LDA (Linear Discriminant Analysis) → objective : LDA aims to find a linear combination of features that best seprates two classes of objects or event.

↓

Supervised machine learning technique

↓

Classification (single class & multi-class) & Dimensionality Reduction

↓

developed By R.A Fisher

It projects the data onto a lower dimensional space with good class-separability.

reduce complexity of equn

$y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + w_0$ ← noise.

$w_1, w_2, \dots, w_n$: associated weights

$x_1, x_2, \dots, x_n$: input features

- Assumptions-
- Normality → The data follows normal (or gaussian) distribution
- Homeginity of variances → The covariance of every classes is same.
- Independence → The observations are assumed to be independently sampled

- Aim → ① Suppose if we have any point x_i, dimension D

$$\begin{bmatrix} x_i^1 \\ x_i^2 \\ x_i^3 \\ \vdots \\ x_i^D \end{bmatrix} \qquad \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_D \end{bmatrix}$$

② And any plane →

$$w_1 x^1 + w_2 x^2 + w_3 x^3 + \dots + w_D x^D = 0$$

passing through origin.

③ So projection of any point

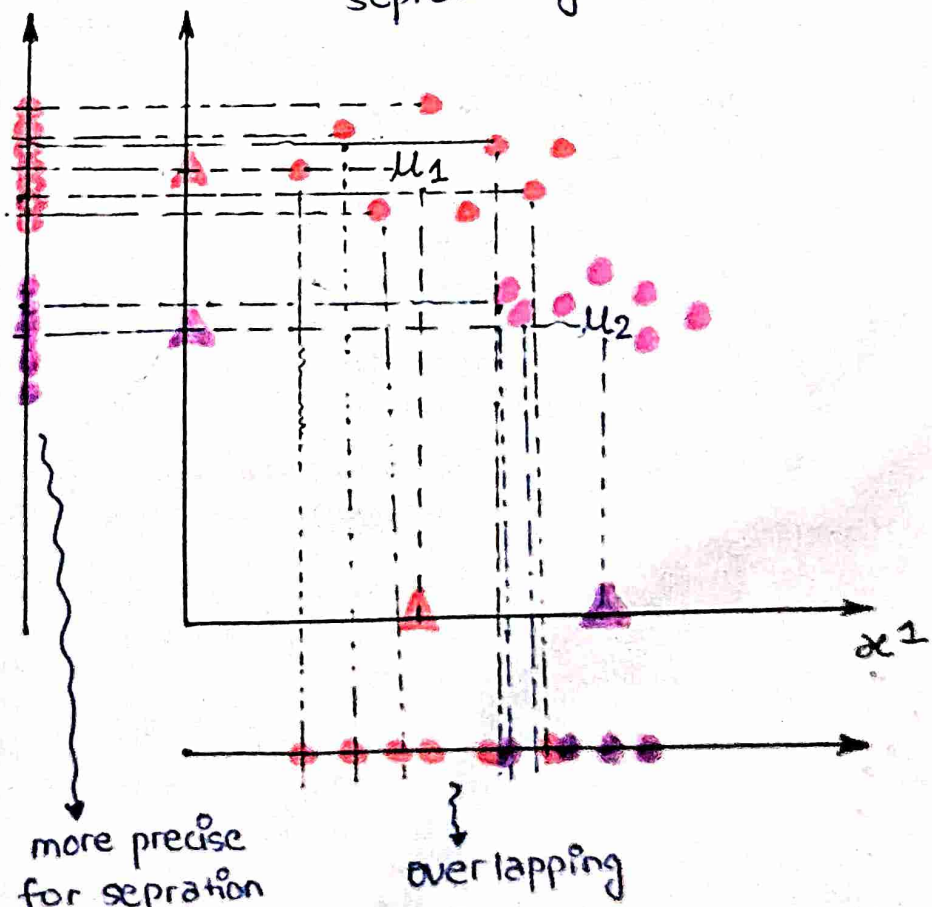
$w^T x_i$

projection of $\vec{B}$ on $\vec{A}$ is (dot-product)

$$[w_1\ w_2 \dots w_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

The main task in LDA is to project the data onto a line or plane such that the distance between the points is maximized-

