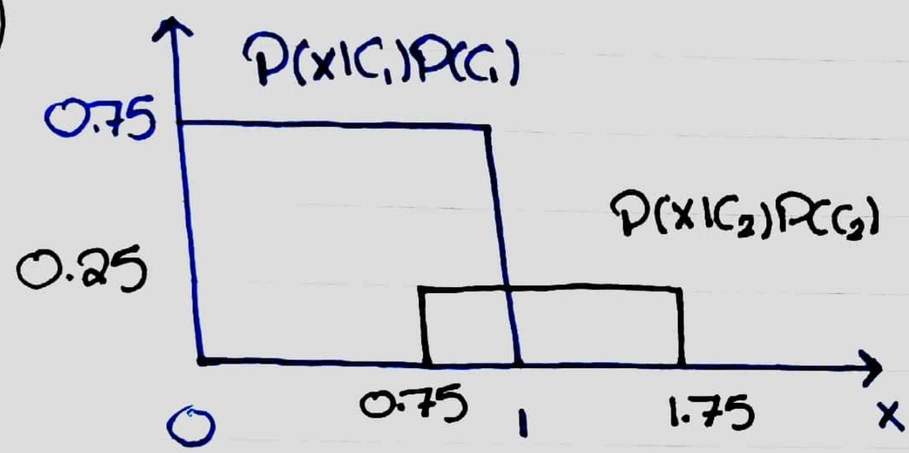
NN Sheet 3

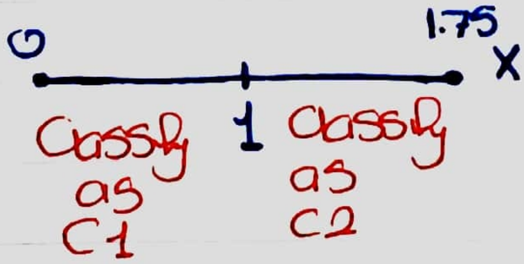
1)



- $P(x|C_2)P(C_2) > P(x|C_1)P(C_1) \quad \forall x > 1$ and
  $P(x|C_2)P(C_2) < P(x|C_1)P(C_1) \quad \forall x < 1$

Hence,

decision boundary is $x = 1$



.at $x=1$ and all other x its a tie

$$P(\text{error}) = 1 - P(\text{Correct}) = 1 - \underbrace{\int_{-\infty}^{\infty} \max_{1 \le i \le 2} P(x|C_i)P(C_i)}_{\text{area under Outer envelope}}$$

Hence, $P(\text{error}) = \text{OverlaP area}$

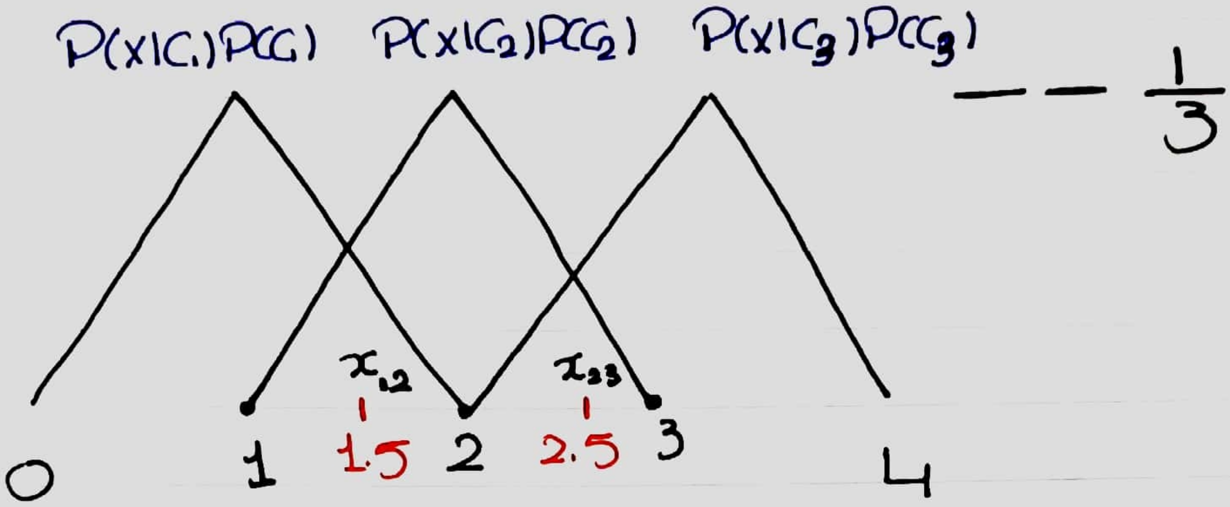
$= 0.25(1 - 0.75) = 0.0625$

.Can alsoCheck by

$$P(\text{error}) = P(x>1|C_1)\overset{P(C_1)}{} + P(x<1|C_2)P(C_2)$$

$0 \quad + \quad 0.0625$

2.

