2.2.3 Minimum error rate and maximum a posteriori

One special case of MBR, when using

0/1 loss : $l_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$

$$BR = \sum_i \sum_j l_{ij} P_{ij} \longrightarrow ER = \sum\sum_{i \neq j} P_{ij}$$

$$\min BR \longrightarrow \min ER$$

2.3 (cont) Two class decision

$$LR = \frac{P(\underline{x} | \omega_2)}{P(\underline{x} | \omega_1)} \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \gamma = \frac{P(\omega_1)}{P(\omega_2)}$$

Another name for min Error rate:

$$R(\hat{\omega} = \omega_i | \underline{x}) = \sum_{j=1}^{c} l_{ij} P(\omega_j / \underline{x})$$

for 0/1 loss

$$\underbrace{\sum_{\substack{j=1 \\ j \neq i}}^{c} P(\omega_j / \underline{x}) + P(\omega_i | \underline{x})}_{=1} - P(\omega_i | \underline{x})$$

$$= 1 - P(\omega_i / \underline{x})$$

Hence, when we $\min_{\omega_i} ER \triangleq \underbrace{\max_{\omega_i} P(\omega_i / \underline{x})}_{\text{max aposteriori (MAP)}}$ due to

0/1 error loss max aposteriori (MAP) $\hat{\omega}_{MAP}(\underline{x})$

2.2.4 Maximum likelihood

One special case of MAP, when assuming equal priors.

$$P(\omega_i) = \frac{1}{C} \quad \forall i$$

$$\Rightarrow P(\omega_i \mid \underline{x}) = P(\underline{x} \mid \omega_i) \cdot \frac{P(\omega_i)}{P(\underline{x})}$$

when $P(\omega_i) = \frac{1}{C}$ then $\frac{P(\omega_i)}{P(\underline{x})}$ is independent of ω_c

and is a scaling factor.

$$\Rightarrow P(\omega_i \mid \underline{x}) \sim P(\underline{x} \mid \omega_i)$$

$$\therefore \max_{\omega_i} P(\omega_i \mid \underline{x}) \triangleq \underbrace{\max_{\omega_i} P(\underline{x} \mid \omega_i)}_{\text{maximum likelihood ML}}$$

$$\hat{\omega}_{ML}(\underline{x})$$

E 2.3 cont: Two class decision.

$$LR = \frac{P(\underline{x} \mid \omega_2)}{P(\underline{x} \mid \omega_1)} \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \gamma = 1$$

By def^n ML decision minimizes the balanced error rate. (BER)

$$ER = \sum_{i \neq j}\sum P_{ij} = \sum_{i \neq j}\sum P(\hat{\omega} = \omega_i, \omega = \omega_j) \quad \boxed{P_{ij}}$$

Sum of all off-diagonal elements

BER is a special case of ER

$$BER = ER\Big|_{P(\omega_j) = \frac{1}{C}} = \sum_{i \neq j}\sum P(\hat{\omega} = \omega_i \mid \omega = \omega_j) \cdot \underbrace{P(\omega = \omega_j)}_{=\frac{1}{C}}$$

$$= \frac{1}{C} \sum_{i \neq j}\sum P(\hat{\omega} = \omega_i \mid \omega = \omega_j) = \frac{1}{C} \sum_{i \neq j}\sum \tilde{P}_{ij}$$

If $c = 2$:

$ER = P_{12} + P_{21}$

$BER = \frac{1}{2} [\bar{P}_{12} + \bar{P}_{21}]$

Slide 2-7

BER → performance of classifier

ER → " " classification

2.2.5 Discriminant functions

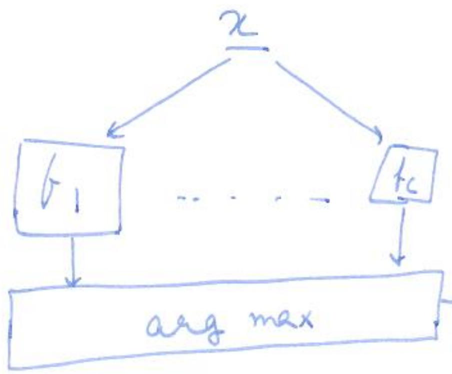
MBR / MAP / ML decision:

