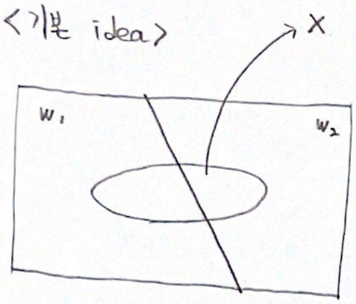
① LDA

* 분류를 위한 베이즈 정리 사용.

<기본 idea>



① 데이터 (X)로 부터 사전 확률 $\overset{1)}{P(W_1)}$, $\overset{2)}{P(W_2)}$ 도출.

· 1) $P(w_1) = \dfrac{w_1}{w_1+w_2}$, 2) $P(w_2) = \dfrac{w_2}{w_1+w_2}$

(범주가 w_1인 데이터 개수, w_2인 데이터 개수를 각각 전체 데이터 개수로 나눠주기)

② 새로운 데이터 (x)가 주어졌을 때, x가 어느 클래스에 속할지 맞추기.

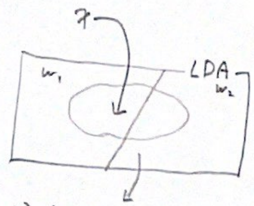
$$P(w_i|x) = \frac{P(x|w_i)\cdot P(w_i)}{P(x)} = \frac{P(x|w_i)\cdot P(w_i)}{P(x|w_1)\cdot P(w_1) + P(x|w_2)\cdot P(w_2)}$$

<분류 메커니즘>

· 새로운 데이터 x 등장.

· LDA 모델 → $P(w_1|x)$, $P(w_2|x)$ 각각 도출.

· 둘 중에 높은 확률이 나오는 곳에 x를 배정.



1) $P(w_1|x) > P(w_2|x) \rightarrow x = w_1$

2) $P(w_1|x) < P(w_2|x) \rightarrow x = w_2$

<개념>

① 벡터, ② 공분산, ③ 행렬, ④ 정규분포 밀도함수

선형대수학 · 전치행렬 (행과 열을 서로 맞바꾼 행렬) — transposed

역행렬 = AB = BA = E (A^{-1} 역행렬, 단위행렬)

평균 = μ, 분산 = σ^2 → 표준화 $\mu=0$, $\sigma^2=1$ → 다변량 정규분포 : 정규분포를 다차원으로 확장.

* 정규분포를 따르는 N개의 데이터가 K개 범주로 구성되어 있는 P차원 (변수 개수 = P개)의 데이터 X의 평균과 공분산

① 벡터

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}$$

column vector

② 공분산

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \cdots & \cdots & \cdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{bmatrix}$$

* $COV(X,Y) = E((X-\mu_X)(Y-\mu_Y))$

↳ X와 Y 두 변수의 변동을 나타냄.

$COV(X,Y) > 0$: X가 증가할 때 Y도 증가 (정의 관계)

〃 < 0 : 〃 Y 감소 (부의 관계)

〃 $= 0$: 서로 독립.

③ 행렬

1) Transposed Matrix ⇒ W^T, W 행렬의 전치행렬.

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & -5 & 6 & 8 \end{bmatrix} \quad A^T = \begin{bmatrix} 1 & -3 \\ 1 & -5 \\ 1 & 6 \\ 1 & 8 \end{bmatrix}$$

2) Inverse Matrix

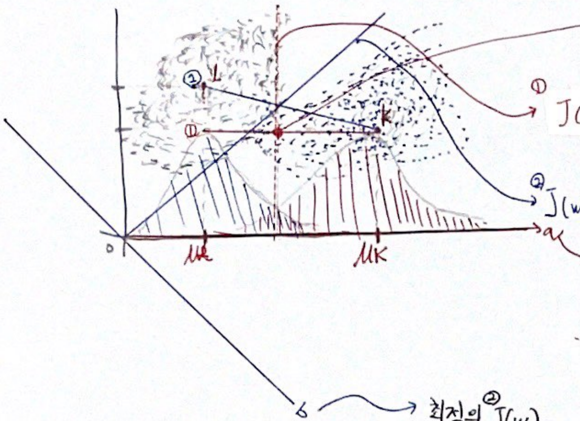
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$ 단, $\dfrac{1}{ad-bc} \neq 0$.

$$A \cdot A^{-1} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{단위행렬}} = E$$

② LDA. (쉬운 ver)

① $J(w) = |\tilde{m}_1 - \tilde{m}_2| = |w^T(\mu_1 - \mu_2)|$: 중심(평균)간의 거리 → 목적함수 = $\boxed{\text{'평균의 차이' 최대}}$ ✻

→ ①평균만 고려한다면, ②표준편차가 고려되지 않으므로 좋은 척도가 아님



b : K와 l의 확률 평균

① $J(w)$: 목적함수, $\left(\log\frac{\pi_k}{\pi_l}\right) - \underbrace{\frac{1}{2}(\mu_k+\mu_l)^T}\Sigma^{-1}(\mu_k-\mu_l) + \underset{x}{x^T}\Sigma^{-1}(\mu_k-\mu_l)$

[μ_k : K그룹의 평균
μ_l : l 그룹의 평균

$\Downarrow$

$Y = ax - ab$

$Y = a(x-b)$

② $J(w)$

$a = \Sigma^{-1}(\mu_k - \mu_l)$: a라는 벡터축

∴ LDA란 (분류) 최적의 (클래스를 나누는) 목적함수 $J(w)$를 찾아 내어 분류하는 것.
〃
decision boundary, 결정경계

최적의 ②$J(w)$

<확률모형적 접근 – 베이지안 룰>

② · Y가 K개의 범주를 가질 때, $k = 1, 2, \cdots, K$

LDA 가정. ① 각 클래스 집단은 정규분포 형태의 확률분포를 갖는다.
② · 〃 비슷한 형태의 공분산 구조를 갖는다.

