Week 5 : Session 3.

8. Linear Discriminant Analysis (LDA)

- Linear Discriminant Analysis (LDA) is another ML technique for Binary classification problem, which relies on the concept of Discriminant.
- This technique is also called as Gaussian Discriminant Analysis (GDA).

What are Discriminant Functions ?

Consider a classifier built using L functions

$g_i(\bar{x})$, $i = 1, 2, \ldots, L$.

$$\Rightarrow \underbrace{g_1(\bar{x}), g_2(\bar{x}), \ldots, g_L(\bar{x})}_{\text{L Discriminant functions}}$$

The input / feature vector $\bar{x}$ is assigned to class l if

$$g_l(\bar{x}) = \max_{1 \le i \le L} g_i(\bar{x})$$

(i) Assign class l for which discriminant function is maximum.

(ii) Evaluate each discriminant function for $\bar{x}$

(i) $g_i(\bar{x})$, $i = 1, 2, \ldots, L$ and assign the class l corresponding to which the discriminant function is maximum.

If $L = 2$, implies Binary classification.

- These $g_i(\bar{x})$ are termed as Discriminant functions.

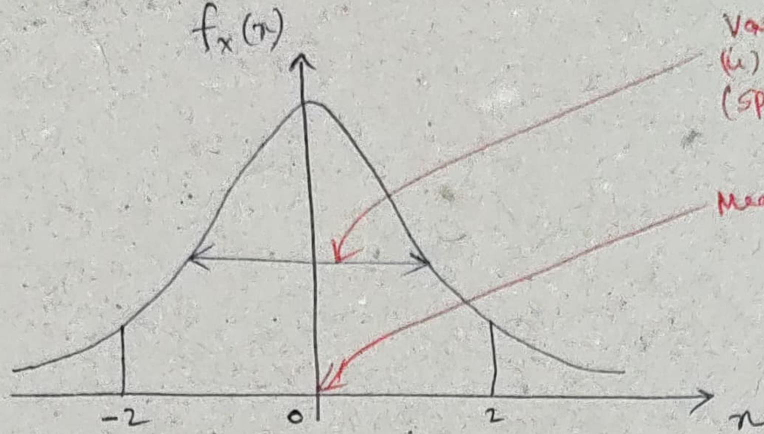
To understand LDA, we start with a Gaussian Density corresponding to a Gaussian Random variable.

Recall, the expression for the Gaussian PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

where, $\mu = E\{X\}$ = Mean

$\sigma^2 = E\{(X-\mu)^2\}$ = Variance.



Variance
(i.e) Spread around the Mean
(Spread is characterized by variance)

Mean, $\mu = 0$

Plot: Gaussian Density Function /
Normal Probability density

Multivariate Gaussian Density

Gaussian Random Vector is a collection of Gaussian Random variable. Recall, the PDF of a Gaussian Random vector

$$\bar{x} = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{\text{N components / N Features}}$$ is given as

Feature vector

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

where, $\bar{\mu} \rightarrow$ Mean, which is a Vector.

$R \rightarrow$ Covariance Matrix.

$|R| \rightarrow$ Determinant of R.

The Mean and covariance matrix are defined as

$$\bar{\mu} = E\{\bar{x}\} = E\left\{\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}\right\} = \begin{bmatrix} E\{x_1\} \\ \vdots \\ E\{x_N\} \end{bmatrix} = \text{Mean}$$

$$R = E\left\{(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\right\} = \text{Covariance Matrix.}$$

Example :

Find multivariate Gaussian PDF, given Mean $\bar{\mu} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\underbrace{\qquad}_{N=2}$

Covariance Matrix, $R = \begin{bmatrix} 7 & 2 \\ 2 & 1 \end{bmatrix}$.

$$|R| = 7 - 4 = 3$$

$$R^{-1} = \frac{1}{3}\begin{bmatrix} 1 & -2 \\ -2 & 7 \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$(\bar{x} - \bar{\mu})^T = \begin{bmatrix} x_1 - 1 & x_2 - 2 \end{bmatrix}$$

$$\bar{x} - \bar{\mu} = \begin{bmatrix} x_1 - 1 \\ x_2 - 2 \end{bmatrix}$$

$$(\bar{x} - \bar{\mu})^T R^{-1} (\bar{x} - \bar{\mu}) = \begin{bmatrix} x_1 - 1 & x_2 - 2 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 1 & -2 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 2 \end{bmatrix}$$

$$= \frac{1}{3}\left(x_1^2 + 7x_2^2 + 6x_1 - 24x_2 - 4x_1x_2 + 21\right)$$

Therefore, the Multivariate Gaussian PDF,

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

$$\boxed{f_{\bar{X}}(\bar{x}) = \frac{1}{\sqrt{12\pi^2}} e^{-\frac{1}{6}\left(x_1^2 + 7x_2^2 + 6x_1 - 24x_2 - 4x_1x_2 + 21\right)}}$$

This is the Multivariate Gaussian PDF for the given example.

