

통계적 결정이론과 확률밀도함수의 추정

Likelihood Ratio Test : 우도비 검증 (LRT)

■ 가우시안 혼합 모델(GMM)

$$g_m(x), g_f(x)$$

$$\text{유사 정도} : \frac{g_m(x)}{g_f(x)}$$

$P(\omega_i | x)$ (a posterior probability : 클래스의 사후 확률)

■ 만약, $P(\omega_1 | x) > P(\omega_2 | x)$ 라면 ω_1 을 선택, 아니면 ω_2 선택

$$\begin{array}{c} \omega_1 \\ > \\ P(\omega_1 | x) & P(\omega_2 | x) \\ < \\ \omega_2 \end{array}$$

■ 베이즈 정리 이용

$$\begin{array}{c} \omega_1 \\ > \\ \frac{P(x|\omega_1)P(\omega_1)}{P(x)} & \frac{P(x|\omega_2)P(\omega_2)}{P(x)} \\ < \\ \omega_2 \end{array}$$

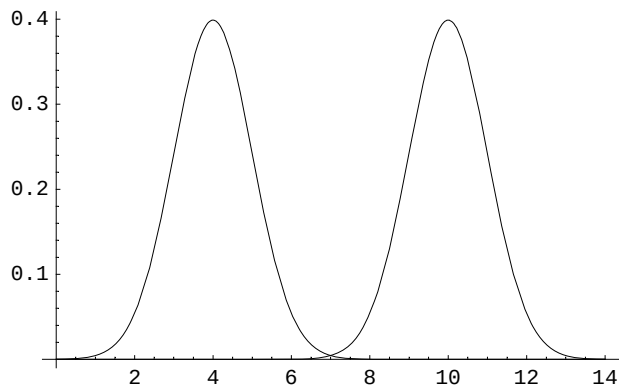
■ $P(x)$ 결정 규칙

$$\begin{array}{c} \omega_1 \\ > \\ \Lambda(x) = \frac{P(x|\omega_1)}{P(x|\omega_2)} & \frac{P(\omega_2)}{P(\omega_1)} \\ < \\ \omega_2 \end{array}$$

$\Lambda(x)$: Likelihood Ratio (우도비)

ex

`Plot[{ $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-4)^2}$ ,  $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-10)^2}$ }, {x, 0, 14}, PlotRange -> All]`



- Graphics -

오류 확률 (Bayes Error Rate)

두 영역 (R_1, R_2) 으로 잘못 분류되는 경우
 ω_1 에 속하는 특징벡터 x 를 R_2 로 결정할 경우
 ω_2 에 속하는 특징벡터 x 를 R_1 로 결정할 경우
 이 사건은 상호 배타적이므로 총 오류 발생 확률

$$\begin{aligned} P(\text{error}) &= P(x \in R_2, \omega_1) + P(x \in R_1, \omega_2) \\ &= P(x \in R_2 | \omega_1) P(\omega_1) + P(x \in R_1 | \omega_2) P(\omega_2) \\ &= \int_{R_2} P(x | \omega_1) P(\omega_1) dx + \int_{R_1} P(x | \omega_2) P(\omega_2) dx \end{aligned}$$

$$P(\text{error}) = \sum_{i=1}^c P(\text{error} | \omega_i) P(\omega_i)$$

$$P(\text{error} | \omega_i) = P(\text{choose } \omega_j | \omega_i) = \int_{R_j} P(x | \omega_i) dx$$

$$P[\text{error}] = P[\omega_1] \int_{R_2} P[x | \omega_1] dx + P[\omega_2] \int_{R_1} P[x | \omega_2] dx$$

$$\epsilon_i = \int_{R_j} P[x | \omega_i] dx$$

$$P(\text{error}) = (\epsilon_1 + \epsilon_2) / 2$$

■ 오류 확률 : $P(\text{error})$

사후 확률 : $P(\text{error}|x)$

$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error} | x) P(x) dx$$

최적의 결정경계는 오류 확률이 최소가 되는 결정경계를 구하는 것이므로
 위의 적분이 최소가 되기 위해서는 $P(\text{error} | x)$ 가 최소가 되어야 할 것이다.