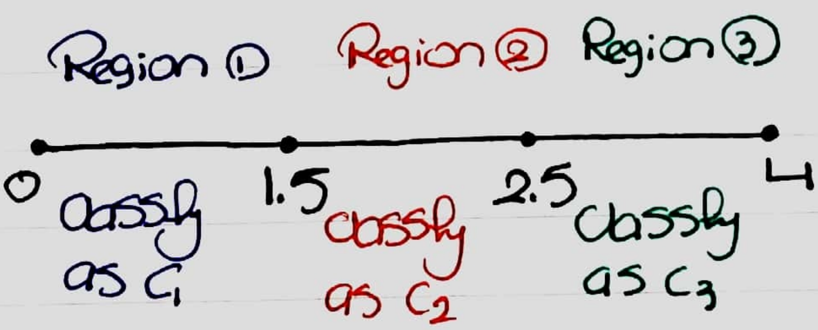


→ Since there are 3 classes, need 3 decision regions ⟷ 2 decision boundaries in 1D

- By symmetry, $P(x|C_1)P(C_1) = P(x|C_2)P(C_2)$ whenever $x = 1.5$ $\left(\frac{1+2}{3}\right)$

- Likewise, $P(x|C_2)P(C_2) = P(x|C_3)P(C_3)$ whenever $x = 2.5$

• Can also do it via straight line eqn.s



← Decision Regions

• Treat $x = 1.5$, $x = 2.5$, $x \in$ Jaue strata tie

$P(\text{correct})$ = Area under envelope

$P = \frac{1}{3}x - \frac{1}{3}$ & $P = -\frac{1}{3}x + \frac{2}{3}$

Hence, $P = \frac{1}{6}$

• also by symmetry

$= 3 \times \triangle - 2 \times \triangle$

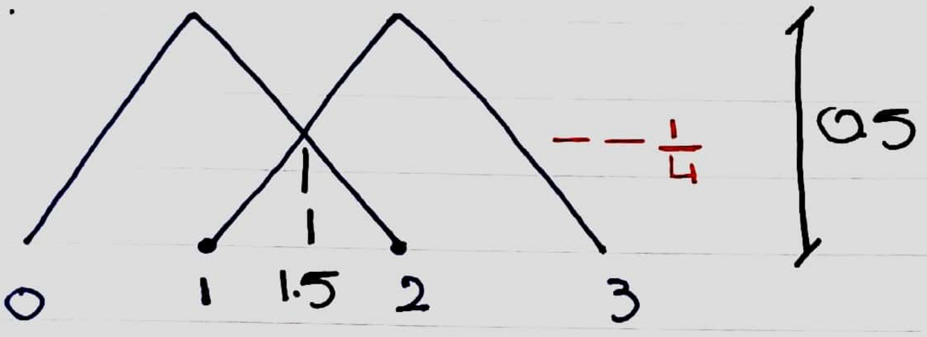
$= 3 \times \left(\frac{1}{2} \times 2 \times \frac{1}{3}\right) - 2 \times \left(\frac{1}{2} \times 1 \times \frac{1}{6}\right)$

$= \frac{5}{6}$

$P(\text{error}) = 1 - P(\text{correct}) = \frac{1}{6} = \underset{(2\Delta)}{\text{total overlap}}$

• Does not always hold.

