$\underline{\Sigma_i = \Sigma}$:

$$g_i(\underline{x}) = -\tfrac{1}{2}(\underline{x}-\underline{\mu}_i)^T \Sigma^{-1}(\underline{x}-\underline{\mu}_i) + \ln P_Y(i)$$

Quantity $(\underline{x}-\underline{\mu})^T \Sigma^{-1}(\underline{x}-\underline{\mu})$ called <u>Mahalanobis Distance</u>

If $P_Y(i) = 1/M$, then $\hat{\alpha}(\underline{x}) = \arg\min_i (\underline{x}-\underline{\mu}_i)^T \Sigma^{-1}(\underline{x}-\underline{\mu}_i)$

<u>minimum dist.</u> using M. distance

Equivalently

$$g_i(\underline{x}) = \underline{\mu}_i^T \Sigma^{-1} \underline{x} - \tfrac{1}{2} \underline{\mu}_i^T \Sigma^{-1} \underline{\mu}_i + \ln P_Y(i)$$

<u>linear</u> <u>discriminant</u>

<u>General Σ_i</u> (Decision region boundaries still hyperplanes but not $\perp$ to line bet. means)

$$g_i(\underline{x}) = -\tfrac{1}{2}(\underline{x}-\underline{\mu}_i)^T \Sigma_i^{-1}(\underline{x}-\underline{\mu}_i) - \tfrac{1}{2}\ln(\det \Sigma_i) + \ln P_Y(i)$$

Equivalently,

$$g_i(\underline{x}) = \underline{\mu}_i^T \Sigma_i^{-1} \underline{x} - \tfrac{1}{2} \underline{\mu}_i^T \Sigma_i^{-1} \underline{\mu}_i - \tfrac{1}{2} \underline{x}^T \Sigma_i^{-1} \underline{x} - \tfrac{1}{2}\ln(\det \Sigma_i) + \ln P_Y(i)$$

<u>quadratic</u> <u>discriminant fns.</u>

Decision regions

surfaces can be hyperplanes, hyperspheres, hyperellipsoids, hyperparaboloids, etc.

Decision regions may not be simply connected.

$Y = i$ w.p. $P_Y(i)$ $i = 1, \dots, m$

L_{ij} = loss in deciding i when $Y = j$

$= L(\alpha(\underline{x}) = i, Y = j)$

<u>Recall</u>: Bayes decision

$$\hat{\alpha}(\underline{x}) = \underset{i=1,\dots,m}{\operatorname{argmin}} \sum_{j=1}^{m} L_{ij} \, P_{Y|\underline{X}}(j|\underline{x})$$

$$= \underset{i=1,\dots,m}{\operatorname{argmin}} \sum_{j=1}^{m} L_{ij} \, P_Y(j) \, f_{\underline{X}|Y}(\underline{x}|j) \qquad (1)$$

Consider likelihood ratio

$$\Lambda_i(\underline{x}) = \frac{f_{\underline{X}|Y}(\underline{x}|i)}{f_{\underline{X}|Y}(\underline{x}|1)}, \quad i = 1, \dots, m$$

Thus, $\Lambda_1(\underline{x}) = 1$. Dividing by $f_{\underline{X}|Y}(\underline{x}|1)$ in (1):

$$\hat{\alpha}(\underline{x}) = \underset{i=1,\dots,m}{\operatorname{argmin}} \left[L_{i1} P_Y(1) + \sum_{j=2}^{m} L_{ij} P_Y(j) \Lambda_j(\underline{x}) \right]$$

Decision based on $m-1$-dim vector $\begin{bmatrix} \Lambda_2(\underline{x}) \\ \vdots \\ \Lambda_m(\underline{x}) \end{bmatrix}$