Gaussian Discriminant Analysis (GDA)

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

C_0 : Mean $\bar{\mu}_0$ and Covariance R

C_1 : Mean $\bar{\mu}_1$ and Covariance R.

Given a vector $\bar{x}$, the class which has the larger value of the probability density, has the higher likelihood. So, PDF can be used as a likelihood function.

Thus, the likelihoods of two classes are

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

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

Now, choose the class that maximizes the likelihood.

(ii) ML Rule / ML Classifier.

Therefore, choose C_0 if

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

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

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

due to '-ve' sign in the exponent

The discriminant function can be simplified as

$$\text{choose } C_0 : \bar{h}^T (\bar{x} - \tilde{\mu}) \geq 0$$

$$\text{choose } C_1 : \bar{h}^T (\bar{x} - \tilde{\mu}) < 0$$

where, $\tilde{\mu} = \frac{\bar{\mu}_0 + \bar{\mu}_1}{2}$ ← Mid point of both classes

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

$$\Rightarrow C_0 : (\bar{\mu}_0 - \bar{\mu}_1)^T R^{-1} \left(\bar{x} - \frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right) \geq 0$$

$$C_1 : (\bar{\mu}_0 - \bar{\mu}_1)^T R^{-1} \left(\bar{x} - \frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right) < 0$$

aka Linear Classifier.

Thus, the classifier is LINEAR. It is characterized by the hyperplane.

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

$$\Rightarrow a_1 x_1 + a_2 x_2 + \dots + a_n x_n \geq b.$$

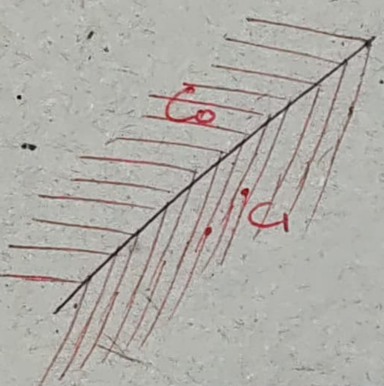
Example:

$$2x_1 + 3x_2 \geq -7 \leftarrow \text{Two Dimensions (LINE)}$$

$$7x_1 - 8x_2 + 17x_3 \geq -2 \leftarrow \text{Three Dimensions (PLANE)}$$

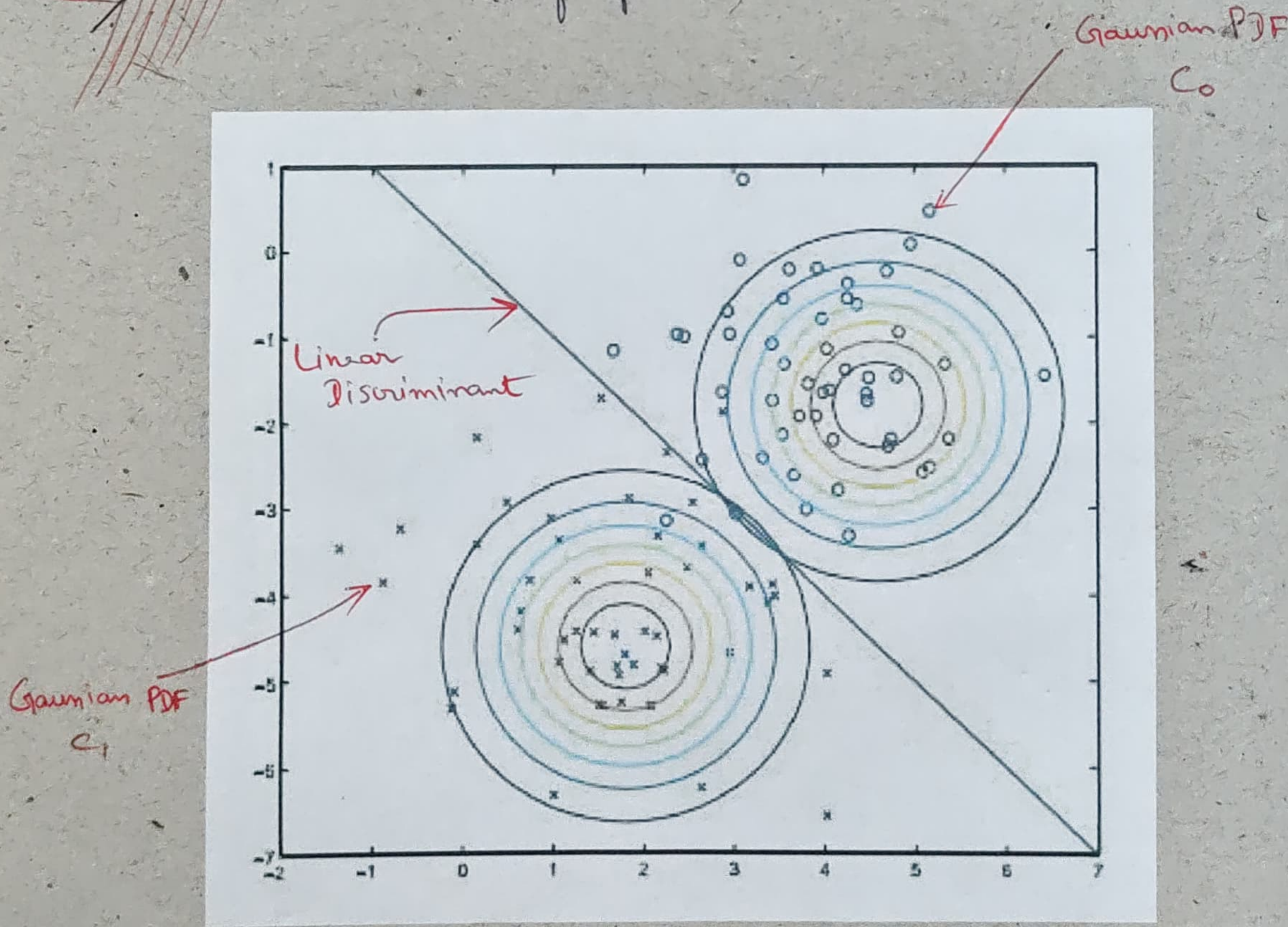
For N Dimensions : HYPERPLANE.

$\bar{h}^T (\bar{x} - \tilde{\mu})$ is Linear. Since the Discriminant function is Linear, hence termed as "Linear Discriminant Analysis".



← Hyperplane.

The Hyperplane divides the space into two equal parts, each of which is a Half space.



Week 5 Session 4

Special case of LDA / GDA

Consider the special case of Covariance Matrix, $R = \sigma^2 I$

(ie) $R \propto$ Identity Matrix

It follows that,

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

$$= \frac{1}{\sigma^2} I (\bar{\mu}_0 - \bar{\mu}_1)$$

Constant

$$\equiv \bar{\mu}_0 - \bar{\mu}_1$$

Recall, the Hyperplane reduces to

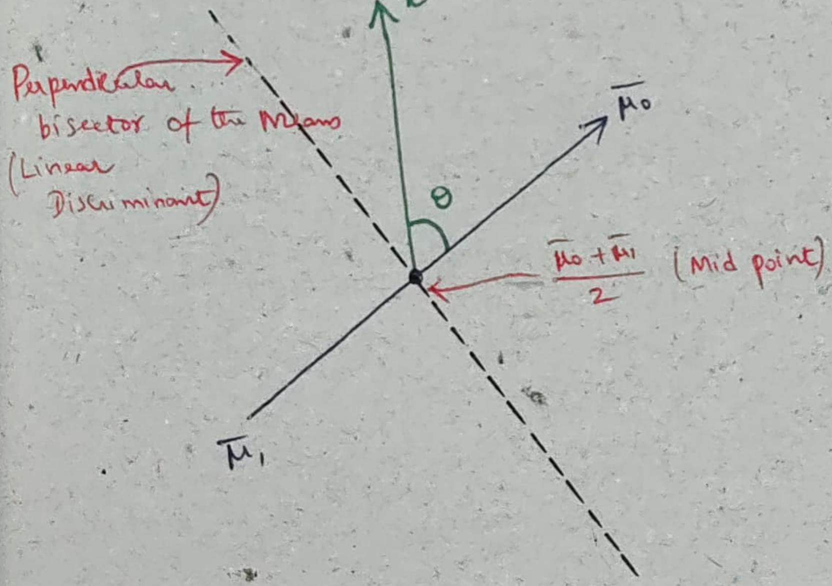
Choose C_0 : $\bar{h}^T(\bar{x} - \tilde{\mu}) \geq 0$