· $y = k$ 일 때, 공분산 구조를 갖는 Σ_K를 갖는 P개의 정규분포 변수의 분포.

$$f_k(x) = \frac{1}{(\sqrt{2\pi})^P \cdot \sqrt{|\Sigma_k|}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \cdot \Sigma_k^{-1}(x-\mu_k)\right)$$

✻ 정규분포 밀도함수

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

✻ 다변량 정규분포, μ(평균), Σ(공분산)

$$f(x) = |2\pi\Sigma|^{-\frac{1}{2}} \cdot \exp\left\{\frac{-(x-\mu)^T \cdot \Sigma^{-1}(x-\mu)}{2}\right\}$$

→ 이때, 아래의 로그 값이 클수록, l 대비 k 범주에 속할 확률이 높아진다.

$$\log\frac{P(k|x)}{P(l|x)} = \log\frac{f_k(x)}{f_l(x)} + \log\frac{\pi_k}{\pi_l}$$

→ 위의 식은, k와 관계없는 공분산 구조 Σ를 갖기 때문에, 아래와 같은 x에 대한 1차식으로 정리 可.

$$\log\frac{P(k|x)}{P(l|x)} = \log\frac{f_k(x)}{f_l(x)} + \log\frac{\pi_k}{\pi_l} = \log\frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k+\mu_l)^T \cdot \Sigma^{-1}(\mu_k-\mu_l) + x^T \cdot \Sigma^{-1}(\mu_k-\mu_l).$$

· (델타) $\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log\pi_k$ 라고 하면,

$$\log\frac{P(k|x)}{P(l|x)} = \delta_k(x) - \delta_l(x)$$

= ①분산은 최소화, ②평균은 최대화

③ LDA & QDA (어려운 ver.)

· 다변량 정규분포.

$$P(x|w_i) = \frac{1}{(2\pi)^{P/2} \cdot |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right]$$

· 판별함수 $\overset{\sim\text{델타}}{\delta_i}(x) = P(w_i|x) \to P(x|w_i)\cdot P(w_i) \to \ln P(x|w_i) + \ln P(w_i)$

↳ 새로운 데이터 x에 대한 범주를 확인.

· 판별함수를 다변량 정규분포에 대입.

판별함수 Sol) $\ln P(x|w_i) = \ln \frac{1}{(2\pi)^{P/2}|\Sigma_i|^{1/2}} - \frac{1}{2}(x-\mu_i)^T \cdot \Sigma^{-1}(x-\mu_i)$

$$= \ln\frac{1}{(2\pi)^{P/2}} + \ln\frac{1}{|\Sigma|^{1/2}} - \frac{1}{2}(x-\mu_i)^T\cdot\Sigma^{-1}(x-\mu_i)$$

$$= -\frac{P}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-\mu_i)^T\cdot\Sigma^{-1}(x-\mu_i) \quad : \text{상수.}$$

알고있는값

결정경계 Sol) · $\delta_1(x) - \delta_2(x) = 0$

· $-\frac{P}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_1| - \frac{1}{2}(x-\mu_1)^T\cdot\Sigma_1^{-1}(x-\mu_1) + \ln P(w_1) - \left(-\frac{P}{2}\ln 2\pi - \frac{1}{2}|\Sigma_2| - \frac{1}{2}(x-\mu_2)^T\cdot\Sigma_2^{-1}(x-\mu_2) + \ln P(w_2)\right) = 0$

· $(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) - (x-\mu_2)^T\cdot\Sigma_2^{-1}(x-\mu_2) - \ln\frac{|\Sigma_2|}{|\Sigma_1|} + 2\ln\frac{P(w_1)}{P(w_2)} = 0$

1) $\Sigma_1 = \Sigma_2$ 가정. (각 범주의 공분산이 동일하다고 가정) $\to$ LDA

by: Pooled Variance (합동분산) $\Sigma_1 = \Sigma_2 = \Sigma = \left[\frac{n_1 - 1}{(n_1-1)(n_2-1)}\right]\cdot\Sigma_1 + \left[\frac{n_2-1}{(n_1-1)(n_2-1)}\right]\Sigma_2$

위의 식을 판별함수에 대입.

$$(\mu_1-\mu_2)^T\Sigma^{-1}x - \frac{1}{2}(\mu_1-\mu_2)^T\cdot\Sigma^{-1}(\mu_1+\mu_2) - \ln\frac{P(w_1)}{P(w_2)} = 0$$

$\Rightarrow$ $Ax + b = 0$ 의 선형식 (Linear Equation) 도출.

: x를 넣어 0보다 큰 값이 나오면 w_1 범주 그렇지 않으면 w_2 범주로 분류.

2) $\Sigma_1 \neq \Sigma_2$ 가정. $\to$ QDA (Quadratic Discriminant Analysis)

Sol)

$$-\frac{1}{2}x^T(\Sigma_1^{-1} - \Sigma_2^{-1})\cdot x + (\mu_1\Sigma_1^{-1} + \mu_2\Sigma_2^{-1})x - \frac{1}{2}(\mu_1^2\Sigma_1^{-1} - \mu_2^2\Sigma_1^{-1}) + \frac{1}{2}\ln\frac{|\Sigma_2|}{|\Sigma_1|} - \ln\frac{P(w_1)}{P(w_2)} = 0$$

$\Rightarrow Ax^2 + bx + c = 0$ 의 2차식 처럼 보임.