Bayes Risk (베이지 위험)

베이지 분류기에 Cost (비용) 적용

class ω_1 , class ω_2 분류시 벌점 C_{ij} 적용

C_{ij} 는 ω_j 가 실제 속한 클래스일 때, ω_i 클래스를 선택하는데 따른 비용

$$R = E[C] = \sum_{i=1}^2 \sum_{j=1}^2 C_{ij} \cdot P[\text{Choose } \omega_i, X \in \omega_j] = \sum_{i=1}^2 \sum_{j=1}^2 C_{ij} \cdot P[X \in R_i | \omega_j] \cdot P[\omega_j]$$

Bayes Risk을 최소화하는 결정규칙은 어떻게 표현될수 있나?

$$P[X \in R_i | \omega_j] = \int_{R_i} P[X | \omega_j] dX$$

$$R = \int_{R_1} [C_{11} P[\omega_1] P[X | \omega_1] + C_{12} P[\omega_2] P[X | \omega_2]] dX + \int_{R_2} [C_{21} P[\omega_1] P[X | \omega_1] + C_{22} P[\omega_2] P[X | \omega_2]] dX$$

Likelihood (우도) 의 합은 항상 1

$$\int_{R_1} P[X | \omega_1] dX + \int_{R_2} P[X | \omega_1] dX = \int_{R_1 \cup R_2} P[X | \omega_1] dX = 1$$

$$\begin{aligned} R &= C_{11} P[\omega_1] \int_{R_1} P[X | \omega_1] dX + C_{12} P[\omega_2] \int_{R_1} P[X | \omega_2] dX \\ &+ C_{21} P[\omega_1] \int_{R_2} P[X | \omega_1] dX + C_{22} P[\omega_2] \int_{R_2} P[X | \omega_2] dX \\ &+ C_{21} P[\omega_1] \int_{R_1} P[X | \omega_1] dX + C_{22} P[\omega_2] \int_{R_1} P[X | \omega_2] dX \\ &- C_{21} P[\omega_1] \int_{R_1} P[X | \omega_1] dX - C_{22} P[\omega_2] \int_{R_1} P[X | \omega_2] dX \end{aligned}$$

$$\begin{aligned} R &= C_{21} P[\omega_1] + C_{22} P[\omega_2] \\ &+ (C_{11} - C_{21}) P[\omega_1] \int_{R_1} P[X | \omega_1] dX + (C_{12} - C_{22}) P[\omega_2] \int_{R_1} P[X | \omega_2] dX \end{aligned}$$

$$\begin{aligned} R_1 &= \text{ArgMin} \left[\int_{R_1} [(C_{12} - C_{22}) P[\omega_2] P[X | \omega_2] - (C_{21} - C_{11}) P[\omega_1] P[X | \omega_1]] dX \right] \\ &= \text{ArgMin} \left[\int_{R_1} g[X] dX \right] \end{aligned}$$

$$g[X] < 0$$

$$(C_{21} - C_{11}) P[\omega_1] P[X | \omega_1] \stackrel{\omega_1}{>} (C_{12} - C_{22}) P[\omega_2] P[X | \omega_2]$$

$$\frac{P[X | \omega_1]}{P[X | \omega_2]} \stackrel{\omega_1}{>} \frac{(C_{12} - C_{22}) P[\omega_2]}{(C_{21} - C_{11}) P[\omega_1]}$$

■ LRT 결정규칙의 변형

■ Bayes Criterion (베이즈 기준)

정의 : Bayes Risk 를 최소화하는 LRT 결정규칙을 Bayes Criterion (베이즈 기준)

$$\Delta[X] = \frac{P[X | \omega_1]}{P[X | \omega_2]} \begin{matrix} > \\ < \end{matrix} \begin{matrix} \omega_1 \\ \omega_2 \end{matrix} \frac{(C_{12} - C_{22}) P[\omega_2]}{(C_{21} - C_{11}) P[\omega_1]}$$

■ MAP (Maximum A Posterior) 기준

대칭적 혹은 비용 값이 0 아니면 1 인 제로 -

원 비용 함수를 사용하면 베이즈 기준이 사후 확률 ($P[\omega_i | X]$) 의 비로 표현

이 기준을 사후 확률 최대화한다.