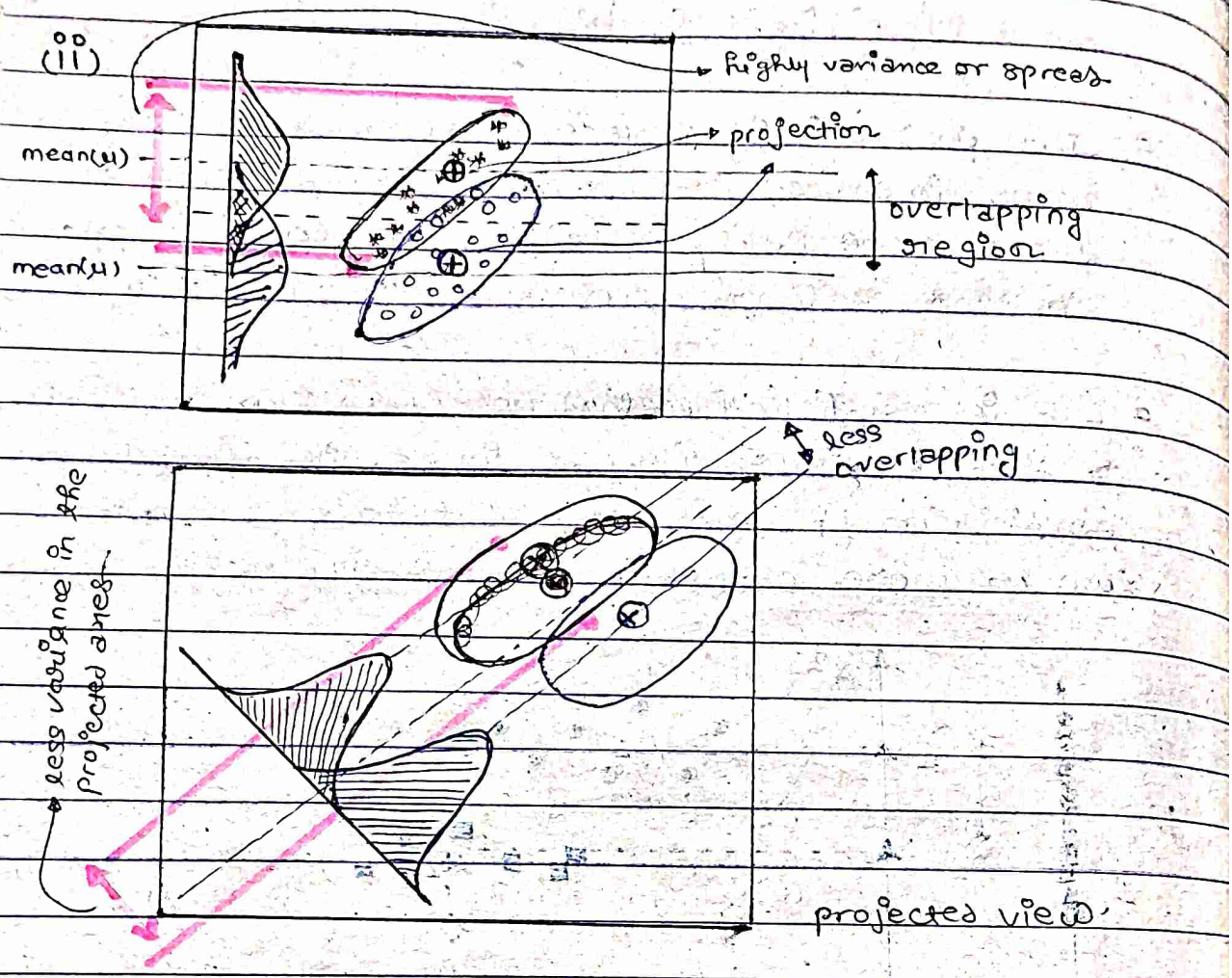
★ maximize class seprability.



μ → original data class mean values

$\tilde{\mu}$ → projected data class mean values-

Thus we want that projection mean location of different classes should be far away from each other.



