Week 5 : Session 1

Gaussian Discriminant Analysis (GDA)

Also known as Linear Discriminant Analysis (LDA)

Consider a n dimensional Gaussian Random vector ($R^{n \times 1}$) with mean $E\{\bar{x}\} = \bar{\mu}$ and covariance Matrix $R = E\{(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\}$, then the PDF of the Gaussian Random vector is given as

$$f_{\bar{X}}(\bar{x}) = \frac{1}{\sqrt{(2\pi)^n |R|}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T R^{-1} (\bar{x}-\bar{\mu})}$$

This is also known as Multivariate Gaussian PDF.

If the samples are zero (ie) $\bar{\mu} = 0$, $R = \sigma^2 I$: (iid zero mean), $|R| = (\sigma^2)^n$, then

$$f_{\bar{X}}(\bar{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{\|\bar{x}\|^2}{2\sigma^2}} \quad \text{(for iid zero mean)}$$

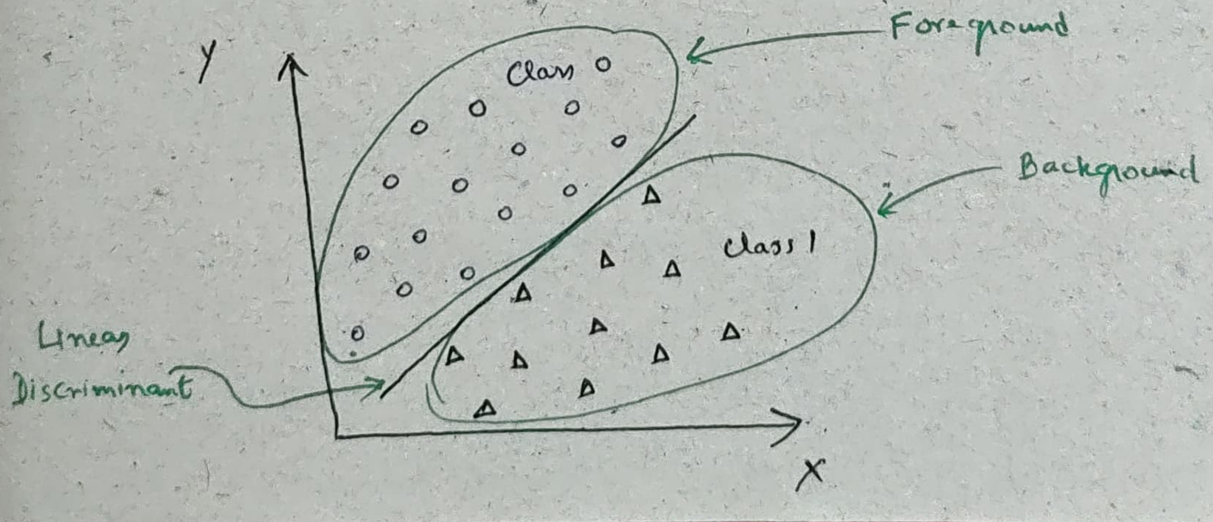
Gaussian classification

Let us consider the problem of classification with two Gaussian classes with different Mean and same covariance.

$C_0 : \bar{\mu}_0, R$

$C_1 : \bar{\mu}_1, R$

Example : Foreground pixels — C_0
Background pixels — C_1



Consider the input vectors $\bar{x}$ drawn from two Gaussian classes

$$C_0 : \mathcal{N}(\bar{\mu}_0, R)$$

$$C_1 : \mathcal{N}(\bar{\mu}_1, R)$$

(i) How to separate the points belonging to these two different classes?

(ii) How to assign a new point to one of the two different classes?

Essentially, these are the problems of the Gaussian discriminant analysis.

Thus, the Likelihoods of the two classes are

⊙ Likelihood of class 0

$$p(\bar{x}; C_0) = \frac{1}{\sqrt{(2\pi)^n |R|}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_0)^T R^{-1} (\bar{x}-\bar{\mu}_0)}$$

⊙ Likelihood of class 1

$$p(\bar{x}; C_1) = \frac{1}{\sqrt{(2\pi)^n |R|}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_1)^T R^{-1} (\bar{x}-\bar{\mu}_1)}$$

where, $|R| \rightarrow$ Determinant of R

Now, choose the class that Maximizes the Likelihood (ML)

(i.e.) Choose the class that has highest likelihood.

Therefore, choose C_0 if

$$p(\bar{x}; C_0) \geq p(\bar{x}; C_1)$$

choose C_1 if

$$p(\bar{x}; C_0) < p(\bar{x}; C_1)$$

This can be simplified as follows.

