DPR 22.4.2022

2. Bayesian Decision theory

decision: model based detection, learning based classification

Bayesian theory

Bayesian decision theory for classification

ch 2

Bayesian estimation theory. for regression

2.1 Bayesian theorem

slide 2.1

need $P(\omega = \omega_i)$, $P(\underline{x} \mid \omega = \omega_i)$, $P(\omega = \omega_i \mid \underline{x})$

apriori likelihood posteriori

we assume $P(\omega_i)$ and likelihood $p(\underline{x} \mid \omega_i)$

Slide 2.3

$$P(\omega_j \mid \underline{x}) = \frac{P(\underline{x}, \omega_j)}{P(\underline{x})} = \frac{P(\underline{x} \mid \omega_j) P(\omega_j)}{P(\underline{x})}$$

$$= \frac{P(\underline{x} \mid \omega_j) P(\omega_j)}{\sum_{k=1}^{c} P(\underline{x} \mid \omega_k) P(\omega_k)}$$

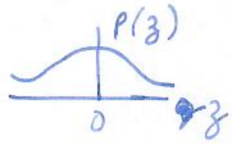
$$\text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}$$

E 2.1 Bit detection

$$X = S + z$$

$$S \in \{-1, 1\} \hat{=} \omega$$

$$z \sim N(0, \sigma^2), \quad p(z) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{z^2}{2\sigma^2}\right)$$



Prior: $P(S=1) = P(S=-\frac{1}{2}) = \frac{1}{2}$ or not.

Likelihood:

$$\left.\begin{aligned} p(x \mid S=-1) &\sim N(-1, \sigma^2) \\ p(x \mid S=+1) &\sim N(1, \sigma^2) \end{aligned}\right\} \text{shifting of mean by } S$$

Evidence:

$$p(x) = p(S=1) \cdot P(x \mid S=1) + P(S=-1) P(x \mid S=-1)$$

$$= \frac{1}{2}\left[N(-1, \sigma^2) + N(1, \sigma^2)\right]$$

2.2 Minimum Bayesian risk decision

2.2.1 Error and loss

Detailed performance criterion:

Confusion matrix $[n_{ij}]$, $C \times C$ matrix

→ joint probability matrix $[P_{ij}]$, $C \times C$

$$P_{ij} = P(\hat{\omega} = \omega_i, \omega = \omega_j) = \frac{n_{ij}}{N}$$

Adv: a complete picture about all combinations of $(\hat{\omega}, \omega)$

Disadv: $C^2 \underbrace{-1}_{\text{sum of } C^2 \text{ element} = 1}$ individual measures, not practical → no general sol^{n}

→ need to condense to single term.

Wish 1: A scalar performance measure

$$1 = \sum_i \sum_j P_{ij}$$

$$= \sum P_{ii} + \sum_{i \neq j}\sum P_{ij}$$

$$= P(\hat{\omega} = \omega) + P(\hat{\omega} \neq \omega)$$

[correct decision] + [wrong decision]

recognition rate/ accurracy

error rate (ER)

$0 \leq ER \leq 1$: average error rate over all $c^2 - c$ elements (errors)

Wish 2 : We need to do a subjective ~~weigting~~ weighting of different errors with loss

loss (cost, price, risk):

$l_{ij} = l(\hat{\omega} = \omega_i, \hat{\omega} = \omega_j) \geq 0, 1 \leq i, j \leq c$

loss matrix $[l_{ij}]$ $c \times c$

To be subjective based on application

Intuition

* loss - bad, to be minimized
* $l_{ii} = 0$: no loss for correct decision
* $l_{ij} > 0 \ \forall \ i, j$; $l_{ij} \uparrow$ for "big" errors
  $l_{ij} \downarrow$ for "small" errors

Bayesian risk (loss) : BR: average loss for $\forall(\hat{\omega}, \omega)$

$$BR = \sum_i \sum_j l_{ij} P_{ij}$$

$$= \underset{\substack{\uparrow \\ \text{Expectation over } (\hat{\omega}, \omega)}}{E_{\hat{\omega}, \omega}} \; l(\hat{\omega}, \omega) \geq 0$$

One important loss:

0/1 loss $\quad l_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$; all errors are equally bad. (Digital communication)

$$BR = \sum\sum_{i \neq j} P_{ij} = \text{Error Rate } (ER)$$

ER is a special case of BR.

BR defined over $(\hat{\omega}, \omega)$

$$\downarrow \hat{\omega} = \hat{\omega}(\underline{x})$$

BR defined over $(\underline{x}, \underline{\omega})$: average loss over all $(\underline{x}, \underline{\omega})$

$$BR = E_{\underline{x}, \omega} \, l(\hat{\omega}(\underline{x}), \omega)$$

$$= \int_{\mathbb{R}^d} \sum_{j=1}^{c} l(\hat{\omega}(\underline{x}), \omega = \omega_j) \, p(\underline{x}, \omega_j) \, d\underline{x}$$

$$= \int_{\mathbb{R}^d} \underbrace{\sum_{j=1}^{c} l(\hat{\omega}(\underline{x}), \omega = \omega_j) \, P(\omega_j | \underline{x})}_{R(\hat{\omega} | \underline{x}) \,:\, \text{Conditional risk}} \, p(\underline{x}) \, d\underline{x}$$

$$= \int_{\mathbb{R}^d} R(\hat{\omega} | \underline{x}) \, p(\underline{x}) \, dx$$

i.e. $\min_{\hat{\omega}} BR \mathrel{\hat{=}} \min_{\hat{\omega}} R(\hat{\omega} | \underline{x})$

$$= \sum_{j=1}^{c} l(\hat{\omega}(\underline{x}), \omega = \omega_j)\, P(\omega_j | \underline{x})$$

$$= E_\omega \left(l(\hat{\omega}(\underline{x}), \omega) \,\middle|\, \underline{X} \right) \quad \forall\ \underline{x}$$