3. $P(x|C_1)P(C_1)$ $P(x|C_2)P(C_2)$



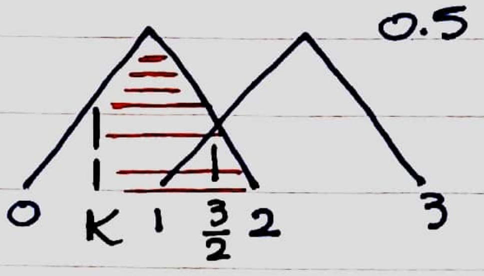
$P(\text{error}) = \Delta + \Delta$
$= \frac{1}{2} \times 1 \times \frac{1}{4}$
$= \frac{1}{8}$

• Bayes decision rule has $x = 1.5$
• Suppose another decision rule comes up with a decision boundary $x = K$ and $K \neq 1.5$

$P(\text{error}) = P(x > K|C_1)P(C_1) + P(x < K|C_2)P(C_1)$

to the right under the 1st Δ + to the left under the left Δ

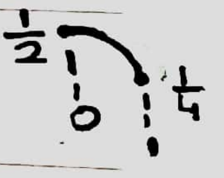
• Suppose $K \in [0,1]$



$0.5x$

$P(\text{error}) = \underset{0 \quad 2}{\overset{\frac{1}{2}}{\Delta}} - \underset{K}{\overset{0.5K}{\Delta}} = \frac{1}{2} \times 2 \times \frac{1}{2} - \frac{1}{2} \times K \times 0.5K$

$= \frac{1}{2}\left(1 - \frac{K^2}{2}\right)$

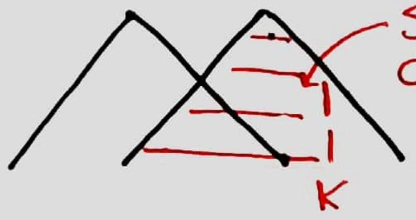


$P(\text{error})\Big|_{\min\,(K=1)} = \frac{1}{2}\left(1 - \frac{1^2}{2}\right) = \frac{1}{4}$

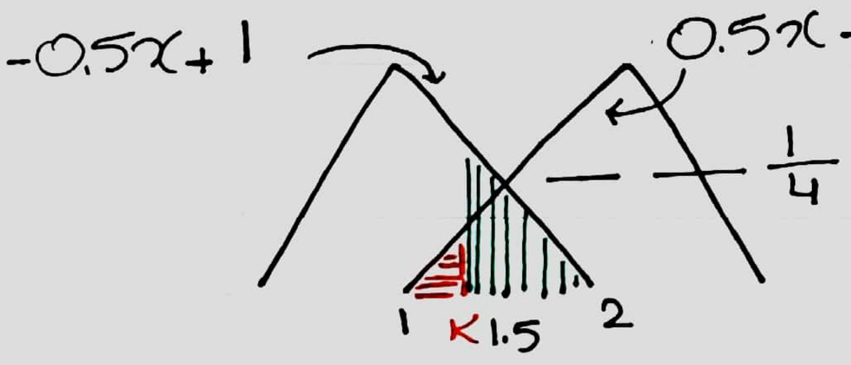
• $P(\text{error}) \in \left[\frac{1}{4}, \frac{1}{2}\right]$

(worse than bayes)

• The case is also clearly symmetric with $K \in [2,3]$



Now if $K \in (1, 1.5]$ but $K \neq 1.5$ (or $K \in (1.5, 2)$)
by symmetry



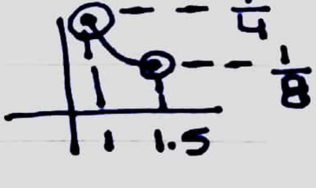
P(error) = to the left under 2nd Δ + to the right under 1st Δ

$$= \frac{1}{2} \times (K-1) \times (0.5K - 0.5)$$

$$+ \frac{1}{2} \times (2-K) \times (-0.5K + 1)$$

$$= 0.5(K^2 - 3K + 2.5)$$

$$= 0.5((K-1.5)^2 + 0.25)$$



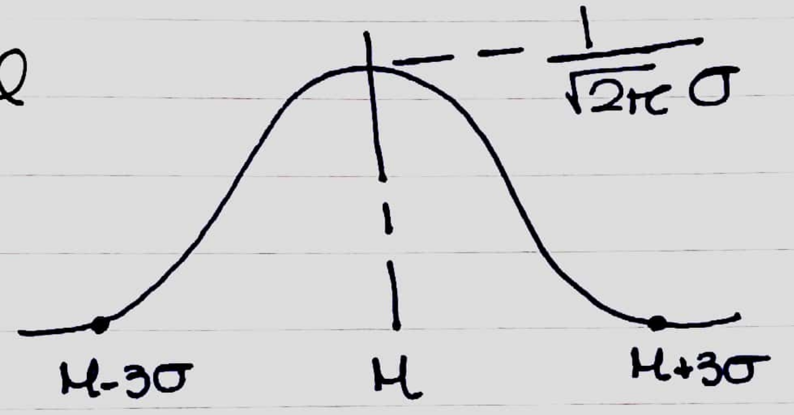
• Clearly have $P(\text{error}) \in (\frac{1}{8}, \frac{1}{4})$ for $K \in (1, 1.5)$
(or the symmetric case $K \in (1.5, 2)$)