Choose C_0 if

$$p(\bar{x}; C_0) \geq p(\bar{x}; C_1)$$

$$\Rightarrow \frac{1}{\cancel{\sqrt{(2\pi)^n|R|}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_0)^T R^{-1}(\bar{x}-\bar{\mu}_0)} \geq \frac{1}{\cancel{\sqrt{(2\pi)^n|R|}}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_1)^T R^{-1}(\bar{x}-\bar{\mu}_1)}$$

$$\Rightarrow (\bar{x}-\bar{\mu}_0)^T R^{-1}(\bar{x}-\bar{\mu}_0) \leq (\bar{x}-\bar{\mu}_1)^T R^{-1}(\bar{x}-\bar{\mu}_1)$$

$$\Rightarrow \left(\bar{x}^T\bar{x} + \bar{\mu}_0^T\bar{\mu}_0 - 2\bar{x}\bar{\mu}_0^T\right)R^{-1} \leq \left(\bar{x}^T\bar{x} + \bar{\mu}_1^T\bar{\mu}_1 - 2\bar{x}\bar{\mu}_1^T\right)R^{-1}$$

$$\Rightarrow \cancel{\bar{x}^T R^{-1}\bar{x}} + \bar{\mu}_0^T R^{-1}\bar{\mu}_0 - 2\bar{x}R^{-1}\bar{\mu}_0^T \leq \cancel{\bar{x}^T R^{-1}\bar{x}} + \bar{\mu}_1^T R^{-1}\bar{\mu}_1 - 2\bar{x}R^{-1}\bar{\mu}_1^T$$

$$\Rightarrow \bar{\mu}_1^T R^{-1}\bar{\mu}_1 - 2\bar{x}R^{-1}\bar{\mu}_1^T - \bar{\mu}_0^T R^{-1}\bar{\mu}_0 + 2\bar{x}R^{-1}\bar{\mu}_0^T \geq 0$$

$$\Rightarrow 2\bar{x}R^{-1}\left(\bar{\mu}_0^T - \bar{\mu}_1^T\right) + \left(\bar{\mu}_1^T R^{-1}\bar{\mu}_1 - \bar{\mu}_0^T R^{-1}\bar{\mu}_0\right) \geq 0$$

Divide both sides by 2

$$\Rightarrow (\bar{\mu}_0-\bar{\mu}_1)^T R^{-1}\bar{x} + \frac{1}{2}\left(\bar{\mu}_1^T R^{-1}\bar{\mu}_1 - \bar{\mu}_0^T R^{-1}\bar{\mu}_0\right) \geq 0$$

$$\Rightarrow (\bar{\mu}_0-\bar{\mu}_1)^T R^{-1}\bar{x} - \frac{1}{2}\left(\bar{\mu}_0^T R^{-1}\bar{\mu}_0 - \bar{\mu}_1^T R^{-1}\bar{\mu}_1\right) \geq 0$$

$$\Rightarrow (\bar{\mu}_0-\bar{\mu}_1)^T R^{-1}\bar{x} - \frac{1}{2}\left[R^{-1}(\bar{\mu}_0+\bar{\mu}_1)(\bar{\mu}_0-\bar{\mu}_1)^T\right] \geq 0$$

$$\Rightarrow \underbrace{(\bar{\mu}_0-\bar{\mu}_1)^T R^{-1}}_{\bar{h}^T}\left(\bar{x} - \underbrace{\left(\frac{\bar{\mu}_0+\bar{\mu}_1}{2}\right)}_{\tilde{\mu}}\right) \geq 0$$

$\Rightarrow \bar{h}^T(\bar{x}-\tilde{\mu}) \geq 0$, which is an N dimensional plane.

Thus, choose C_0 if

$$\bar{h}^T(\bar{x}-\tilde{\mu}) \geq 0$$

choose C_1 if

$$\bar{h}^T(\bar{x}-\tilde{\mu}) < 0$$

where, $\bar{h} = R^{-1}(\bar{\mu}_0 - \bar{\mu}_1)$

$$\tilde{\mu} = \frac{\bar{\mu}_0+\bar{\mu}_1}{2}$$

Mid point of $\bar{\mu}_0, \bar{\mu}_1$

Another way to write the same classifier.

$$\Rightarrow -(\bar{\mu}_0-\bar{\mu}_1)^T R^{-1}\left(\bar{x} - \left(\frac{\bar{\mu}_0+\bar{\mu}_1}{2}\right)\right) \leq 0$$

$$\Rightarrow (\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}\bar{x} \leq \frac{1}{2}(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1+\bar{\mu}_0)$$

