Machine Learning

5장 Discriminant analysis와 Naive Bayes model

고려대학교 통계학과
박유성

Contents

01 Bayes rule을 이용한 classification

- $d \times 1$인 입력변수 $\boldsymbol{x}$ 가 주어져 있을 때 목표변수 y가 class k $(k = 1, \ldots, K)$ 에 속할 확률은 Bayes rule에 의해,

$$P(y=k|\boldsymbol{x}) = \frac{P(\boldsymbol{x}|k)P(k)}{P(\boldsymbol{x})} = \frac{P(\boldsymbol{x}|k)P(k)}{\sum_{k=1}^{K} P(\boldsymbol{x}|k)P(k)} \tag{5.1}$$

  이 됨.

- $P(\boldsymbol{x}|k)$ 분포 가정에 따라 방법이 달라짐.

02 Discriminant analysis

- $P(\boldsymbol{x}|y=k) \sim N_d(\boldsymbol{\mu}_k, \Sigma_k)$ 라고 가정

  1) $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \Sigma$

     → LDA (Linear discriminant analysis)

  2) $i \neq j$ 에 대하여 $\Sigma_i \neq \Sigma_j$ 가 존재

     → QDA (Quadratic discriminant analysis)

LDA (Linear Discriminant Analysis)

- $P(\boldsymbol{x}|y=k) \sim N_d(\boldsymbol{\mu}_k, \Sigma_k)$ 이므로 (5.1)식에서

$$P(\boldsymbol{x}|k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T\Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)) \qquad (5.2)$$

- $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_k = \Sigma$ 이므로 class k와 l에 대하여

$$\log\frac{P(y=k|\boldsymbol{x})}{P(y=l|\boldsymbol{x})} = \log\frac{P(\boldsymbol{x}|k)P(y=k)}{P(\boldsymbol{x}|l)P(y=l)}$$

$$= \log\frac{P(y=k)}{P(y=l)} - \frac{1}{2}(\boldsymbol{\mu}_k + \boldsymbol{\mu}_l)^T\Sigma^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_l) + \boldsymbol{x}^T\Sigma^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_l) \qquad (5.3)$$

LDA (Linear Discriminant Analysis)

- $P(y=k)=\pi_k$ 로 표기하고

$$\delta_k(\boldsymbol{x})=\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{\mu}_k-\frac{1}{2}\boldsymbol{\mu}_k^T\Sigma^{-1}\boldsymbol{\mu}_k+\log\pi_k \tag{5.4}$$

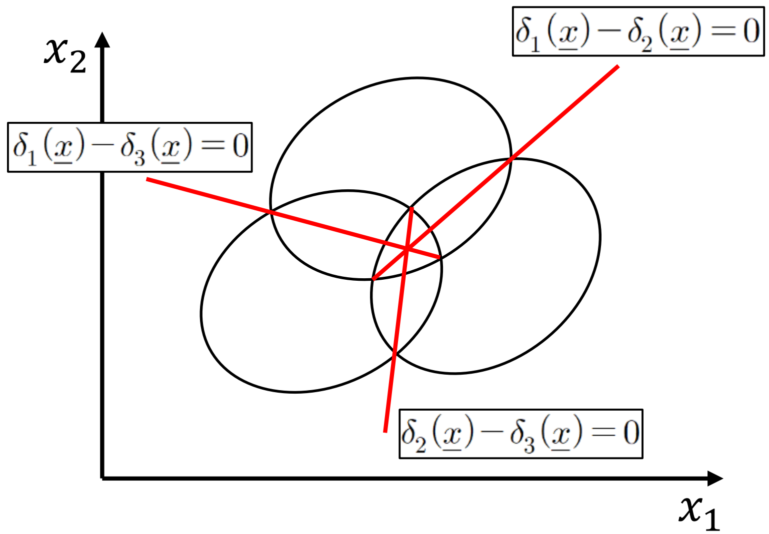
로 정의하면

- (5.3) 식은 $\log\frac{P(y=k|\boldsymbol{x})}{P(y=l|\boldsymbol{x})}=\delta_k(\boldsymbol{x})-\delta_l(\boldsymbol{x})$로 표현할 수 있음.
- $\delta_k(\boldsymbol{x})$ 는 $\boldsymbol{x}$ 에 linear이므로 선형판별함수