Have shown that for all values of K where $K \neq 1.5$
$P(\text{error}) \in \left(\frac{1}{8}, \frac{1}{2}\right]$

↳ for $K > 3$ or $K < 0$
Clearly have $P(\text{error}) = \frac{1}{2}$

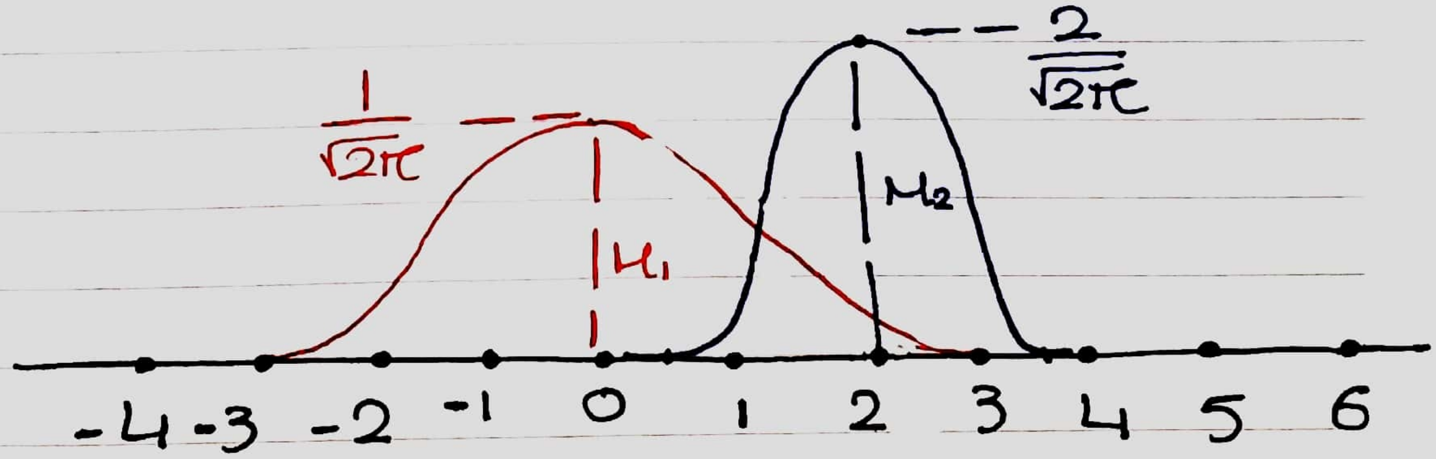
Hence, any other classification rule will result in higher probability of error.

4. Recall



99.73% area within $\mu \mp 3\sigma$

Let $\mu = 0$, $\sigma = 1$ and $\mu = 2$, $\sigma = 0.5$



- $P(\text{error})$ = overlap area
  → Clearly decreases when μ_1 gets farther apart from μ_2 (less overlap)
  → Decreases as well if any of the σs decrease

$P(\text{error}) = P(C_1) \times$ + $P(C_1) \times$

$\int_{-k}^{\infty} P_1(x)\,dx$ $\qquad$ $\int_{-\infty}^{k} P_2(x)\,dx$

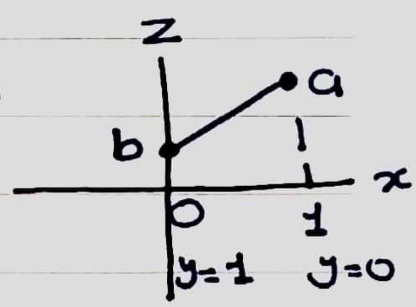
. The two integrals are const, minimize
a, b

$P(C_1)$ $\quad$ $P(C_2)$

$$ax + by \text{ such that } x + y = 1$$

let $a \geq b$ w.l.o.g

$$z = ax + (1-x)b = (a-b)x + b$$



b

- minimized at $x=0$ and maximized when $x=1$

$b < (a+b)$

At $x = y = 0.5$, $z = \frac{1}{2}(a+b)$ (nothing super interesting happens but P(error) decreases as the prior probability of the class having more area in the overlap decreases & vice versa)