$R_i = \{ \underline{x} : \hat{\alpha}(\underline{x}) = i \}$ = set of pts satisfying

$$L_{i1} P_Y(1) + \sum_{j=2}^{m} L_{ij} P_Y(j) \Lambda_j(\underline{x}) \le L_{k1} P_Y(1) + \sum_{j=2}^{m} L_{kj} P_Y(j) \Lambda_j(\underline{x})$$

$\forall k \ne i$

or $(L_{i1} - L_{k1}) P_Y(1) + \sum_{j=2}^{m} (L_{ij} - L_{kj}) P_Y(j) \Lambda_j(\underline{x}) \leq 0$
$\forall k \neq i$

In the space of likelihood ratios, inequality corresp. to halfspace defined by the affine space

$$(L_{i1} - L_{k1}) P_Y(1) + \sum_{j=2}^{m} (L_{ij} - L_{kj}) P_Y(i) \Lambda_j(\underline{x}) = 0$$

This separates R_i and R_k.

R_i = convex region defined by $m-1$ affine halfspaces.

- If $L_{ii} = 0 \ \forall i$ and $L_{ij} = g_j \quad \forall i \neq j$

(i.e. cost of making error when $Y=j$ indep of how error was made). Then,

$$\sum_{j=1}^{m} L_{ij} P_Y(j) \Lambda_j(\underline{x}) = \sum_{j \neq i} g_j P_Y(j) \Lambda_j(\underline{x})$$

$$= \sum_{j=1}^{m} g_j P_Y(j) \Lambda_j(\underline{x}) - g_i P_Y(i) \Lambda_i(\underline{x})$$

Thus, $\hat{\alpha}(\underline{x}) = \underset{i}{\operatorname{argmin}} \sum_{j=1}^{m} L_{ij} P_Y(j) \Lambda_j(\underline{x})$

$$= \underset{i}{\operatorname{argmin}} \underbrace{\sum_{j=1}^{m} g_j P_Y(j) \Lambda_j(\underline{x})}_{\text{same for all } i} - g_i P_Y(i) \Lambda_i(\underline{x})$$

$$= \underset{i=1,..,m}{\operatorname{argmax}} \ g_i P_Y(i) \Lambda_i(\underline{x})$$

View as set of binary threshold comparisons,

$$g_i\, p_Y(i)\, \Lambda_i(\underline{x}) \underset{\hat{\alpha} \neq i}{\overset{\hat{\alpha} \neq j}{\gtrless}} g_j\, p_Y(j)\, \Lambda_j(\underline{x})$$

$$\frac{\Lambda_j(\underline{x})}{\Lambda_i(\underline{x})} = \frac{f_{\underline{X}|Y}(\underline{x}|j)}{f_{\underline{X}|Y}(\underline{x}|i)} \underset{\hat{\alpha} \neq j}{\overset{\hat{\alpha} \neq i}{\gtrless}} \frac{g_i\, p_Y(i)}{g_j\, p_Y(j)} \equiv \eta_{ji}$$