Maximum A Posterior (MAP) Criterion

$$C_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \Rightarrow \Delta[X] = \frac{P[X | \omega_1]}{P[X | \omega_2]} \begin{matrix} > \\ < \end{matrix} \begin{matrix} \omega_1 \\ \omega_2 \end{matrix} \frac{P[\omega_2]}{P[\omega_1]} \Leftrightarrow \frac{P[X | \omega_1]}{P[X | \omega_2]} \begin{matrix} > \\ < \end{matrix} \begin{matrix} \omega_1 \\ \omega_2 \end{matrix} 1$$

■ ML (Maximum Likelihood) 기준

사전 확률 ($P[\omega_1]$) 이 같고 제로 - 원 비용 함수의 경우, 베이즈 기준은 우도 ($P[X | \omega_i]$) 의 비로 표현

$$C_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases} \Rightarrow \Delta[X] = \frac{P[X | \omega_1]}{P[X | \omega_2]} \begin{matrix} > \\ < \end{matrix} \begin{matrix} \omega_1 \\ \omega_2 \end{matrix} 1$$

$$P[\omega_i] = \frac{1}{c} \forall i$$

■ 다중 클래스 결정규칙

■ 오류확률 ($P[\text{error}]$) 을 이용한 다중 클래스 결정규칙 ($P[\text{correct}]$)

$$P[\text{error}] = 1 - P[\text{correct}]$$

$$P[\text{correct}] = \sum_{i=1}^c P[\omega_i] \int_{R_i} P[X | \omega_i] dX$$

$$P[\text{correct}] = \sum_{i=1}^c P[\omega_i] \int_{R_i} P[X | \omega_i] dX = \sum_{i=1}^c \int_{R_i} P[X | \omega_i] P[\omega_i] dX = \sum_{i=1}^c \int_{R_i} P[\omega_i | X] P[X] dX$$

$$\int_{R_i} P[\omega_i | X] P[X] dX : \text{최대화}$$

$P[\omega_i | X]$ 를 최대로 하는 클래스 ω_i 를 선택

$P[\omega_i | X]$ 가 최대가 되는 영역을 R_i 로 선택

■ 베이지 위험을 이용한 다중 클래스 결정규칙

클래스 ω_i 를 선택하는 결정을 α_i 라고 정의

특징 x 를 클래스 ω_i 로 매핑하는 전체 결정규칙을 $\alpha[x]$ 라고 정의한다.

$$\alpha[x] \rightarrow \{\alpha_1, \alpha_2, \dots, \alpha_c\}$$

특징 x 를 클래스 ω_i 로 할당하는 베이지 위험 $R[\alpha_i | x]$

$$R[\alpha[x] \rightarrow \alpha_i] = R[\alpha_i | x] = \sum_{j=1}^c C_{ij} P[\omega_j | x]$$

$$R[\alpha[x]] = \int R[\alpha[x] | x] P[x] dx$$

■ Discriminant Function (판별함수) $g_i[x]$

결정규칙

만약, $\forall j \neq i, g_i[x] > g_j[x]$ 이라면, 특징벡터 x 를 클래스 ω_i 에 속한다고 결정

c 개의 클래스 중의 하나로 결정하는 시스템은 c 개의 판별함수로 구성된 네트워크로 표현하여 가장 큰 값을 출력하는 카테고리를 선택하는 구조를 가짐

기준	판별함수
Bayes	$g_i[x] = -R[\alpha_i x]$
MAP	$g_i[x] = P[\omega_i x]$
ML	$g_i[x] = P[x \omega_i]$

■ 최우추정법에 의한 확률밀도함수의 추정

$$P[\omega_j | x] = \frac{P[x | \omega_j] P[\omega_j]}{\sum_{k=1}^N P[x | \omega_k] P[\omega_k]} = \frac{P[x | \omega_j] P[\omega_j]}{P[x]}$$

a Posterior Probability (사후 확률)

a Prior Probability (사전 확률)

Likelihood (우도)

Parametric (모수적법) - MLE (최우추정)

Nonparametric (비모수적법) - 히스토그램, k - NNR, KDE (커널밀도추정)