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

where, $\bar{\mu}_0 - \bar{\mu}_1 \rightarrow$ vector from $\bar{\mu}_1$ to $\bar{\mu}_0$

$\bar{x} - \frac{\bar{\mu}_0 + \bar{\mu}_1}{2} \rightarrow$ Vector from $\frac{\bar{\mu}_0 + \bar{\mu}_1}{2}$ to $\bar{x}$

$(\bar{\mu}_0 - \bar{\mu}_1)^T\left(\bar{x} - \frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right) \rightarrow$ Dot product between vectors $(\bar{\mu}_0 - \bar{\mu}_1)$ and $\left(\bar{x} - \frac{\bar{\mu}_0 + \bar{\mu}_1}{2}\right)$.



It can be seen that,

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

when $-90^\circ \leq \theta \leq 90^\circ$

Dot product b/w $\bar{a}$ and $\bar{b}$

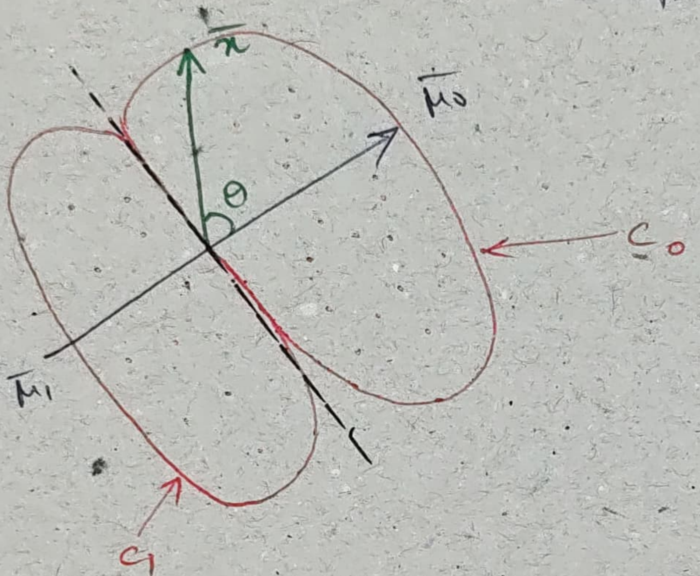
$\Rightarrow \bar{a} \cdot \bar{b}^T \geq 0$

$\Rightarrow \|a\| \|b\| \cos\theta \geq 0$

$\Rightarrow \cos\theta \geq 0$

$\Rightarrow -90^\circ \leq \theta \leq 90^\circ$

(ii) All the points lying on one side of the perpendicular bisector towards the side of $\bar{\mu}_0$ will be classified as C_0. And, all the points that are lying on the other side of the perpendicular bisector will be classified as C_1.



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

<u>Example</u>:

Consider the Gaussian Classification Problem, where the two Classes C_0, C_1 are distributed as

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

What is the Linear Discriminant Classifier?

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

$$R = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/4 \end{bmatrix}, \quad R^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

- Calculate $\bar{h}$

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

$$= \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \left(\begin{bmatrix} -1 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)$$

$$= \begin{bmatrix} -6 \\ 4 \end{bmatrix}.$$

- Further, calculate $\tilde{\mu}$.

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

$$= \frac{1}{2}\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right)$$

$$= \frac{1}{2}\begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

- The classifier chooses C_0 if

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

$$\Rightarrow \begin{bmatrix} -6 & 4 \end{bmatrix} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 3 \end{bmatrix} \right) \geq 0$$

$$\Rightarrow -6x_1 + 4x_2 + 3 - 6 \geq 0$$

$$\Rightarrow -6x_1 + 4x_2 \geq 3.$$

Thus, the Linear Discriminant classifier is given by

choose C_0 if $6x_1 - 4x_2 \leq -3$

choose C_1 if $6x_1 - 4x_2 > -3$.