min/max function $(\hat{\omega}(x) = \omega_i)$ $1 \leq i \leq c$

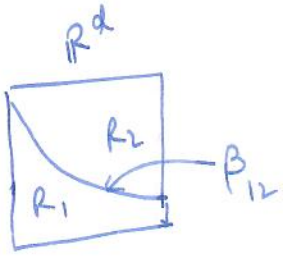
→ a unified framework:

a set of c discriminant functions $f_i(\underline{x})$ $1 \leq i \leq c$

$\hat{\omega}(\underline{x}) = \arg\max_i f_i(\underline{x})$



$\hat{\omega}(\underline{x}) \in \{\omega_1 \dots \omega_c\}$



decision region:

$$R_i = \{ \underline{x} \mid f_i(\underline{x}) > f_j(\underline{x}) \ \forall j \neq i \}$$

decision boundary $\beta_{ij} = \{ \underline{x} \mid f_i(\underline{x}) = f_j(\underline{x}) > f_k(\underline{x}) \ \forall k \neq i,j \}$

If $\phi()$ is a monotonically increasing f^n eg $\ln()$, then $\phi(f_i(x_i))$ is an equivalent set of discriminant f^n.

$$\text{MBR} : f_i(\underline{x}) = -R(\hat{\omega} = \omega_i \mid \underline{x}) = -\sum_{j=1}^{C} l_{ij} P(\omega_j \mid \underline{x})$$

$$\text{MAP} : f_i(\underline{x}) = P(\omega_i \mid \underline{x}) \sim p(\underline{x} \mid \omega_i) P(\omega_i) \text{ or}$$

$$\ln p(\underline{x} \mid \omega_i) + \ln P(\omega_i)$$

$$\text{ML} : f_i(x) = p(\underline{x} \mid \omega_i) \text{ or } \ln p(\underline{x} \mid \omega_i)$$

2.2.6 MAP decision for Gaussian likelihood

Slide 2-10

Assumptions:

*) $C \geq 2$

*) $\underline{x} \in \mathbb{R}^d$

*) MAP: 0/1 loss

*) Known priors $P(\omega_i) = p_i$

*) Known Gaussian likelihood

$$\underline{X} \mid \omega_i \sim N(\underline{\mu}_i, \underline{\underline{C}}_i)$$

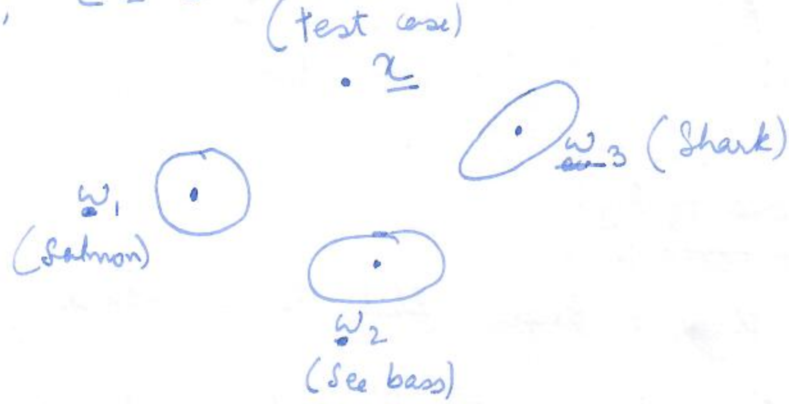
$$p(\underline{x} \mid \omega_i) = \frac{1}{(2\pi)^{d/2}\sqrt{\underline{\underline{C}}_i}}\, e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)^T \underline{\underline{C}}_i^{-1} (\underline{x}-\underline{\mu}_i)}$$