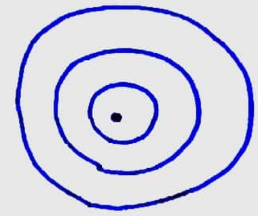
#should be intuitive

Can say that if it was > 0.5 then it will decrease as Prob.s get closer to 0.5 (but it can get better)

$a > \frac{a}{2} + \frac{b}{2}$

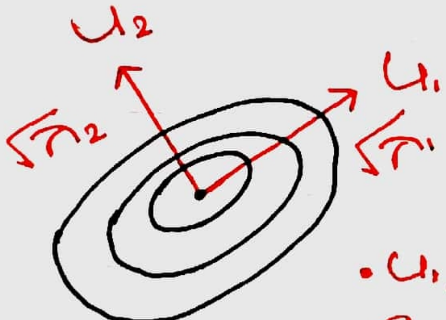
5.

Top View



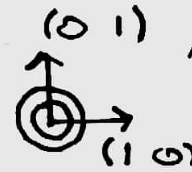
Same PDF value for each circle (symmetry)

• Center at (μ_1, μ_2)

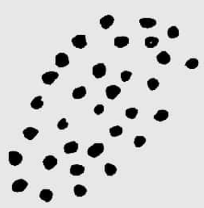


Generally

• u_1, u_2 are eigenvectors of the Covariance matrix when $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ we have



This would correspond to a dataset such as



$$* \; P(\underline{x}|C_i) = \frac{e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})}}{(2\pi)^{N/2} |\Sigma|^{0.5}}$$

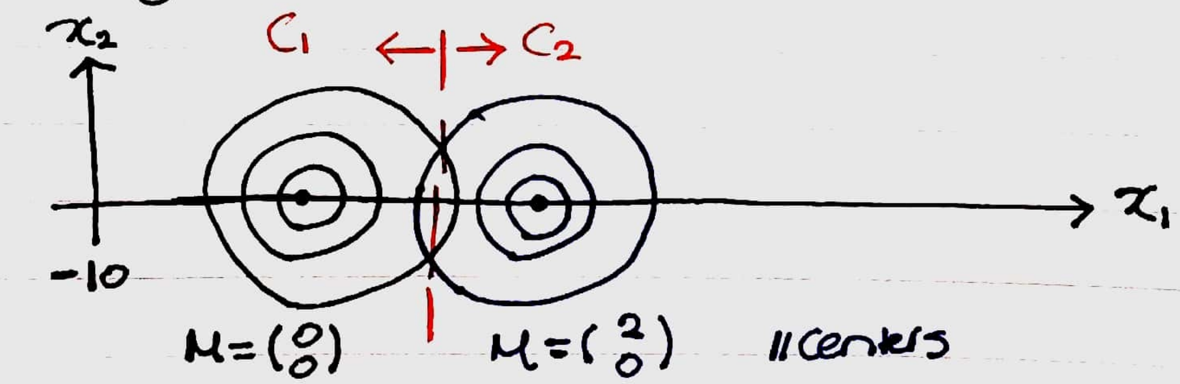
$\underline{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ C_1

C_2 $\underline{\mu}_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

$$P(\underline{x}|C_1) = \frac{e^{-\frac{1}{2}x^T x}}{(2\pi)^{N/2}}$$

$$P(\underline{x}|C_2) = \frac{e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_2)^T (x-\underline{\mu}_2)}}{(2\pi)^{N/2}}$$

To get decision regions



Can confirm the boundary by assuming equal Prior Probabilities and $P(\underline{x}|C_1)P(C_1) = P(\underline{x}|C_2)P(C_2)$ which yields

$(x_1, x_2)^t = \underline{x}$

$$-\frac{1}{2}\underline{x}^T \underline{x} = -\frac{1}{2}(\underline{x}-\underline{\mu}_2)^T(\underline{x}-\underline{\mu}_2)$$

$$x_1^2 + x_2^2 = (x_1-2)^2 + x_2^2$$

$$x_1^2 + x_2^2 = x_1^2 - 4x_1 + 4 + x_2^2$$

$$x_1 = 1$$

- The two decision regions are

$C_1 \rightarrow \underbrace{x_1 < 1}_{A_1}$ and $C_2 \rightarrow \underbrace{x_1 > 1}_{A_2}$

