오류에 둔감한 이점)
- $M=2$ 인 이진분류 경우에는 분별함수 하나만 사용!

$x \in g_{1,2}(x) > 0$ 이면 w_1 , $g_{1,2}(x) < 0$ 이면 w_2 로 분류

$$g_{1,2}(x) = g_1(x) - g_2(x)$$

2.4 정규분포에서 베이시언 분류기

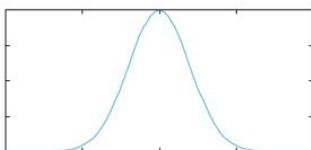
사후 확률 계산을 위해 사전 확률 $P(w_i)$ 와 우도 $P(x|w_i)$ 의 곱으로 대치해 계산한다. 이제 training set을 이용해 사전 확률과 우도를 알아내야 한다.

training set이 정규분포로 모델링이 가능한 형태라는 전제 하에 베이시언 분류기를 만들어 본다.

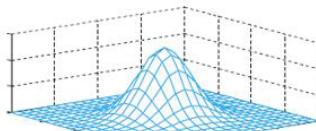
- 정규분포 (가우시언 분포)
 - 현실 세계에 맞는 경우가 있음
 - 평균과 분산이라는 두 종류의 매개 변수만으로 표현 가능

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

→ d차원에서의 $\boldsymbol{\mu}$ 는 평균 벡터이고, $\boldsymbol{\Sigma}$ 는 $d \times d$ 크기의 공분산 행렬이다



(a) 1 차원 정규 분포



(b) 2 차원 정규 분포

그림 2.9 정규 분포의 예

• 유도를 다시 쓰면,

e^{\square}

$$p(x|w_i) = N(\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right)$$

• 로그를 취하여 분별함수를 만들어 보면,

$$g_i(x) = \ln(p(x|w_i) p(w_i))$$

$$= \ln N(\mu_i, \Sigma_i) + \ln p(w_i)$$

$$= -\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln p(w_i)$$

$$= -\frac{1}{2}\left(x^T \Sigma_i^{-1} x - x^T \Sigma_i^{-1} \mu_i - \mu_i^T \Sigma_i^{-1} x + \mu_i^T \Sigma_i^{-1} \mu_i\right) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln p(w_i)$$

$$= -\frac{1}{2}(x^T \Sigma_i^{-1} x - 2\mu_i^T \Sigma_i^{-1} x + \mu_i^T \Sigma_i^{-1} \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln p(w_i)$$

$g_i(x)$ 는 변수 x 에 대한 2차식!

예제 2.4

□ $d=2$ 이고 아래와 같다고 가정

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

□ 분별 함수를 유도해 보면,

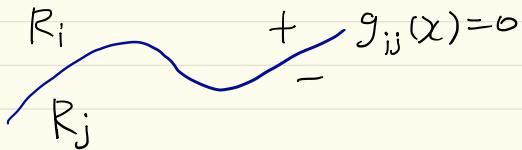
$$\begin{aligned} g_i(\mathbf{x}) &= -\frac{1}{2} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right)^T \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right) - \frac{2}{2}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

결정 경계

- 두 부류가 차지하는 영역의 경계

$g_i(x) = g_j(x)$ 인 점, 즉 $g_{ij}(x) = 0$ 인 점

$$g_{ij}(x) = g_i(x) - g_j(x)$$



2.4. 2 선형분류

모든 부류의 공분산 행렬이 같은 경우, $\leq_i = \leq$ 로 표기

변수 x 에 대한 1차식이 됨!

$$= (\Sigma^{-1} \mu_i)^T x + (\ln p(w_i) - \frac{1}{2} \mu_i^T \Sigma^{-1} \mu_i)$$

$$= w_i^T x + b_i$$

$$g_{ij}(x) = g_i(x) - g_j(x)$$

$$= (\zeta^{-1}(\mu_i - \mu_j))^T x + (\ln p(w_i) - \ln p(w_j)) - \frac{1}{2} M_i^T \zeta^{-1} \mu_i + \frac{1}{2} M_j^T \zeta^{-1} \mu_j$$

$$= \frac{(\mu_i - \mu_j)^\top}{w} \left(x - \left(\frac{1}{2} (\mu_i + \mu_j) - \frac{\mu_i - \mu_j}{(\mu_i - \mu_j)^\top \Sigma^{-1} (\mu_i - \mu_j)} \ln \frac{p(w_i)}{p(w_j)} \right) \right)$$

$$= w^T(w - w_0)$$

결정 경계상의 점 x 는 $g_{ij}(x) = 0$ 이어야 하는데, $g_{ij}(x)$ 가 1차식이므로 결정 경계가 초평면이라는 사실을 알 수 있다.

두 부류 W_i 와 W_j 가 공분산이 같아서 모양이 같은 타원으로 표현했다. 앞의 수식에 따르면 결정직선은 x_0 를 지나며, $\Sigma^{-1}(\mu_i - \mu_j)$ 에 수직이어야 한다.