$\underline{\mu}_i$ = mean / center

$\underline{\underline{C}}_i$ = covariance matrix / contour lines

$\Big\}$ of class ω_i

$d = 2, \quad c = 3$



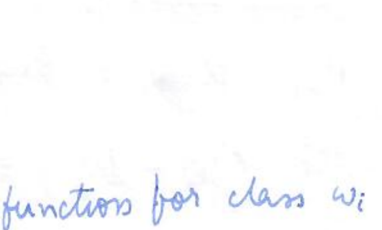
MAP decision rule / MAP discriminant functions for class ω_i :

$$f_i(\underline{x}) = \ln P(\underline{x} \mid \omega_i) + \ln P(\omega_i)$$

$$= \cancel{-\frac{d}{2}\ln 2\pi} - \frac{1}{2}\ln |\underline{\underline{C}}_i| - \frac{1}{2}(\underline{x}-\underline{\mu}_i)^T \underline{\underline{C}}_i^{-1}(\underline{x}-\underline{\mu}_i)$$

(constant, not needed) $\quad + \ln P_i$

Case A: $\underline{\underline{C}}_i = \sigma^2 \underline{\underline{I}} \quad \forall i$



$$f_i(\underline{x}) = \cancel{-\frac{1}{2\sigma^2}}^{\text{(ignored)}} - \frac{1}{2\sigma^2}\|\underline{x}-\underline{\mu}_i\|^2 + \ln P_i$$

$$= -\frac{1}{2\sigma^2}\left(\overset{\text{(ignored)}}{\cancel{\|\underline{x}\|^2}} - 2\underline{\mu}_i^T \underline{x} + \|\mu_i\|^2\right) + \ln P_i$$

$$= -\frac{1}{2\sigma^2}\left(-2\mu_i^T \underline{x} + \|\mu_i\|^2\right) + \ln P_i$$

$$= \frac{1}{\sigma^2}\mu_i^T \underline{x} + \underbrace{\ln P_i - \frac{1}{2\sigma^2}\|\mu_i\|^2} = \underline{w}_i^T \underline{x} + \underline{w}_{i0}$$

Note:

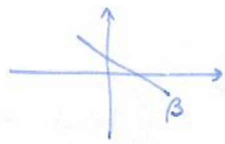
$f(\underline{x})$ is a linear f^n of $\underline{x}$ if $f(\underline{x}) = \underline{\omega}^T \underline{x}$

" " an affine " " " " $f(x) = \underline{\omega}^T \underline{x} + \omega_0$

Decision boundary β_{ij} between $f_i(\underline{x})$ and $f_j(\underline{x})$:

$$f_i(\underline{x}) - f_j(\underline{x}) = (\underline{\omega_i} - \underline{\omega_j})^T \underline{x} + (\omega_{i0} - \omega_{j0}) \overset{!}{=} 0$$

$d = 1$ | $d = 2$ | $d = 3$ | $d \geq 4$



plane

hyper plane

Properties of β_{ij}:

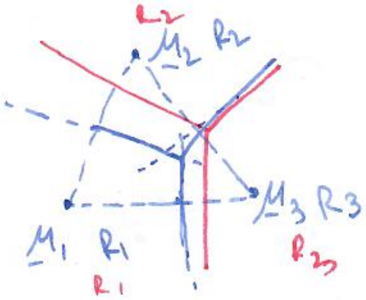
P1. It is a linear (flat) decision boundary

P2. $(\underline{\mu_i} - \underline{\mu_j}) = \sigma^2 (\underline{\omega_i} - \underline{\omega_j}) \perp \beta_{ij}$

P3. If $P_i = P_j$: $\max_i f_i(\underline{x}) \hat{=} \min \|\underline{x} - \underline{\mu_i}\|$ ← due to the -ve sign in front.

" nearest mean solution "

If $P_i > P_j$: large R_i, β_{ij} is closer to $\underline{\mu_j}$



(a) $P_1 = P_2 = P_3 = \frac{1}{3}$

(b) $P_1 = \frac{1}{2}$, $P_2 = P_3 = \frac{1}{4}$

(b) β_1 moves closer to μ_2, μ_3 parallel to the previous one. The other 2 boundaries remain the same.

Slide 2-12.