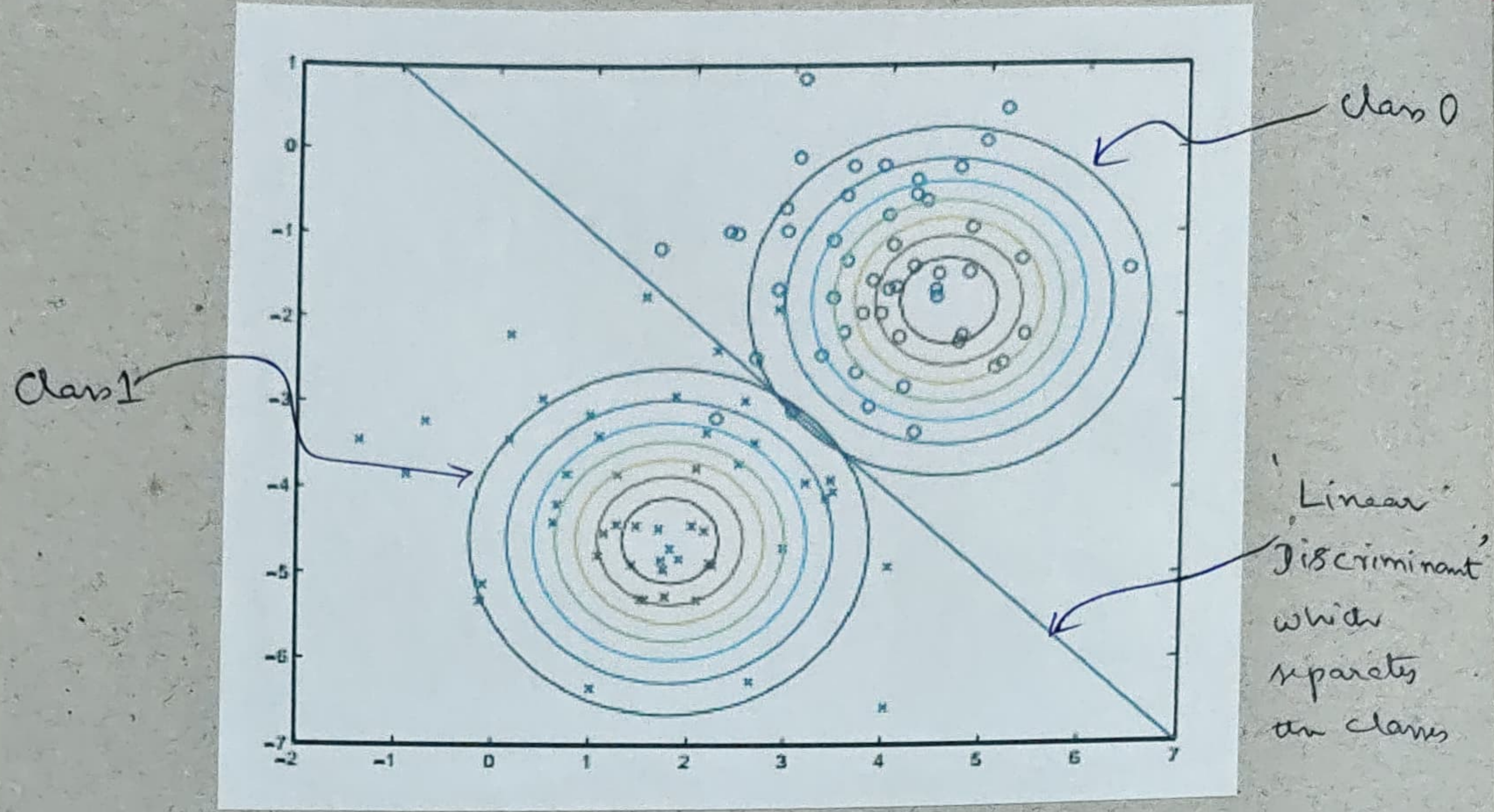


Fig. Gaussian classification

Special case:

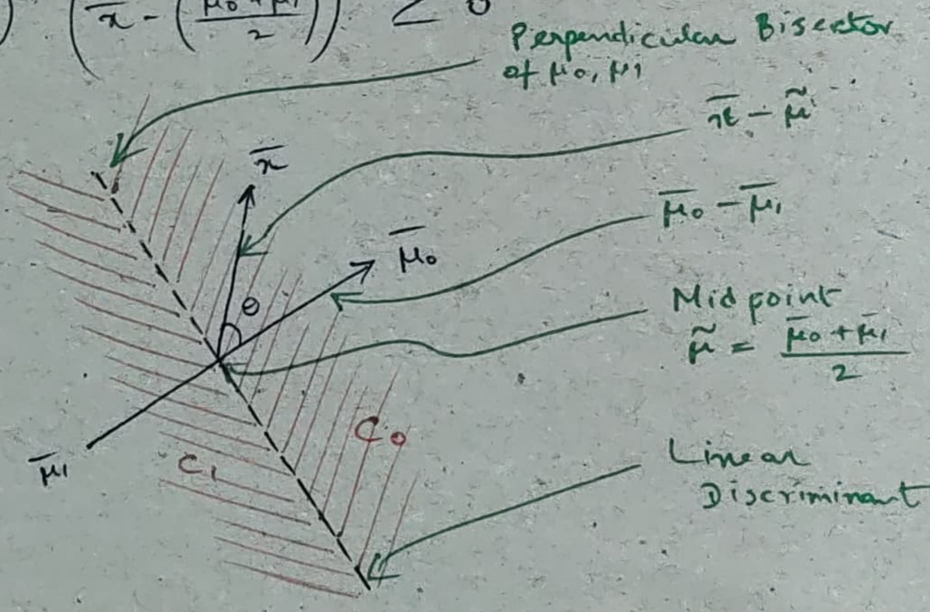
Consider the special case $R = \sigma^2 I$, it follows that $\bar{h} = R^{-1}(\bar{\mu}_0 - \bar{\mu}_1) = \frac{1}{\sigma^2}(\bar{\mu}_0 - \bar{\mu}_1)$. Thus, the hyperplane reduced to: choose C_0 if

$$(\bar{\mu}_0 - \bar{\mu}_1)^T R^{-1}\left(\bar{x} - \left(\frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right)\right) \geq 0$$

$$\Rightarrow (\bar{\mu}_0 - \bar{\mu}_1)^T \frac{1}{\sigma^2}\left(\bar{x} - \left(\frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right)\right) \geq 0$$

$$\Rightarrow (\bar{\mu}_0 - \bar{\mu}_1)^T \left(\bar{x} - \left(\frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right)\right) \geq 0$$

Pictorially,



The points on the right side of the perpendicular bisector (i.e.) closer to $\bar{\mu}_0$, will be classified as C_0, whereas the points on the left side of the perpendicular bisector (i.e.) closer to $\bar{\mu}_1$, will be classified as C_1. Thus, the Linear Discriminant Analysis (LDA) reduces to Nearest Neighbour Decoding Rule.

So, for IID Gaussian noise samples (mean = 0, Variance = σ^2) (or) if the covariance Matrix $R = \sigma^2 I$, the LDA reduces to Nearest Neighbour Decoding Rule.

It can be seen that the dot product

- $(\bar{\mu}_0 - \bar{\mu}_1)^T(\bar{x} - \tilde{\mu}) = \|\bar{\mu}_0 - \bar{\mu}_1\| \, \|\bar{x} - \tilde{\mu}\| \cos\theta \geq 0$ when $-90^\circ \leq \theta \leq 90^\circ$, as $\cos\theta \geq 0$.
- $(\bar{\mu}_0 - \bar{\mu}_1)^T(\bar{x} - \tilde{\mu}) = \|\bar{\mu}_0 - \bar{\mu}_1\| \, \|\bar{x} - \tilde{\mu}\| \cos\theta < 0$ when $90^\circ < \theta < 270^\circ$, as $\cos\theta < 0$.

Thus, the hyperplane is the perpendicular bisector of $\bar{\mu}_0$ and $\bar{\mu}_1$.

Week 5 : Session 2

Example.

Determine the classifier for the Binary Gaussian classification problem, with two classes C_0, C_1 distributed as

$$C_0 \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} 2 \\ -4 \end{bmatrix}}_{\bar{\mu}_0}, \underbrace{\begin{bmatrix} 1/4 & 0 \\ 0 & 1/8 \end{bmatrix}}_{R}\right), \quad C_1 \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} -4 \\ 2 \end{bmatrix}}_{\bar{\mu}_1}, \underbrace{\begin{bmatrix} 1/4 & 0 \\ 0 & 1/8 \end{bmatrix}}_{R}\right)$$

