Binary Classification in Additive Gaussian Noise

$$Y = 1 \iff X \sim \mathcal{N}(a, \sigma^2)$$

$$Y = 2 \iff X \sim \mathcal{N}(b, \sigma^2)$$

$$f_{X|Y}(x|1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]$$

$$f_{X|Y}(x|2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-b)^2}{2\sigma^2}\right]$$

use 0-1 loss function:

$$\frac{f_{X|Y}(x|2)}{f_{X|Y}(x|1)} = \exp\left[\frac{(x-a)^2-(x-b)^2}{2\sigma^2}\right]$$

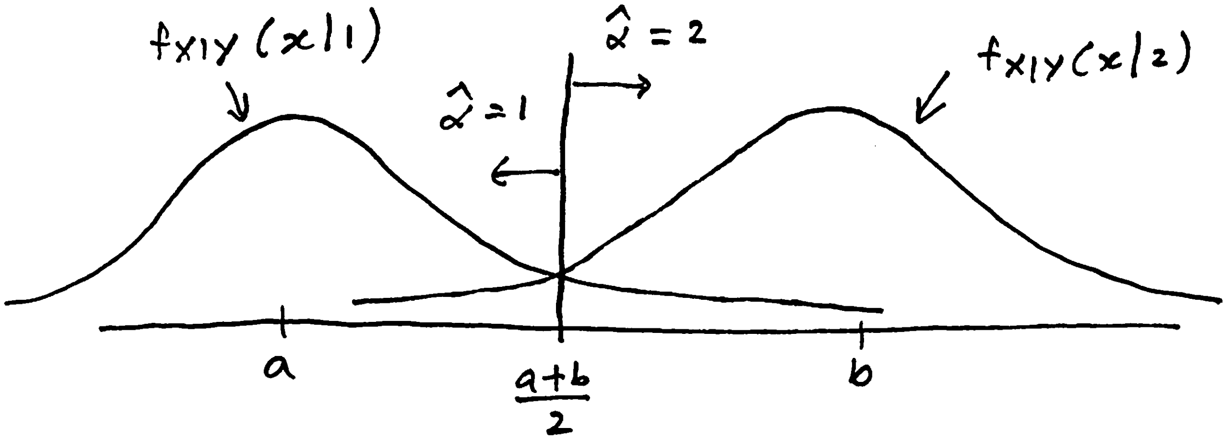
$$= \exp\left[\left(\frac{b-a}{\sigma^2}\right)\left(x-\frac{b+a}{2}\right)\right] \underset{\hat{\alpha}(x)=1}{\overset{\hat{\alpha}(x)=2}{\gtrless}} \frac{p_Y(1)}{p_Y(2)} = T_L$$

Take logarithm (natural):

$$\ln\left(\frac{f_{X|Y}(x|2)}{f_{X|Y}(x|1)}\right) = \left(\frac{b-a}{\sigma^2}\right)\left(x-\frac{b+a}{2}\right) \underset{\hat{\alpha}(x)=1}{\overset{\hat{\alpha}(x)=2}{\gtrless}} \ln\left(\frac{p_Y(1)}{p_Y(2)}\right)$$

$$\iff \quad x \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}} \frac{\sigma^2 \ln\left(\frac{p_Y(1)}{p_Y(2)}\right)}{b-a} + \frac{b+a}{2} \equiv \eta \quad \underline{MAP}$$

If $p_Y(1) = p_Y(2) = \frac{1}{2}$, $\quad x \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}} \frac{b+a}{2} \quad \underline{ML}$



From MAP rule,

$$\Pr(e \mid Y=1) = \Pr(X \geq \eta \mid Y=1)$$

Given $Y=1$, $\quad X \sim \mathcal{N}(a, \sigma^2)$, $\quad \frac{X-a}{\sigma} \sim \mathcal{N}(0,1)$

$$\Pr(X \geq \eta \mid Y=1) = \Pr\left(\frac{X-a}{\sigma} \geq \frac{\eta-a}{\sigma} \,\Big|\, Y=1\right)$$

$$= Q\left(\frac{\eta-a}{\sigma}\right)$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$ ← not in closed form, tabulated.

$$\Pr(e \mid Y=1) = Q\left(\frac{\sigma^2 \ln\left(\frac{P_Y(1)}{P_Y(2)}\right)}{b-a} + \frac{b-a}{2\sigma}\right)$$