So, for better projection we need ⟹

- The variance of the projected points of the different classes should be minimum.

Outcome

⇓

1) $\max [\tilde{\mu}_1 - \tilde{\mu}_0]^2$ $\tilde{\mu}_1$ ⇒ projected mean of class 1 points.

$\tilde{\mu}_0$ ⇒ projected mean of class 2 points.

2) $\min \tilde{S}_1^2$ ⇒ variance of projected points on class 1

$\tilde{S}_0$ ⇒ —||— class 0.

So, we need to find ω vector $\omega \equiv \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_B \end{bmatrix}$

Objective function $\Rightarrow \max \left[\frac{(\tilde{\mu}_1 - \tilde{\mu}_0)^2}{s_0^2 + s_1^2} \right]$

Hence, LDA
(Fisher Linear Discriminant)

To create this new axis & reduce dimensionality —

maximize the dist b/w projected mean of two classes
reduce the variance b/w projection ← ←

→ How to find the projection of any point on a

→ Algorithm:

x_i^D → $l \times D$ matrix here (6×2)

dim 1	dim 2	class
3 x_1^1	4 x_1^2	1
2 x_2^1	3 x_2^2	1
1 x_3^1	1 x_3^2	1
5	5	0
6	6	0
7	7	0
x_N^{D-1}	x_N^D	

N1 points

No points

6/3

8/3

18/3

$$\mu_1 \Rightarrow \begin{bmatrix} \frac{1}{N_1}\sum_{i=1}^{N} x_i^2 \\ \frac{1}{N_1}\sum_{i=1}^{N} x_i^{(2)2} \\ \vdots \\ \frac{1}{N_1}\sum_{i=1}^{N} x_i^{(D)2} \end{bmatrix} \quad D \times 1 \text{ matrix}$$

$$\tilde{\mu}_1 = \omega^T \mu_1 \quad (1 \times D \text{ and } D \times 1 \rightarrow 1 \times 1)$$

$$\mu_1 \Rightarrow \begin{bmatrix} 2 \\ 8/3 \end{bmatrix}_{2\times 1} ; \quad \mu_0 \Rightarrow \begin{bmatrix} 6 \\ 6 \end{bmatrix} ;$$

$$\tilde{\mu}_1 - \tilde{\mu}_0 = \omega^T (\mu_1 - \mu_0)$$

gap b/w the mean location in original feature space.

$(\tilde{\mu}_1 - \tilde{\mu}_0)^2 \Rightarrow$

← $\omega\omega^T \rightarrow \frac{1}{2}\|\omega\| \rightarrow \sqrt{\omega_1^2+\omega_2^2+\cdots+\omega_D^2}$

$(\omega^T(\mu_1-\mu_0))(\omega^T(\mu_1-\mu_0))^T$ ⤵ $(AB)^T = B^T A^T$

$(\omega^T)_{1\times D} (\mu_1-\mu_0)_{D\times 1} (\mu_1-\mu_0)^T_{1\times d} \omega_{d\times 1}$

original data class mean values available if training data is known.

$$(\tilde{\mu}_1 - \tilde{\mu}_0)^2 = \omega^T S_B \omega$$

S_B → scatter-matrix between class-

between-class scatter matrix (SB) – measures the scatter B/w the means of different classes.

5D case

$\Rightarrow (\mu_1-\mu_0)(\mu_1-\mu_0)^T$

$$\begin{bmatrix}-4\\-10/3\end{bmatrix}\begin{bmatrix}-4 & -10/3\end{bmatrix} = \begin{bmatrix}16 & 40/3\\ 40/3 & 100/9\end{bmatrix}$$

Class 1 :

dim 1 : 3 2 1

dim 2 : 4 3 1 $\equiv X_1$

seperate N_1 & N_0 points.