Given: $\bar{\mu}_0 = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$, $\bar{\mu}_1 = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$, $R = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/8 \end{bmatrix} \Rightarrow R^{-1} = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}$

classifier chooses C_0 if

$$(\bar{\mu}_0 - \bar{\mu}_1)^T R^{-1} \left(\bar{x} - \frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right) \geq 0$$

$$\bar{\mu}_0 - \bar{\mu}_1 = \begin{bmatrix} 2 \\ -4 \end{bmatrix} - \begin{bmatrix} -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \end{bmatrix}$$

$$\frac{\bar{\mu}_0 + \bar{\mu}_1}{2} = \frac{1}{2}\left(\begin{bmatrix} 2 \\ -4 \end{bmatrix} + \begin{bmatrix} -4 \\ 2 \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

Classifier chooses C_0 if

$$\begin{bmatrix} 6 & -6 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} -1 \\ -1 \end{bmatrix} \right) \geq 0$$

$$\Rightarrow \begin{bmatrix} 6 & -6 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} x_1+1 \\ x_2+1 \end{bmatrix} \geq 0$$

$$\Rightarrow \begin{bmatrix} 6 & -6 \end{bmatrix} \begin{bmatrix} 4(x_1+1) \\ 8(x_2+1) \end{bmatrix} \geq 0$$

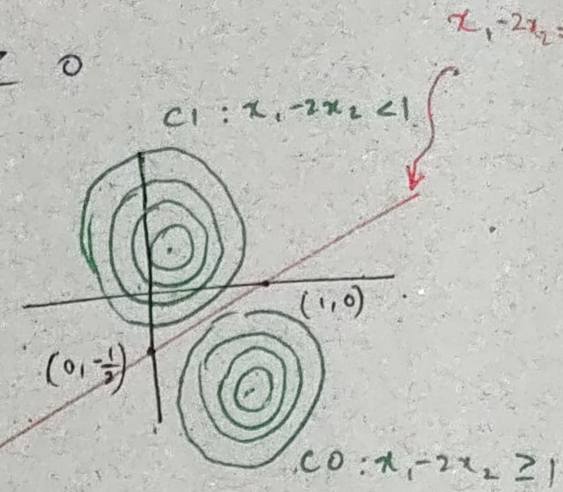
$$\Rightarrow (6)(4)(x_1+1) - (6)(8)(x_2+1) \geq 0$$

$$\Rightarrow 24x_1 + 24 - 48x_2 - 48 \geq 0$$

$$\Rightarrow 24x_1 - 48x_2 - 24 \geq 0$$

$$\Rightarrow x_1 - 2x_2 - 1 \geq 0$$

$\Rightarrow x_1 - 2x_2 \geq 1$, which is called as Linear Discriminant Function.



Analysis

Classifier chooses C_0 if $x_1 - 2x_2 \geq 1$

C_1 if $x_1 - 2x_2 < 1$

Recall, the classifier chooses C_0 if

$$(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x} \leq \frac{1}{2} (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 + \bar{\mu}_0)$$

classifier chooses C_1 if

$$(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x} > \frac{1}{2} (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 + \bar{\mu}_0)$$

(i) What is P_{FA} ?

- Under H_0, what is probability decision $= H1$?
- Under H_0, $\bar{x} \sim \mathcal{N}(\bar{\mu}_0, R)$

The Gaussian Random vector under C_0 is $(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x}$.

Mean: $E\{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x}\} = (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} E\{\bar{x}\}$

$$= (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{\mu}_0$$

Variance:

$$E\{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{x} - \bar{\mu}_0)(\bar{x} - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)\}$$

$$= (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \underbrace{E\{(\bar{x} - \bar{\mu}_0)(\bar{x} - \bar{\mu}_0)^T\}}_{R} R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)$$

$$= (\bar{\mu}_1 - \bar{\mu}_0)^T \cancel{R^{-1}} \cancel{R} R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)$$

$$= (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)$$

Therefore,

$$(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x} \sim \mathcal{N}\Big(\underbrace{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{\mu}_0}_{\text{Mean } (\mu)}, \underbrace{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}_{\text{Variance } (\sigma^2)}\Big)$$

FA occurs when under, H_0, if decision is H_1.

$$\Rightarrow P_{FA} = P_r\left((\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x} > \frac{1}{2}(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 + \bar{\mu}_0)\right)$$

Now, we use the property

$$P_r(X > \alpha) = P_r\Big(\underbrace{\frac{X-\mu}{\sigma}}_{\text{Standard Gaussian RV.}} > \frac{\alpha - \mu}{\sigma}\Big) = Q\left(\frac{\alpha-\mu}{\sigma}\right)$$

$$\Rightarrow Q\left(\frac{\frac{1}{2}(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 + \bar{\mu}_0) - (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{\mu}_0}{\sqrt{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}}\right)$$

$$\Rightarrow Q\left(\frac{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \left(\frac{\bar{\mu}_1}{2} + \frac{\bar{\mu}_0}{2} - \bar{\mu}_0\right)}{\sqrt{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}}\right)$$

$$\Rightarrow Q\left(\frac{\frac{1}{2}(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}{\sqrt{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}}\right)$$

$$\Rightarrow Q\left(\frac{1}{2}\sqrt{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}\right)$$