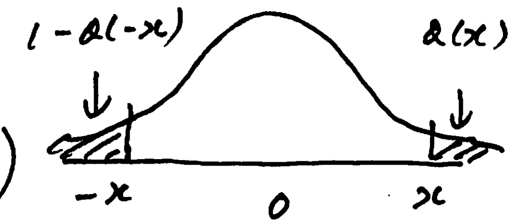
by using $\quad \eta = \frac{\sigma^2 \ln(P_Y(1)/P_Y(2))}{b-a} + \frac{b+a}{2}$

Similarly, given $Y=2$, $\quad X \sim \mathcal{N}(b, \sigma^2)$

$$\Pr(X < \eta \mid Y=2) = \Pr\left(\frac{X-b}{\sigma} < \frac{\eta-b}{\sigma} \,\Big|\, Y=2\right)$$

$$= 1 - Q\left(\frac{\eta-b}{\sigma}\right)$$

Now $Q(x) = 1 - Q(-x)$ for any x



$$\Rightarrow \quad Pr(e|Y=2) = Q\left(\frac{-\sigma \ln\left(\frac{P_Y(1)}{P_Y(2)}\right)}{b-a} + \frac{b-a}{2\sigma} \right)$$

$Pr(e|Y=i)$ are functions of $\frac{b-a}{\sigma}$ and $\frac{P_Y(1)}{P_Y(2)}$

Let $T_L = P_Y(1)/P_Y(2)$

$$\gamma = \frac{b-a}{\sigma}$$

$$Pr(e|Y=1) = Q\left(\frac{\ln T_L}{\gamma} + \frac{\gamma}{2}\right)$$

$$Pr(e|Y=2) = Q\left(-\frac{\ln T_L}{\gamma} + \frac{\gamma}{2}\right)$$

$P_Y(1) = P_Y(2) = 1/2$: $\quad T_L = 1$ $\qquad$ ML case.

$$Pr(e|Y=1) = Pr(e|Y=2) = Q\left(\frac{\gamma}{2}\right) = Q\left(\frac{b-a}{2\sigma}\right)$$

Note: $\quad \gamma^2 = \dfrac{(b-a)^2}{\sigma^2} = \dfrac{\text{energy in signal diff}}{\text{noise energy}}$

Vector version

$Y=1$: $\quad \underline{X} \sim \mathcal{N}(\underline{a}, \sigma^2 I) \qquad \underline{X} = \underline{a} + \underline{z}$

$Y=2$: $\quad \underline{X} \sim \mathcal{N}(\underline{b}, \sigma^2 I) \qquad \underline{X} = b + \underline{z}$

note $\quad \underline{z} \sim \mathcal{N}(0, \sigma^2 I) = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$ $\qquad \|\cdot\|$ 2-norm

z_i iid $\sim \mathcal{N}(0, \sigma^2)$ $\quad i = 1, \dots, n$

$$f_{\underline{z}}(\underline{z}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^{n} \exp\left[-\frac{z_i^2}{2\sigma^2}\right] = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \underbrace{\exp \sum_{i=1}^{n} -\frac{z_i^2}{2\sigma^2}}_{\exp\left[-\|\underline{z}\|^2/2\sigma^2\right]}$$

$$f_{\underline{x}|y}(\underline{x}|1) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \sum_{i=1}^{n} -\frac{(x_i - a_i)^2}{2\sigma^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{\|\underline{x}-\underline{a}\|^2}{2\sigma^2}\right]$$

Contour surfaces are spheres

$$f_{\underline{x}|y}(\underline{x}|2) = \quad " \quad \exp\left[-\frac{\|\underline{x}-\underline{b}\|^2}{2\sigma^2}\right]$$

Likelihood ratio

$$\Lambda(\underline{x}) = \frac{f_{\underline{x}|y}(\underline{x}|2)}{f_{\underline{x}|y}(\underline{x}|1)} = \exp\left[\frac{\|\underline{x}-\underline{a}\|^2 - \|\underline{x}-\underline{b}\|^2}{2\sigma^2}\right]$$

$$= \exp\left[\frac{\underline{x}^T(\underline{b}-\underline{a})}{\sigma^2} + \frac{\|\underline{a}\|^2 - \|\underline{b}\|^2}{2\sigma^2}\right]$$