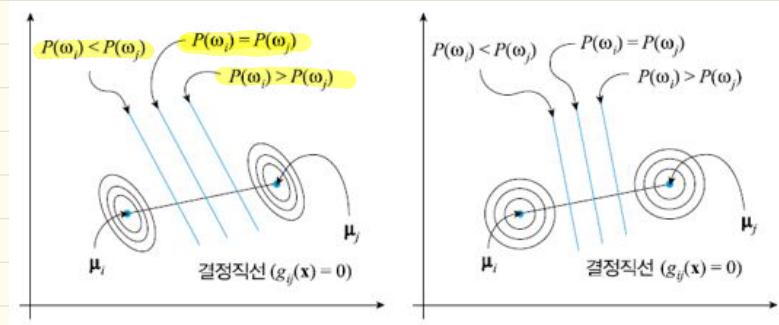


그림 2.11 선형 분류기

* 두 부류의 사전확률 같다($\ln P(W_i)/P(W_j) = 0$)

$x_0 = \frac{1}{2}(\mu_i + \mu_j)$, 즉 결정 직선은 두 평균 벡터 μ_i 와 μ_j 의 중점을 지나야 한다.

LDA

(linear discriminant analysis)

이러한 고장을 거쳐 선형 분류기를 만드는 방법: 선형 별별 분석

공분산 행렬이 $\leq = \Delta^2 I$ 인 경우, 대각선 요소는 Δ^2 을 갖고, 나머지 요소는 0을 갖는 행렬.

$\leq^{-1} = (1/\Delta^2)I$, 이 상황에서는 정규분포의 등고선이 단원이 아닌 원을 만든다. 이제 결정 직선은 $\mu_i - \mu_j$ 에 수직이다.

$$g_{ij}(x) = \frac{1}{\Delta^2} (\mu_i - \mu_j)^T \left(x - \left(\frac{1}{2}(\mu_i + \mu_j) - \frac{\Delta^2(\mu_i - \mu_j)}{\|\mu_i - \mu_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} \right) \right)$$

$$= W^T (x - x_0)$$

■ 예제 2.5

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

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

$$\mu_1 = \frac{1}{4} \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 5 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \end{pmatrix} \right) = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 8/3 & 0 \\ 0 & 2/3 \end{pmatrix}$$

$$\mu_2 = \frac{1}{4} \left(\begin{pmatrix} 6 \\ 6 \end{pmatrix} + \begin{pmatrix} 8 \\ 5 \end{pmatrix} + \begin{pmatrix} 10 \\ 6 \end{pmatrix} + \begin{pmatrix} 8 \\ 7 \end{pmatrix} \right) = \begin{pmatrix} 8 \\ 6 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 8/3 & 0 \\ 0 & 2/3 \end{pmatrix}$$

$$g_{12}(x) = \left(\begin{pmatrix} 3/8 & 0 \\ 0 & 3/2 \end{pmatrix} \begin{pmatrix} 3-8 \\ 2-6 \end{pmatrix} \right)^T x$$

$$+ \left(\ln P(\omega_1) - \ln P(\omega_2) - \frac{1}{2} (3-2) \begin{pmatrix} 3/8 & 0 \\ 0 & 3/2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} + \frac{1}{2} (8-6) \begin{pmatrix} 3/8 & 0 \\ 0 & 3/2 \end{pmatrix} \begin{pmatrix} 8 \\ 6 \end{pmatrix} \right)$$

$$= \begin{pmatrix} -15/8 & -6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (\ln P(\omega_1) - \ln P(\omega_2) + 34.3125)$$

$$= -15/8 x_1 - 6x_2 + (\ln P(\omega_1) - \ln P(\omega_2) + 34.3125)$$

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

$$P(\omega_1) = 0.8, P(\omega_2) = 0.2 \text{ 일 경우: } 5x_1 + 16x_2 - 95.197 = 0$$

$$P(\omega_1) = 0.2, P(\omega_2) = 0.8 \text{ 일 경우: } 5x_1 + 16x_2 - 87.803 = 0$$

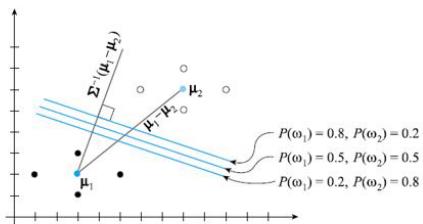


그림 2.12 선형 분별 분석 (LDA) 사례

2.4.3 2차 분별

공분산 제약 없을 때

$$g_{ij}(x) = g_i(x) - g_j(x)$$

2차식 = 2차식 - 2차식

결정곡선은 쌍곡선도 타원도 된다.