Substituting $R = \sigma^2 I$

$$\Rightarrow Q\left(\frac{1}{2}\sqrt{(\bar{\mu}_1 - \bar{\mu}_0)^T \frac{1}{\sigma^2} I (\bar{\mu}_1 - \bar{\mu}_0)}\right)$$

$$\Rightarrow Q\left(\frac{\|\bar{\mu}_1 - \bar{\mu}_0\|}{2\sigma}\right),$$ which is the P_{FA} for Gaussian classification.

(ii) What is P_D?

- Under H_1, what is probability that decision $= H_1$?
- Under H_1, $\bar{x} \sim \mathcal{N}(\bar{\mu}_1, R)$

The Gaussian Random vector under C_1 is $(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x}$.

<u>Mean</u>:

$$E\{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{x}\} = (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} E[\bar{x}]$$
$$= (\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} \bar{\mu}_1$$

Variance:

$$E\left\{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{x}-\bar{\mu}_1)(\bar{x}-\bar{\mu}_1)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)\right\}$$

$$= (\bar{\mu}_1-\bar{\mu}_0)^T R^{-1} \underbrace{E\left\{(\bar{x}-\bar{\mu}_1)(\bar{x}-\bar{\mu}_1)^T\right\}}_{R} R^{-1}(\bar{\mu}_1-\bar{\mu}_0)$$

$$= (\bar{\mu}_1-\bar{\mu}_0)^T \cancel{R^{-1}}\,\cancel{R}\, R^{-1}(\bar{\mu}_1-\bar{\mu}_0)$$

$$= (\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)$$

Therefore,

$$(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}\bar{x} \sim \mathcal{N}\left(\underbrace{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}\bar{\mu}_1}_{\text{Mean }(\mu)},\ \underbrace{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}_{\text{Variance }(\sigma^2)}\right)$$

Detection occurs when under H_1, if decision is H_1.

$$\Rightarrow P_D = P_r\left((\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}\bar{x} > \frac{1}{2}(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1+\bar{\mu}_0)\right)$$

We use the property,

Standard Gaussian RV

$$P_r(X > \alpha) = P_r\left(\frac{X-\mu}{\sigma} > \frac{\alpha-\mu}{\sigma}\right) = Q\left(\frac{\alpha-\mu}{\sigma}\right)$$

$$\Rightarrow Q\left(\frac{\frac{1}{2}(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1+\bar{\mu}_0) - (\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}\bar{\mu}_1}{\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}}\right)$$

$$\Rightarrow Q\left(\frac{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}\left(\frac{\bar{\mu}_1}{2}+\frac{\bar{\mu}_0}{2}-\bar{\mu}_1\right)}{\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}}\right)$$

$$\Rightarrow Q\left(\frac{-\frac{1}{2}(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}{\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}}\right)$$

$$\Rightarrow Q\left(-\frac{1}{2}\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}\right)$$, which is the P_D for Gaussian classification.

(iii) P_{MD} is given as

$$P_{MD} = 1 - P_D = 1 - Q\left(-\frac{1}{2}\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}\right)$$

$$= Q\left(\frac{1}{2}\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1}(\bar{\mu}_1-\bar{\mu}_0)}\right)$$

$$= P_{FA}$$

$$\boxed{1 - Q(-x) = Q(x)}$$

Thus, the Probability of Error for Gaussian Classification problem.

$$P_e = \frac{1}{2} P_{FA} + \frac{1}{2} P_{MD}$$

$$= Q\left(\frac{1}{2}\sqrt{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}\right)$$

where, $\bar{\mu}_0 \rightarrow$ Mean of class 0

$\bar{\mu}_1 \rightarrow$ Mean of class 1

$R \rightarrow$ Covariance Matrix.

Example.