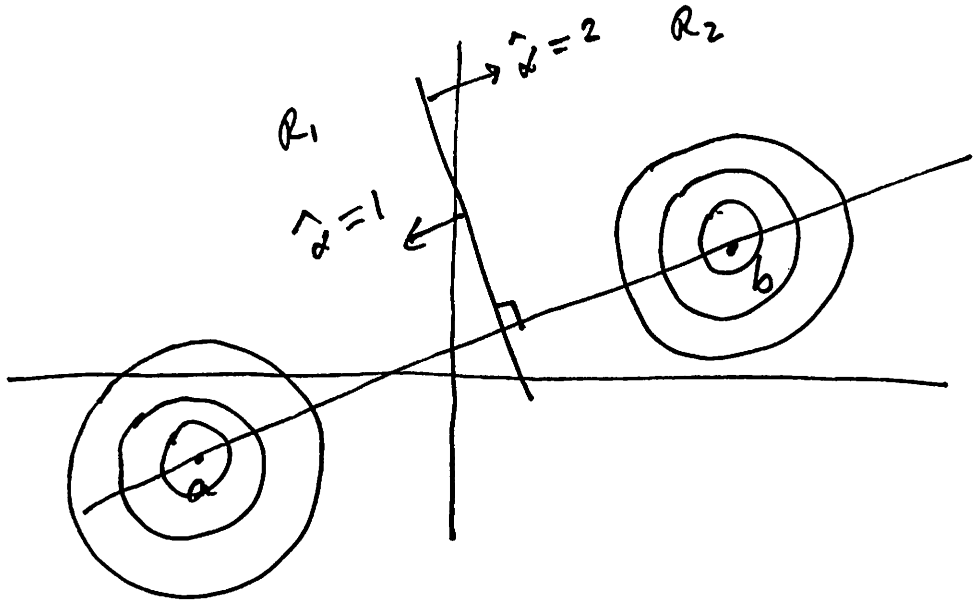
$$\ln\Lambda(\underline{x}) = \frac{\underline{x}^T(\underline{b}-\underline{a})}{\sigma^2} + \frac{\|\underline{a}\|^2 - \|\underline{b}\|^2}{2\sigma^2} \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}} \underbrace{\ln\frac{P_y(1)}{P_y(2)}}_{\ln T_L}$$

$$\Longleftrightarrow \quad \underline{x}^T(\underline{b}-\underline{a}) \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}} \underbrace{\sigma^2 \ln T_L + \frac{\|\underline{b}\|^2 - \|\underline{a}\|^2}{2}}_{\phi}$$

Another version:

$$\ln\Lambda(\underline{x}) = \frac{\|\underline{x}-\underline{a}\|^2 - \|\underline{x}-\underline{b}\|^2}{2\sigma^2} \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}} \underbrace{\ln\frac{P_y(1)}{P_y(2)}}_{\ln T_L}$$

$$\Longleftrightarrow \quad \|\underline{x}-\underline{a}\|^2 - \|\underline{x}-\underline{b}\|^2 \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}} 2\sigma^2 \ln T_L$$



$$\{\underline{x} : \Lambda(\underline{x}) = c\} = \{\underline{x} : \|\underline{x}-\underline{a}\|^2 - \|\underline{x}-\underline{b}\|^2\} = c'\}$$

$$= \{\underline{x} : \underline{x}^T(\underline{b}-\underline{a}) = c''\}$$

defines equation of <u>affine space</u> or <u>hyperplane</u> in $\mathbb{R}^n$

LLR test compares $\underbrace{\underline{x}^T(\underline{b}-\underline{a})}\ \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}}\ \phi$

<u>correlation</u> bet. $\underline{x}$ and $\underline{b}-\underline{a}$ is a <u>sufficient</u> <u>statistic</u>

<u>Correlation</u> <u>detector</u> - can be implemented using linear filter

- For case $P_Y(1) = P_Y(2) = 1/2$,

$$\|\underline{x}-\underline{a}\|^2\ \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}}\ \|\underline{x}-\underline{b}\|^2$$

<u>Minimum</u> <u>Distance</u> rule

Note $\|\underline{b}\|^2 - \|\underline{a}\|^2 = (\underline{b}-\underline{a})^T(\underline{b}+\underline{a})$

$$LLR(\underline{x}) = \ln \Lambda(\underline{x}) = \frac{(\underline{b}-\underline{a})^T}{\sigma^2}\left(\underline{x} - \frac{\underline{b}+\underline{a}}{2}\right) \underset{\hat{\alpha}=1}{\overset{\hat{\alpha}=2}{\gtrless}} \ln T_L \qquad (1)$$