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

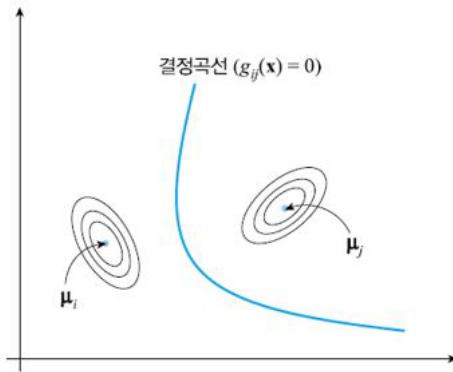


그림 2.13 2차 분별 분석으로 만든 2차 분류기

예제 2.6

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

$$\omega_2: (7,6)^T, (8,4)^T, (9,6)^T, (8,8)^T$$

$$\mu_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} 8/3 & 0 \\ 0 & 2/3 \end{pmatrix}$$

$$\mu_2 = \begin{pmatrix} 8 \\ 6 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 2/3 & 0 \\ 0 & 8/3 \end{pmatrix}$$

$$g_{12}(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}) = \frac{9}{16}x_1^2 - \frac{9}{16}x_2^2 - \frac{87}{8}x_1 + \frac{3}{4}x_2 + \frac{801}{16} + \ln P(\omega_1) - \ln P(\omega_2)$$

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

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

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

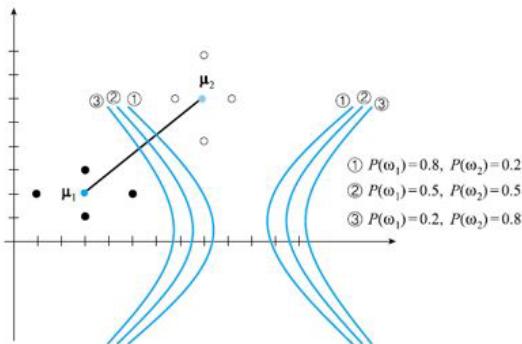


그림 2.14 2차 분별 분석 (QDA) 사례

2.4.4 최소 거리 분류기

- 최소거리 분류기로 다시 해석해 보자

수식 유도 편의를 위해 두 뷰의 사전 확률과 공분산 행렬 같다고 가정.

$$g_i(x) = -\frac{1}{2}(x - \mu_i)^T \Sigma^{-1} (x - \mu_i)$$

• 최소 거리 분류기

$$x \in k = \arg \min \left((x - \mu_i)^T \Sigma^{-1} (x - \mu_i) \right)^{\frac{1}{2}}$$

* 거리 척도

마할라노비스 거리: $((x - \mu_i)^T \Sigma^{-1} (x - \mu_i))^{1/2}$

유클리디언 거리 $((x - \mu_i)^T (x - \mu_i))^{1/2} = \|x - \mu_i\|$
 $(\Sigma = \lambda^2 I)$ 인 경우

예제 2.7

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

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

$$x = (8,2)^T$$

$$\mu_1 \text{까지의 마할라노비스 거리} = \left((8-3) \begin{pmatrix} 3/8 & 0 \\ 0 & 3/2 \end{pmatrix} (8-3) \right)^{1/2} = 3.062$$

$$\mu_2 \text{까지의 마할라노비스 거리} = \left((8-8) \begin{pmatrix} 3/8 & 0 \\ 0 & 3/2 \end{pmatrix} (8-8) \right)^{1/2} = 4.899$$

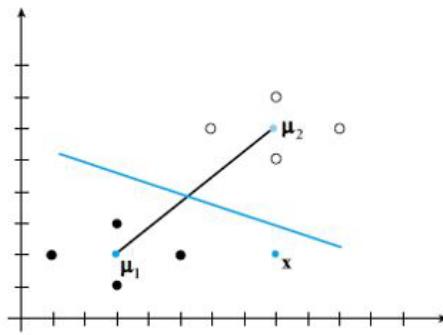


그림 2.15 두 통계 분포까지의 마할라노비스 거리

2.5 베이시언 분류의 특성

1. 비현실성을 내포 : 일반적인 확률분포를 사용하려면 차원의 저주 발생, 정규분포를 가정하려면 실제 확률분포와 차이 발생
2. 실제 확률분포를 있다고 가정하면 오류율 측면에서 최적!
3. M개 부류 각각에 대해 그에 속한 확률 출력, 즉,
 $g_i(x)$ 를 확률로 해석할 수 있다. 따라서 이 확률을 신뢰도 값으로
삼아 후처리 등에 활용할 수 있다.

특징들이 모두 "독립"이라는 가정을 하고, 차원의 저주를
피하는 방법

$$P(x|w_i) = \prod_{j=1}^d P(x_j|w_i)$$

나이브 베이시언 분류기

운하는,

$$w_k = \underset{w_i}{\operatorname{argmax}} P(x_1, x_2, \dots, x_d | w_i) P(w_i)$$

→ 차원의 저주

- 우도 계산을 \uparrow 로 함

차원의 저주를 피했으나, 성능 저하

2.6 기각처리

- 신뢰도가 충분하지 않은 경우는
의사결정 포기

그림에서 두 부류의 확률 차이가 작보다
작으면 기각

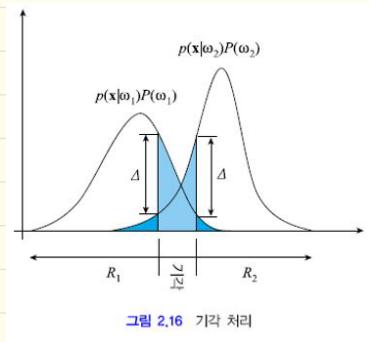


그림 2.16 기각 처리