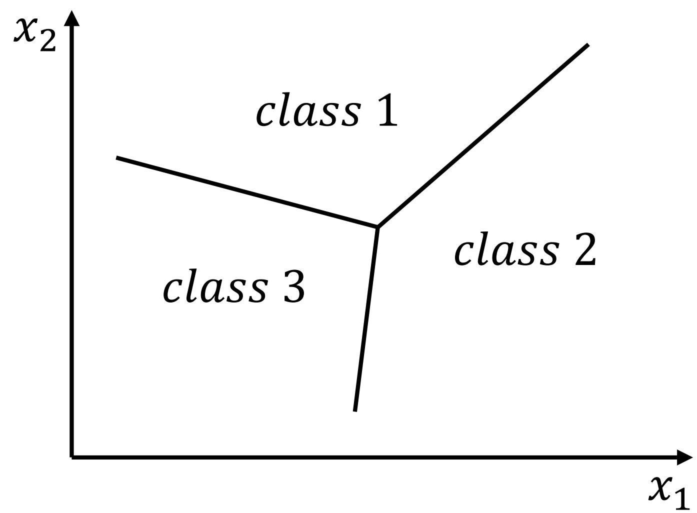
$\delta_k(\boldsymbol{x})-\delta_l(\boldsymbol{x})>0$ 은 $P(y=k|\boldsymbol{x})>P(y=l|\boldsymbol{x})$을 의미

→ $\delta_k(\boldsymbol{x})$를 선형판별함수(Linear discriminant function)이라 부름.

LDA (Linear Discriminant Analysis)



<판별 함수의 경계>



<판별 영역>

- 위 그림은 2개의 feature x_1, x_2가 존재할 때 3개의 class를 구분하는 경계선을 보여주고 있음.

LDA (Linear Discriminant Analysis)

- $\delta_k(\boldsymbol{x})$ 의 추정

  n : training data의 총 관측치 수

  n_k : $class\ k$에 속하는 관측치 수, K : 총 $class$ 수

  $\boldsymbol{x}_i^{(k)}$: $class\ k$에 속한 i번째 $d \times 1$ feature 라고 할 때,

  $$\widehat{\pi}_k = n_k/n, \ \widehat{\boldsymbol{\mu}}_k = \sum_{i=1}^{n_k} \boldsymbol{x}_i^{(k)}/n_k, \ \widehat{\Sigma} = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(\boldsymbol{x}_i^{(k)} - (\widehat{\boldsymbol{\mu}}_k))(\boldsymbol{x}_i^{(k)} - (\widehat{\boldsymbol{\mu}}_k))^T/(n-K)$$

  를 사용하여 $\delta_k(\boldsymbol{x})$ 를 추정하게 됨.

- 이와 같이 선형판별함수를 이용하여 y의 $class$를 판별하는 것이 LDA (Linear Discriminant Analysis)임.

QDA (Quadratic Discriminant Analysis)

- $P(\boldsymbol{x}|y=k) \sim N_d(\boldsymbol{\mu}_k, \Sigma_k)$이면서 $i \neq j$에 대하여 $\Sigma_i \neq \Sigma_j$가 존재한다고 가정
- 그러므로 $\delta_k(\boldsymbol{x}) = -\frac{1}{2}log|\Sigma_k| - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T\Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k) + \log\pi_k$ 가 되며

  이를 이차판별함수(Quadratic discriminant function)이라 함.
- $\delta_k(\boldsymbol{x})$ 에 대해 추정 시 $\widehat{\Sigma_k} = k\ class$ 내 feature들의 표본공분산 행렬임.
- 이와 같이 이차판별함수를 이용하여 y의 $class$를 판별하는 것이 QDA (Quadratic Discriminant Analysis)임.

03 Naïve bayes model

- $\boldsymbol{x} = (x_1, \ldots, x_d)^T$ 가 서로 간에 독립이라고 가정

  (즉, $P(\boldsymbol{x}|k) = \prod_{j=1}^{d} P(x_j|k)$ 라고 가정)

  1) $x_j \sim N(\mu_k, \sigma_k^2)$ $(j = 1, \ldots, d)$라고 가정(즉, LDA에서 $\Sigma_k = \sigma_k^2 I$)

  → Normal naïve bayes model

  2) $\boldsymbol{x}$ 를 d 차원 multinomial distribution으로 가정

  → Multinomial naïve bayes model

Q & A