$\hat{=}$ average loss for all ω, but for a fixed $\underline{x}$

$R(\hat{\omega} | \underline{x})$ depends on

*) $l(\hat{\omega}, \omega)$ - chosen by user

*) $P(\omega_j | \underline{x})$ - Bayesian theorem – can be calculated

*) $\hat{\omega}(\underline{x})$ – unknown - to be determined

2.2.2 Minimum Bayesian Risk (MBR)

1 idea to design $\hat{\omega}(\underline{x})$

MBR decision/detection/classifier

$$\hat{\omega}_{MBR}(\underline{x}) \approx \arg\min_{\hat{\omega}} R(\hat{\omega} | \underline{x})$$

$$= \arg\min_{\omega_i} R(\hat{\omega} = \omega_i | \underline{x})$$

i.e. for any given $\underline{x}$:

*) Calculate c values $R(\hat{\omega} = \omega_i | \underline{x}) = \sum_{j=1}^{c} l_{ij} P(\omega_j | \underline{x})$ $\quad 1 \le i \le c$

*) look for the minimum.

*) index of min is the class

$$\boxed{l_{ij}} \begin{bmatrix} P(\omega_1 | \underline{x}) \\ \vdots \\ P(\omega_c | \underline{x}) \end{bmatrix} = \begin{bmatrix} R(\hat{\omega} = \omega_1 | \underline{x}) \\ \vdots \\ R(\hat{\omega} = \omega_c | \underline{x}) \end{bmatrix}$$

$C \times C$ $\quad$ $C \times 1$ $\quad$ $C \times 1$

known as chosen

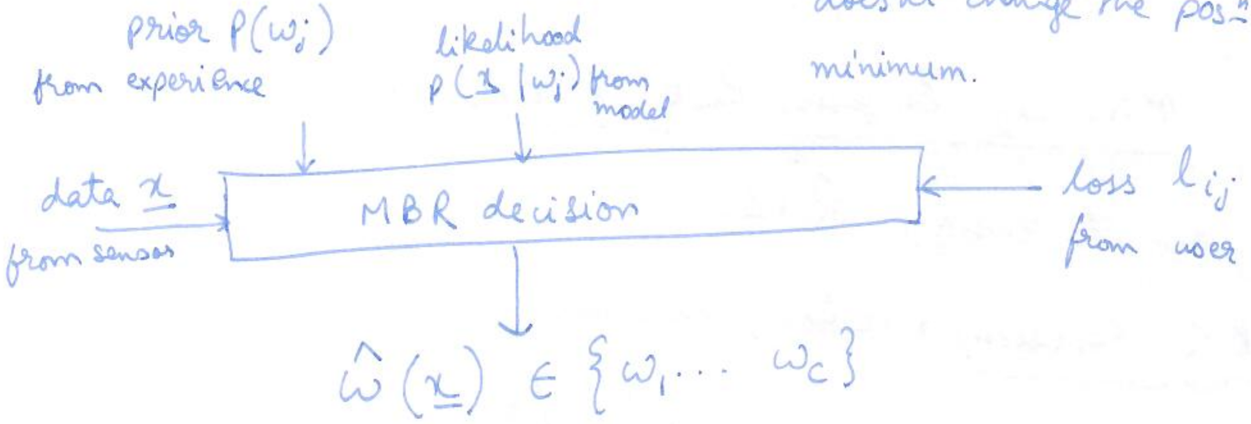
check the min ~~here~~ here

row index of min

Bayesian theorem

$$P(\omega_j | \underline{x}) = \frac{P(\underline{x} | \omega_j)\, P(\omega_j)}{P(\underline{x})}$$

this is independent of ω_j so we can ignore it as it doesn't change the posⁿ of the minimum.



E2.3 Two class MBR

$c = 2$

$$R(\hat{\omega} = \omega_1 | \underline{x}) \underset{\hat{\omega}=\omega_1}{\overset{\hat{\omega}=\omega_2}{\gtrless}} R(\hat{\omega} = \omega_2 | \underline{x})$$

$$= l_{21} P(\omega_1 | \underline{x}) + l_{22} P(\omega_2 | \underline{x})$$

$$= l_{11} P(\omega_1 | \underline{x}) + l_{12} P(\omega_2 | \underline{x})$$

$$= \underbrace{(l_{12} - l_{22})}_{>0} \underbrace{P(\omega_2 | \underline{x})}_{= \frac{P(\underline{x}|\omega_2) P(\omega_2)}{\cancel{P(\underline{x})}}} \underset{\hat{\omega}=\omega_1}{\overset{\hat{\omega}=\omega_2}{\gtrless}} \underbrace{(l_{21} - l_{11})}_{>0} \underbrace{P(\omega_1 | \underline{x})}_{= \frac{P(\underline{x}|\omega_1) P(\omega_1)}{\cancel{P(\underline{x})}}}$$

$\Rightarrow$ likelihood ratio test (LRT)

LR

$$LR = \underbrace{\frac{P(\underline{x} | \omega_2)}{P(\underline{x} | \omega_1)}}_{\text{likelihood ratio}} \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \frac{l_{21} - l_{11}}{l_{12} - l_{22}} \frac{P(\omega_1)}{P(\omega_2)} = \gamma$$

VVI