For $P_Y(1) = P_Y(2) = 1/2$; $\ln T_L = 0$,

R_1, R_2 separated by perpendicular bisector bet. $\underline{a}$ and $\underline{b}$

- Error probabilities: $Pr(e|Y=1)$ and $Pr(e|Y=2)$

Given $Y=1$: $\underline{X} \sim \mathcal{N}(\underline{a}, \sigma^2 I)$

$$E\left[\underline{X} - \frac{\underline{b}+\underline{a}}{2} \,\middle|\, Y=1\right] = \frac{\underline{a}-\underline{b}}{2}$$

By (1), $$E[LLR(\underline{X}) | Y=1] = -\frac{(\underline{b}-\underline{a})^T(\underline{b}-\underline{a})}{2\sigma^2}$$

$$= -\frac{\gamma^2}{2}$$

where $\gamma = \dfrac{\|\underline{b}-\underline{a}\|}{\sigma}$

$(var(a^T\underline{z}) = a^T E[\underline{z}\,\underline{z}^T]\underline{a})$

$$var[LLR(\underline{X})|Y=1] = \frac{(\underline{b}-\underline{a})^T}{\sigma^2}\cancel{\sigma^2} I \frac{(\underline{b}-\underline{a})}{\sigma^2} = \gamma^2$$

Thus given $Y=1$, $LLR(\underline{X}) \sim \mathcal{N}(-\frac{\gamma^2}{2}, \gamma^2)$

$\underline{X}$ is a Gaussian random vector, thus $LLR(\underline{X})$ is a Gaussian r.v.

$$Pr(e|Y=1) = Pr(LLR(\underline{X}) \geq \ln T_L | Y=1)$$

$$= Pr\left(\frac{LLR(\underline{X}) + \frac{\gamma^2}{2}}{\gamma} \geq \frac{\ln T_L + \frac{\gamma^2}{2}}{\gamma} = \frac{\ln T_L}{\gamma} + \frac{\gamma}{2}\right)$$

$$\Pr(e \mid Y=1) = Q\left(\frac{\ln T_L}{\gamma} + \frac{\gamma}{2}\right)$$

similarly, given $Y=2$, $LLR(\underline{x}) \sim \mathcal{N}\left(\frac{\gamma^2}{2}, \gamma^2\right)$

$$\Pr(e \mid Y=2) = Q\left(-\frac{\ln T_L}{\gamma} + \frac{\gamma}{2}\right)$$

Note $\Pr(e \mid Y=i)$ function only of $\gamma = \frac{\|\underline{b}-\underline{a}\|}{\sigma}$

$\|\underline{b}-\underline{a}\|^2$ = energy in difference bet. signals

σ^2 = noise energy per measurement

- View γ^2 as signal-to-noise ratio

Generalize to Gaussian random vector

$\underline{z} \sim \mathcal{N}(\underline{0}, \Sigma_z)$ $\{z_1, \ldots, z_n\}$ a set of jointly Gaussian r.v.'s

$$f_{\underline{z}}(\underline{z}) = \frac{\exp\left[-\frac{1}{2}\underline{z}^T \Sigma_{\underline{z}}^{-1} \underline{z}\right]}{(2\pi)^{n/2}\sqrt{\det(\Sigma_{\underline{z}})}}$$

$\det(\Sigma_z) \neq 0$

contour surfaces are ellipsoids

$E[\underline{z}] = \underline{0}$

symmetric matrix psd $\rightarrow$ $\Sigma_z = E[\underline{z}\,\underline{z}^T]$ $(\Sigma_z)_{ij} = E[z_i z_j]$

$\underline{s}^T \underline{z} = \sum_{i=1}^{n} s_i z_i$ is a Gaussian r.v.

where $\underline{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$

$\underline{u} \sim \mathcal{N}(\underline{m}, \Sigma_u)$

$$f_{\underline{u}}(\underline{u}) = \frac{\exp\left[-\frac{1}{2}(\underline{u}-\underline{m})^T \Sigma_u^{-1} (\underline{u}-\underline{m})\right]}{(2\pi)^{n/2}\sqrt{\det(\Sigma_u)}}$$

$\det(\Sigma_u) \neq 0$

Go to 2-D example (p. 8)

Back to Bayesian decision theory

$Y=1$: $$f_{\underline{X}|Y}(\underline{x}|1) = \frac{\exp[-\frac{1}{2}(\underline{x}-\underline{a})^T \Sigma_z^{-1}(\underline{x}-\underline{a})]}{(2\pi)^{n/2}\sqrt{\det(\Sigma_z)}}$$