$S_1^2 \rightarrow$ variance of projected points of class 1.

$$\frac{1}{N_1}\sum_{i=1}^{N_1}(w^T x_i - \tilde{\mu_1})^2$$

↑ only known parameter

$\Rightarrow (w^T x_i - w^T \mu_1)$ → known

$\Rightarrow w^T(x_i - \mu_1)$.

$$X_1 \equiv \begin{vmatrix} x_1^1-\mu & x_2^1 & \cdots & x_{N_1}^1 \\ x_1^2-\mu & x_2^2 & \cdots & x_{N_1}^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_1^D-\mu & x_3^D & \cdots & x_{N_1}^D \end{vmatrix} \Rightarrow$$

$w^T(x_1-\mu_1)$ → $D \times N_1$

↓ $1 \times D$.

$\Rightarrow [w^T(x_1-\mu_1)\ w^T(x_2-\mu_2)\ \cdots\ (w^T(x_n-\mu_n)]$

$S_1^2 \Rightarrow [w^T(X_1-\mu_1))(w^T(X_1-\mu_1))^T$

similar

$S_0^2 \Rightarrow [w^T(X_0-\mu_0)][w^T(X_0-\mu_0)]^T$

$S_1^2 + S_0^2 \Rightarrow w^T(X_1-\mu_1)(X_1-\mu_1)^T w^T + w^T(X_0-\mu_0)(X_0-\mu_0)^T w^T$

$\Rightarrow w^T[(X_1-\mu_1)(X_1-\mu_1)^T + (X_0-\mu_0)(X_0-\mu_0)^T]w$

$\Downarrow$

S_W

(Scatter matrix within-class)

measure the scatter (variance) within each class.

→ from eqn – (1),

$$\frac{\partial A}{\partial w} = \frac{A}{B}\frac{\partial B}{\partial w}$$

$$\left[S_B w = \left(\frac{A}{B}\right) S_w w \right]$$

(A/B → constant)

$$S_w^{-1} S_B w = \left(\frac{A}{B}\right) w$$

(w, w → same; A/B → Consto)

$Ax = \lambda x$ where x is eigen vector & λ is eigen value

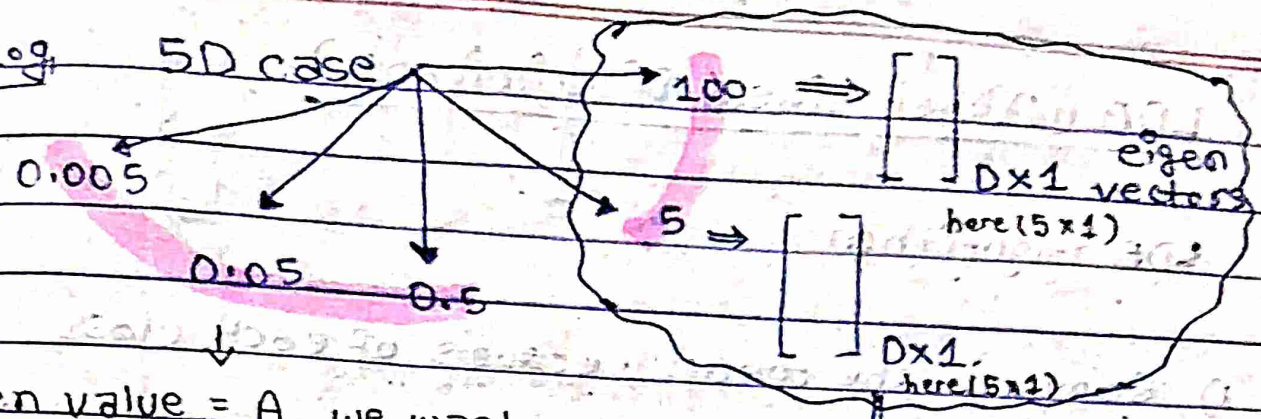
$$\max \frac{w^T S_B w}{w^T S_w w} \Rightarrow S_w^{-1} S_B w = \frac{A}{B} w$$

($S_w^{-1} S_B$ → known; w → eigen vector; $\frac{A}{B}$ → eigen value)

- we have data of D dimension
- S_w, S_B can be found → D × D (size)
- So $\underbrace{S_w^{-1} S_B}_{D \times D} w = \frac{A}{B} w$ → So, we can get D nos of eigen values

we will find eigen vector for each & every corresponding eigen value.

e.g. 5D case



eigen value = $\frac{A}{B}$ we want to maximize, Hence, doesn't take lower values.