■ Maximum Likelihood Estimation (최우추정 : 최대 우도 추정)

$$P[x | \theta], x = \{x_1, \dots, x_n\}, \theta = \{\theta_1, \dots, \theta_m\}$$

$$P[x | \theta] = \prod_{k=1}^n P[x_k | \theta]$$

$$\hat{\theta}$$

■ log-Likelihood (로그-우도 함수)

$$l[\theta] = \log P[x | \theta] = \sum_{i=1}^n \log P[x_i | \theta]$$

$$\frac{\partial}{\partial \theta} \log P[x | \theta] = - \sum_{i=1}^n \frac{\partial}{\partial \theta} \log P[x_i | \theta] = 0$$

■ ∇ 그래디언트 연산자

$$\nabla_{\theta} l = \left(\frac{\partial l}{\partial \theta_1}, \dots, \frac{\partial l}{\partial \theta_M} \right)^T \text{ (최우방정식) } (\hat{\theta} : \text{로그-우도를 최대화 하는 파라미터})$$

$$x = \{x_1, \dots, x_R\}$$

$$x \sim N[\mu, \sigma^2]$$

$$\theta = \{\theta_1, \theta_2\} = \{\mu, \sigma\}$$

$$l[\theta] = \log \left(\prod_{n=1}^R \frac{1}{\sqrt{2\pi} \theta_2} e^{-\frac{1}{2\theta_2^2} (x_n - \theta_1)^2} \right) = \sum_{n=1}^R \left(-\frac{1}{2} \log(2\pi \theta_2^2) - \frac{1}{2\theta_2^2} (x_n - \theta_1)^2 \right)$$

$$\nabla_{\theta} l[\theta] = \nabla_{\theta} \log P[x | \theta] = \sum_{n=1}^R \left(\begin{array}{c} \frac{1}{\theta_2^2} (x_n - \theta_1) \\ -\frac{1}{\theta_2} + \frac{(x_n - \theta_1)^2}{\theta_2^3} \end{array} \right)$$

$$\sum_{n=1}^R \frac{1}{\theta_2^2} (x_n - \theta_1) = 0$$

$$\theta_{1, ML} = \frac{1}{R} \sum_{n=1}^R x_n = \hat{\mu} \text{ (표본평균)}$$

$$-\sum_{n=1}^R \frac{1}{\theta_2} + \sum_{n=1}^R \frac{1}{\theta_2} + \frac{(x_n - \theta_1)^2}{\theta_2^3} = 0$$

$$\theta_{2, ML} = \frac{1}{R} \sum_{n=1}^R (x_n - \hat{\mu})^2 = \hat{\sigma} \text{ (표본분산)}$$

■ 표본평균

$$E[\hat{\mu}] = E\left[\frac{1}{R} \sum_{i=1}^R x_i\right] = \frac{1}{R} E\left[\sum_{i=1}^R x_i\right] = \frac{1}{R} \sum_{i=1}^R \mu = \mu$$

→ unbiased (언바이어스 추정)

■ 표본분산

$$E[\hat{\sigma}^2] = E\left[\frac{1}{R} \sum_{i=1}^R (x_i - \hat{\mu})^2\right] = E\left[\frac{1}{R} \left(\sum_{i=1}^R x_i - \frac{1}{R} \sum_{j=1}^R x_j\right)^2\right] = \frac{R-1}{R} \sigma^2$$

→ bias (바이어스 추정)

$$\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_R\}$$

$$\mathbf{X} \sim \mathbf{N}[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$$

i.i.d 조건

$$\boldsymbol{\mu}_{\text{ML}} = \frac{1}{R} \sum_{n=1}^R \mathbf{x}_n = \hat{\boldsymbol{\mu}}$$

$$\boldsymbol{\Sigma}_{\text{ML}} = \frac{1}{R} \sum_{n=1}^R (\mathbf{x}_n - \hat{\boldsymbol{\mu}}) (\mathbf{x}_n - \hat{\boldsymbol{\mu}})^T = \hat{\boldsymbol{\Sigma}}$$

$$\mathbf{E}[\hat{\boldsymbol{\mu}}] = \boldsymbol{\mu}, \quad \mathbf{E}[\hat{\boldsymbol{\Sigma}}] = \frac{R-1}{R} \boldsymbol{\Sigma}$$