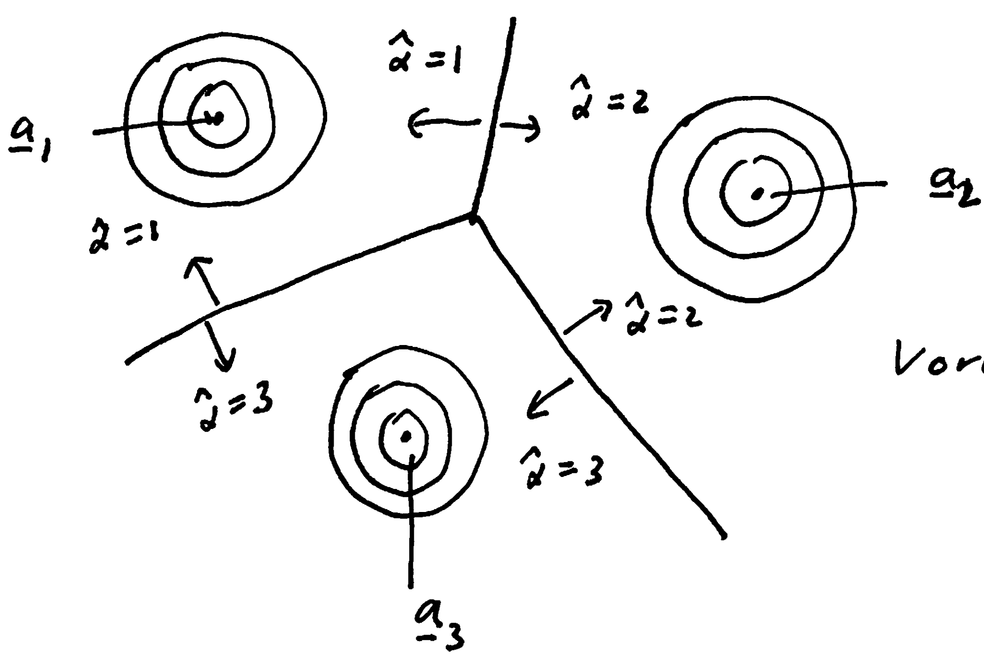
$$\frac{\Lambda_j(\underline{x})}{\Lambda_i(\underline{x})} \geq \eta_{ji} \quad \forall\, i \neq j \Rightarrow j \text{ is Bayes decision}$$

<u>Ex</u>: $\quad Y = i: \quad \underline{X} = \underline{a}_i + \underline{Z} \qquad \underline{Z} \sim \mathcal{N}(\underline{0}, \sigma^2 I)$

i.e. $\quad \underline{X} \sim \mathcal{N}(\underline{a}_i, \sigma^2 I) \qquad \underline{a}_i = \begin{bmatrix} a_{i1} \\ \vdots \\ a_{in} \end{bmatrix}$

From p. 36, each binary threshold comparison has form:

$$(\underline{a}_j - \underline{a}_i)^T \underline{x} \underset{\hat{\alpha} \neq j}{\overset{\hat{\alpha} \neq i}{\gtrless}} \sigma^2 \ln \eta_{ji} + \frac{\|\underline{a}_j\|^2 - \|\underline{a}_i\|^2}{2}$$



$n = 2$

Voronoi regions

- Decision threshold bet. each pair of categories is affine space $\perp$ to line joining 2 signals
- $(\underline{a}_j - \underline{a}_1)^T \underline{x}$, $2 \le j \le m$ is a suff. stat. for m-ary problem since $(\underline{a}_j - \underline{a}_i)^T \underline{x} = (\underline{a}_j - \underline{a}_1)^T \underline{x} - (\underline{a}_i - \underline{a}_1)^T \underline{x}$
- If $n > m-1$, reduce to $m-1$ dimensions by expressing $\underline{a}_i - \underline{a}_1$ in terms of $\le m-1$ basis vectors.
- Calculating $Pr(e)$ in Gaussian m-ary decision problem is hard, even when calculating in $m-1$ dimensional space. Must integrate over voronoi regions
- often good enough to give good upper bound to $Pr(e)$.

Union bound:

$$P\left(\bigcup_{j=1}^{K} E_j\right) \le \sum_{j=1}^{K} P(E_j)$$ for any set of events $E_1, \ldots, E_K$.

Then,

$$Pr(e \mid Y=i) \le \sum_{j \ne i} P\left((\underline{a}_j - \underline{a}_i)^T \underline{x} \ge \sigma^2 \ln \eta_{ji} + \frac{\|\underline{a}_j\|^2 - \|\underline{a}_i\|^2}{2} \right)$$

$$= \sum_{j \neq i} \ell \left(\frac{\sigma \ln \eta_{ji}}{\| \underline{a}_j - \underline{a}_i \|} + \frac{\| \underline{a}_j - \underline{a}_i \|}{2\sigma} \right).$$