Significant or important.

matrix containg eigen vector from both

$W \rightarrow$ D x 2

Original datapoint $(x_i) \Rightarrow$ map $\rightarrow w^T x_i$

D x 1 (here (5x1)) $\rightarrow$ 2 x D (here (2x5)) $\cdot$ D x 1

corresponding

dimensionality reduction

2 x 1

$S_1^2 + S_2^2 \Rightarrow [(X_1 - \mu_1)(X_1 - \mu_1)^T + (X_0 - \mu_0)(X_0 - \mu_0)^T]$

$$\begin{bmatrix} 1 & 0 & -1 \\ 4/3 & 1/3 & -5/3 \end{bmatrix} \begin{bmatrix} 1 & 4/3 \\ 0 & 1/3 \\ -1 & -5/3 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$\Downarrow \qquad \Downarrow$$

$$\begin{bmatrix} 2 & 3 \\ 3 & 14/3 \end{bmatrix} \qquad \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 5 \\ 5 & 20/3 \end{bmatrix}$$

$\Downarrow$

S_W.

$$J(V) \Rightarrow \max \left(\frac{w^T S_B w}{w^T S_W w} \right)$$

$\rightsquigarrow \max [\mu_1 - \mu_0]^2 \Rightarrow$ maximize the dist b/w projected mean of two class

$\rightsquigarrow \min [S_1^2 + S_0^2] \Rightarrow$ reduce the variance b/w projected mean of two class

$$\frac{d}{dw}\left(\frac{w^T S_B w}{w^T S_W w} \right) \rightarrow \frac{d}{dw}\left(w^T S_B w \right) \Rightarrow 2 S_B w$$

$$\rightarrow \frac{d}{dw}\left(w^T S_W w \right) \Rightarrow 2 S_W w$$

So, $S_W^{-1} S_B \Rightarrow \begin{bmatrix} 16 & 40/3 \\ 40/3 & 100/9 \end{bmatrix} * \frac{1}{40 \times \frac{20}{3} - 5 \times 5} \begin{bmatrix} 20/3 & -5 \\ -5 & 4 \end{bmatrix}$

$(S_W^{-1} S_B)\omega = \lambda\omega$

$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \Rightarrow \frac{1}{ad - bc} \begin{matrix} d & -b \\ -c & a \end{matrix}$

$\frac{3}{5} \begin{bmatrix} 40 & 100/3 \\ -\frac{80}{3} & -200/9 \end{bmatrix} \omega = \lambda\omega$

↓ eigen value &
↓ eigen vector

$AX = \lambda X$

$AX - \lambda X = 0$

$(A - \lambda I)X = 0$ → eigen vector X are calculated by solving this eqn

for finding eigen value of A

$\Rightarrow |A - \lambda I| = 0$

LDA (Fisher Linear Discriminant)

LDA Algorithm

1) Compute the mean vectors of each class

2) Compute the scatter matrices S_W & S_B.

3) Compute the eigen vector & eigen values for the matrix $S_W^{-1} S_B$.

4) Sort the eigen vectors by decreasing eigen values and choose the top k eigen vectors to form a matrix W.

5) Transform the samples to new space using W.

Testing

So first transform the point x_t $\longrightarrow$ $(w^T x_t)$

Now, find distance of transformed point from the $\tilde{\mu}_1$, $\tilde{\mu}_0$.

So, if dist to from $\tilde{\mu_1}$ < dist from μ_0

$\hookrightarrow$ Class 1

else class 0.

Questions

① What is the primary goal of LDA?

(a) To maximize the variance of data → PCA
(b) To minimize the covariance b/w classes
(c) To maximize the seprability b/w different classes ✓✓
(d) To find non-linear boundary b/w classes → classification & dimensionality reduction

② Assumption of LDA?

③ In LDA, the within-class scatter Matrix (Sw) measures?

(a) The variance within each class → $Sw = S_1^2 + S_0^2$
(b) The variance b/w the classes. $(x_1 - \mu_1)(x_1 - \mu_1)^T + (x_0 - \mu_0)(x_0 - \mu_0)^T$

④ Key diff b/w LDA → maximizes class seprability & PCA → maximizes variance.