$Y=2$: $$f_{\underline{X}|Y}(\underline{x}|2) = \frac{\exp(-\frac{1}{2}(\underline{x}-\underline{b})^T \Sigma_z^{-1}(\underline{x}-\underline{b}))}{(2\pi)^{n/2}\sqrt{\det(\Sigma_z)}}$$

$$LLR(\underline{x}) = \ln \frac{f_{\underline{X}|Y}(\underline{x}|2)}{f_{\underline{X}|Y}(\underline{x}|1)} = \frac{1}{2}(\underline{x}-\underline{a})^T\Sigma_z^{-1}(\underline{x}-\underline{a}) - \frac{1}{2}(\underline{x}-\underline{b})^T\Sigma_z^{-1}(\underline{x}-\underline{b})$$

$$= (\underline{b}-\underline{a})^T \Sigma_z^{-1}\underline{x} + \frac{1}{2}\underline{a}^T\Sigma_z^{-1}\underline{a} - \frac{1}{2}\underline{b}^T\Sigma_z^{-1}\underline{b}$$

$$= (\underline{b}-\underline{a})^T\Sigma_z^{-1}\left(\underline{x} - \frac{\underline{b}+\underline{a}}{2}\right)$$

Thus, $(\underline{b}-\underline{a})^T\Sigma_z^{-1}\underline{x}$ is a sufficient statistic

~~equivalently,~~

Error probability: $Pr(e|Y=1)$ and $Pr(e|Y=2)$

$$E[LLR(\underline{X}|Y=1)] = -\frac{(\underline{b}-\underline{a})^T\Sigma_z^{-1}(\underline{b}-\underline{a})}{2} = -\frac{\gamma^2}{2}$$

(Given $Y=1$: $\underline{X} \sim \mathcal{N}(\underline{a}, \Sigma_z)$)

where $\gamma = [(\underline{b}-\underline{a})^T\Sigma_z^{-1}(\underline{b}-\underline{a})]^{1/2}$

(cf. $\gamma = \frac{\|\underline{b}-\underline{a}\|}{\sigma}$ for $\underline{z} \sim \mathcal{N}(0, \sigma^2 I)$)

$$Var(LLR(\underline{X})|Y=1) = (\underline{b}-\underline{a})^T\Sigma_z^{-1}\Sigma_z\Sigma_z^{-T}(\underline{b}-\underline{a})$$
$$= \gamma^2$$

Since $\underline{X}$ is a Gaussian random vector,

Given $Y=1$: $\quad LLR(\underline{X}) \sim \mathcal{N}(-\frac{\gamma^2}{2}, \gamma^2)$

$Y=2$: $\quad LLR(\underline{X}) \sim \mathcal{N}(\frac{\gamma^2}{2}, \gamma^2)$

$$Pr(e \mid Y=1) = Q\left(\frac{\ln T_L}{\gamma} + \frac{\gamma}{2}\right)$$

$$Pr(e \mid Y=2) = Q\left(-\frac{\ln T_L}{\gamma} + \frac{\gamma}{2}\right)$$

<u>Multicategory case</u> $\quad \mathcal{Y} = \{1, \ldots, m\}$, 0-1 loss fns.

$Y=i$: $\quad \underline{X} \sim \mathcal{N}(\underline{\mu}_i, \Sigma_i)$

Decide $\hat{y}(\underline{x}) = i$ if $\quad g_i(\underline{x}) \geq g_j(\underline{x}) \quad \forall j \neq i$

where $\quad g_i(\underline{x}) = \ln f_{\underline{X}|Y}(\underline{x} \mid i) + \ln P_Y(i)$

Decision regions $\mathcal{R}_1, \ldots, \mathcal{R}_m$ partitioning $\mathcal{X}$

if $\quad g_i(\underline{x}) \geq g_j(\underline{x}) \quad \forall j \neq i$, then $\underline{x} \in \mathcal{R}_i$

<u>$\Sigma_i = \sigma^2 I$:</u>

$$g_i(\underline{x}) = -\frac{\|\underline{x} - \underline{\mu}_i\|^2}{2\sigma^2} + \ln P_Y(i)$$

equivalently,

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

<u>linear</u> discriminant functions

$\Sigma_i = \Sigma$:

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

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

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

Equivalently

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