$\bar{\mu}_0$ $\quad$ R $\quad$ $\bar{\mu}_1$

For the Gaussian classification problem with classes C_0, C_1, what is P_e?

$$C_0 \sim N\left(\begin{bmatrix} 2 \\ -4 \end{bmatrix}, \begin{bmatrix} 1/4 & 0 \\ 0 & 1/8 \end{bmatrix}\right), \quad C_1 \sim N\left(\begin{bmatrix} -4 \\ 2 \end{bmatrix}, \begin{bmatrix} 1/4 & 0 \\ 0 & 1/8 \end{bmatrix}\right)$$

Given: $\bar{\mu}_0 = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$, $\bar{\mu}_1 = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$, $R = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/8 \end{bmatrix} \Rightarrow R^{-1} = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}$

$$P_e = Q\left(\frac{1}{2}\sqrt{(\bar{\mu}_1 - \bar{\mu}_0)^T R^{-1} (\bar{\mu}_1 - \bar{\mu}_0)}\right)$$

$$= Q\left(\frac{1}{2}\sqrt{\begin{bmatrix} -6 & 6 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} 6 \\ -6 \end{bmatrix}}\right)$$

$$= Q\left(\frac{1}{2}\sqrt{\begin{bmatrix} -6 & 6 \end{bmatrix} \begin{bmatrix} 24 \\ -48 \end{bmatrix}}\right)$$

$$= Q\left(\frac{1}{2}\sqrt{(-6)(24) + (6)(-48)}\right)$$

$$= Q\left(\frac{1}{2}\sqrt{432}\right)$$

$$= Q(6\sqrt{3})$$

Week 5 : Session 3

Signal Design Problem

How to choose the optimal signals $\bar{\mu}_1, \bar{\mu}_0$, so as to minimize the P_e? P_e is given as $Q\left(\frac{1}{2}\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1} (\bar{\mu}_1-\bar{\mu}_0)}\right)$, which is a decreasing function (Complementary Cumulative Distribution function CCDF of the standard Gaussian Random Variable).

Note : CDF increases from 0 to 1
CCDF decreases from 1 to 0.

Thus, $Q\left(\frac{1}{2}\sqrt{(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1} (\bar{\mu}_1-\bar{\mu}_0)}\right)$ is a decreasing function.

To minimize P_e, we've to maximize $(\bar{\mu}_1-\bar{\mu}_0)^T R^{-1} (\bar{\mu}_1-\bar{\mu}_0)$.

Let $(\bar{\mu}_1-\bar{\mu}_0) = \bar{s}$.

$$\Rightarrow \max \ \bar{s}^T R^{-1} \bar{s}$$

Let R have the Eigen value decomposition.

Diagonal matrix of Eigen values of R

$$R = U \Lambda U^T$$

$$= \underbrace{[\bar{u}_1 \ldots \bar{u}_N]}_{\text{Eigen vectors}} \overbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_N \end{bmatrix}}^{\Lambda} \begin{bmatrix} \bar{u}_1^T \\ \vdots \\ \bar{u}_N^T \end{bmatrix}$$

Property / Definition of Eigen value / Eigen vector.

$$R\bar{u}_i = \lambda_i \bar{u}_i$$

Eigen value

Eigen vector

We can find these Eigen values and Eigen vectors by solving the characteristic equation

Determinant

$$|R - \lambda I| = 0$$

Note the following properties.

$$R = U \Lambda U^T$$

$\bar{u}_i^T \bar{u}_j = 0$ if $i \neq j$

$\|\bar{u}_i\|^2 = 1$

where R is a Covariance Matrix, which means R is positive definite. Therefore, the Eigen values are greater than 0. (ie) $\lambda_i > 0$. Also, since R is considered to be Invertible; it means that none of the Eigen values are zero.

The Eigen vectors are Orthonormal (Orthogonal and Unit Norm)

(ie) $UU^T = U^TU = I$

where $U \rightarrow$ Unitary Matrix.

Let $\bar{s} = U\bar{\alpha}$.

Expanding $\bar{s}$ using the basis of orthonormal eigenvectors of R.

$$\Rightarrow \bar{s} = [\bar{u}_1\ \bar{u}_2 \dots \bar{u}_N] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \alpha_1 \bar{u}_1 + \alpha_2 \bar{u}_2 + \dots + \alpha_N \bar{u}_N = \sum_i \alpha_i \bar{u}_i$$

Furthermore, R^{-1} is given as

$$R^{-1} = U\Lambda^{-1}U^T = U \underbrace{\begin{bmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_N} \end{bmatrix}}_{\Lambda^{-1}} U^T$$

Diagonal matrix of Eigen values of R^{-1}

$$\begin{aligned} R^{-1} \cdot R &= U\Lambda^{-1}\cancel{U^T}\cancel{U}\Lambda U^T \\ &= U\cancel{\Lambda^{-1}}\cancel{\Lambda} \cdot U^T \\ &= \cancel{U}\cancel{U^T} \\ &= I \end{aligned}$$

So, R^{-1} can be evaluated by simply replacing Λ by Λ^{-1} in the Eigen value decomposition.