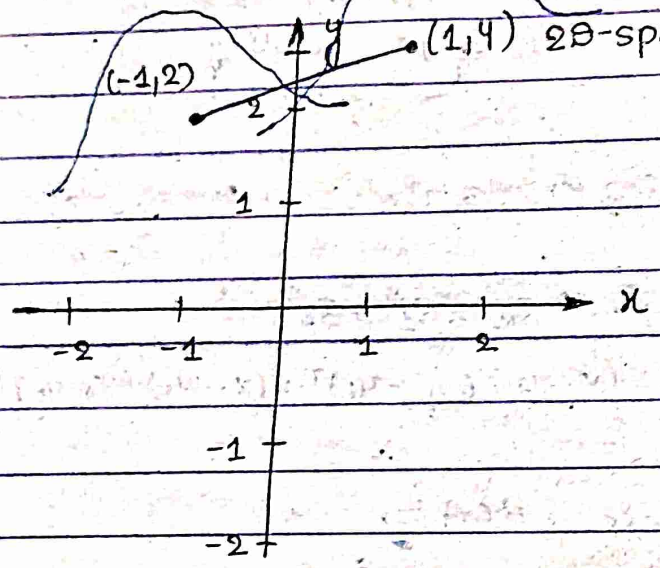
⑤ In context to LDA, primary goal of eigen value & eigen vector
Transform the data into new space with reduced dimensions

⑥ LDA is particularly effective when —

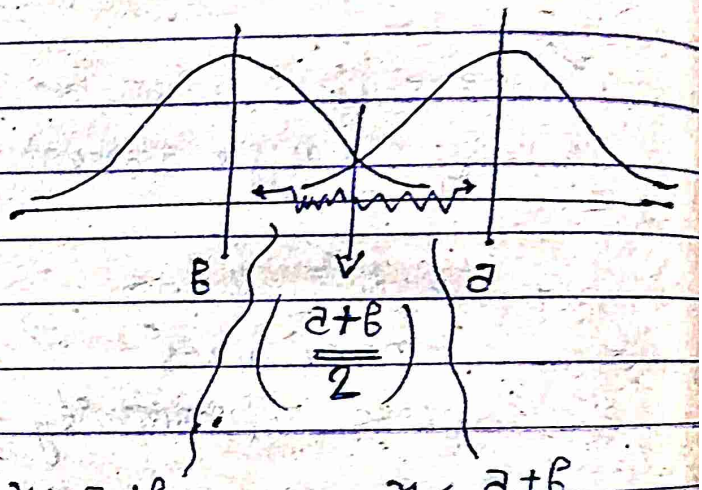
(a) The classes are linearly seprable ✓✓
(b) The classes are not linearly seprable
(c) There are more features than samples
(d) The data follow a non-gaussian distribution ✗

Q) Consider the case where two classes follows gaussian distribution which are centered at (-1,2) & (1,4) and have identity covariance matrix. which of the following is the seprating decision Boundary?

(a) $y-x=3$ (b) $x+y=3$ (c) $x+y=6$ (d) (b) & (c) are possible

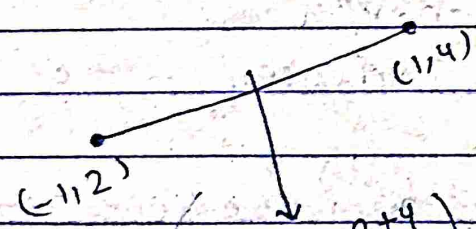


दोनों का gaussian distribution (skewness) same है



$x > \frac{a+b}{2}$ Class 1

$x < \frac{a+b}{2}$ Class 0



$\left(\frac{-1+1}{2}, \frac{2+4}{2}\right)$

$\equiv (0,3)$

classifier should be perpendicular to this line passing through this point (0,3)

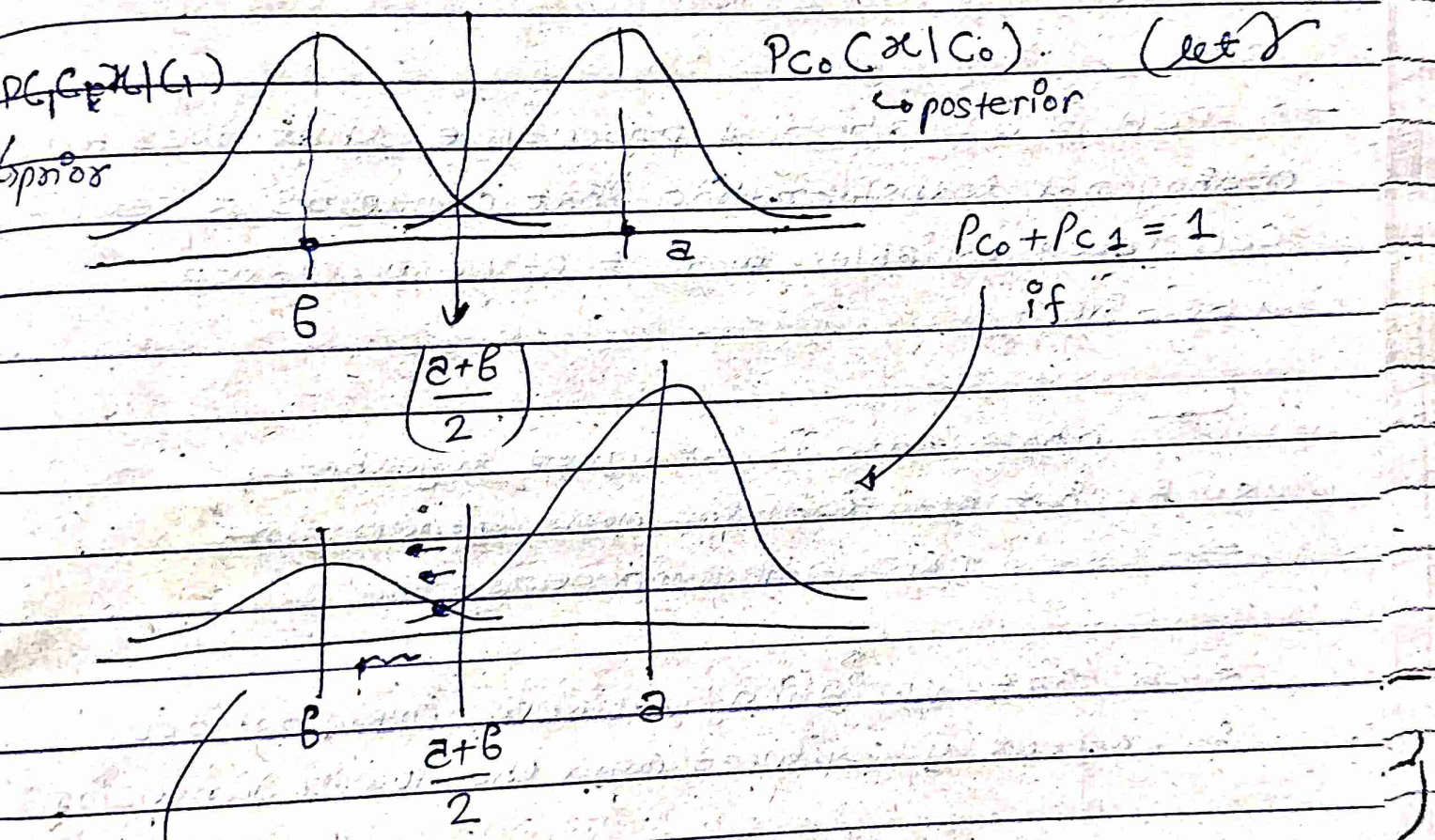
→ so slope = -1

$y - y_1 = m(x - x_1)$ (0, 3)

→ so line : $y = -x + 3$

1. In LDA, what is the effect of class priors on decision Boundary?

(a) class priors have no effect on the decision boundary
(b) Higher prior shift decision boundary towards that class
(c) Lower prior shift the decision boundary towards that class
(d) class prior make the decision boundary curved



Prior lower decision boundary shift toward that class