Therefore, one can formulate the optimization problem as Maximizing $\bar{s}^T R^{-1} \bar{s}$, subject to the constraint $\|\bar{s}\|^2 = 1$.

$$\left.\begin{aligned} &\text{objective function} : \max\ \bar{s}^T R^{-1} \bar{s} \\ &\text{Constraint} : \|\bar{s}\|^2 = 1 \end{aligned}\right\} \rightarrow \text{Constrained Optimization Problem.}$$

The quantity $\bar{s}^T R^{-1} \bar{s}$ is unbounded above, if $\|\bar{s}\|$ increases. As we cannot increase the signal power in an unbounded fashion, therefore we set $\|\bar{s}\|^2 = 1$.

We have $\bar{s} = U\bar{\alpha} \Rightarrow \bar{s}^T = \bar{\alpha}^T U^T$. Thus,

$$\begin{aligned} \bar{s}^T R^{-1} \bar{s} &= \bar{\alpha}^T \cancel{U^T} \cdot \cancel{U} \Lambda^{-1} \cancel{U^T} \cdot \cancel{U} \bar{\alpha} \\ &= \bar{\alpha}^T \Lambda^{-1} \bar{\alpha} \\ &= [\alpha_1 \dots \alpha_N] \begin{bmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_N} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \\ &= \sum_i \frac{\alpha_i^2}{\lambda_i} \end{aligned}$$

$$\|\bar{s}\|^2 = \bar{s}^T \bar{s} = \bar{\alpha}^T U^T U \bar{\alpha}$$
$$= \bar{\alpha}^T \alpha = \sum_i \alpha_i^2 = \|\bar{\alpha}\|^2$$

Thus, the Optimization problem can be modified as

$$\max \ \bar{s}^T R^{-1} \bar{s} \equiv \sum_i \frac{\alpha_i^2}{\lambda_i}, \text{ subject to the}$$

constraint $\|\bar{s}\|^2 = 1 \equiv \sum_i \alpha_i^2 = 1$

Solution is,

- Set $\alpha_i^2 = 1$, for i such that λ_i is minimum!
- Rest $\alpha_i^2 = 0$.

(i)

- $\alpha_i = 1$, $i = \arg\min_j \lambda_j \Rightarrow \bar{s} = \bar{u}_i$
- $\alpha_i = 0$, otherwise.

$\bar{u}_i$ = Orthonormal Eigen vector of R corresponding to minimum Eigen value.

Intuitively, λ_j denotes the noise power along $\bar{u}_j$. Therefore, result says allocate all power to $\bar{u}_i$, such that noise power λ_i is minimum !!!

(ii) Choose $\bar{s}$ along direction $\bar{u}_i$ which noise power λ_i is min!

$$\bar{s} = \bar{\mu}_1 - \bar{\mu}_0 = \bar{u}_i \quad \text{to minimize } P_D.$$

Example:

$$R = \begin{bmatrix} -2 & -2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}$$

Consider the LDA based classifier for the classification of two Gaussian classes $\mathcal{N}(\bar{\mu}_0, R)$, $\mathcal{N}(\bar{\mu}_1, R)$ with the Covariance Matrix R above. The optimal signal $\bar{s} = \bar{\mu}_1 - \bar{\mu}_0$ that minimizes the probability of error is ________ ?

The goal is to compute $\bar{s}$, which minimizes the probability of error in the LDA framework.

The Eigen value decomposition is given as

$$R = \underbrace{\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}}_{U} \begin{bmatrix} 2\times 2\sqrt{2}\times\sqrt{2} & \\ & 4\times 2\sqrt{2}\times\sqrt{2} \end{bmatrix} \underbrace{\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}}_{U^T}$$

$$\Rightarrow R = \begin{bmatrix} \underbrace{\begin{matrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{matrix}}_{\bar{u}_1} & \underbrace{\begin{matrix} \frac{-1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{matrix}}_{\bar{u}_2} \end{bmatrix} \underbrace{\begin{bmatrix} \overset{\lambda_1}{8} & \\ & \overset{\lambda_2}{16} \end{bmatrix}}_{\Lambda} \begin{array}{l} \left[\begin{matrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{matrix}\right] \bar{u}_1^T \\ \left[\begin{matrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{matrix}\right] \bar{u}_2^T \end{array}$$

$$\min \lambda_j = \lambda_1 = 8$$

$$\bar{u}_i = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Therefore, the Eigen vector corresponding to minimum Eigen value is

$$\bar{u}_i = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \bar{\mu}_i - \bar{\mu}_0 = \bar{s}$$

This is the signal that minimizes the P_e of Gaussian classification.