• Called the regions A_1, A_2

$$P(\text{Correct}) = P(C_1)\,P(x_1 < 1 | C_1) + P(C_2)P(x_1 > 1 | C_2)$$

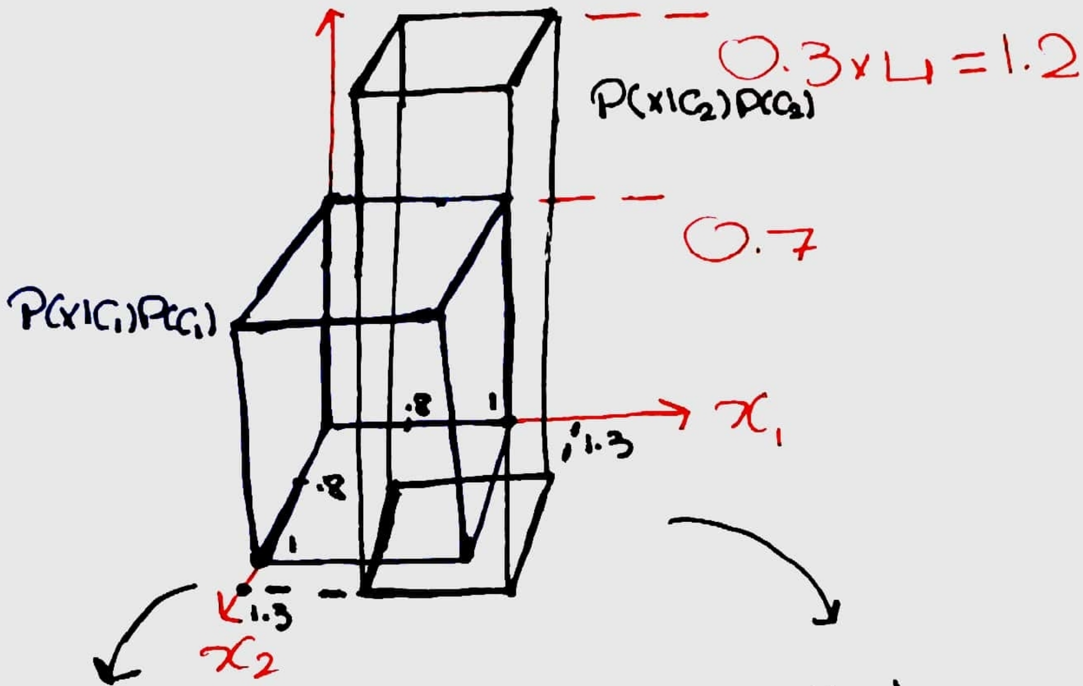
$$= P(C_1) \iint_{\underline{x} \in A_1} \frac{e^{-(x_1^2+x_2^2)/2}}{2\pi}\, dx_1\, dx_2$$

$$+ \iint_{\underline{x} \in A_2} \frac{e^{-((x_1-2)^2+x_2^2)/2}}{2\pi}\, dx_1\, dx_2 \,.\, P(C_2)$$

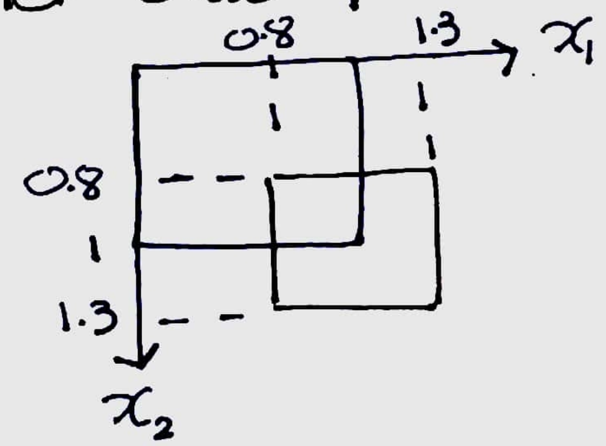
- $\iint_{\underline{x} \in A_1} \longleftrightarrow \int_{x_2=-\infty}^{x_2=\infty} \int_{x_1=-\infty}^{x_1=1}$
- $\iint_{\underline{x} \in A_2} \longleftrightarrow \int_{-\infty}^{\infty} \int_{-\infty}^{x_1}$

$\Bigg\}$ Now rewrite P(Correct)

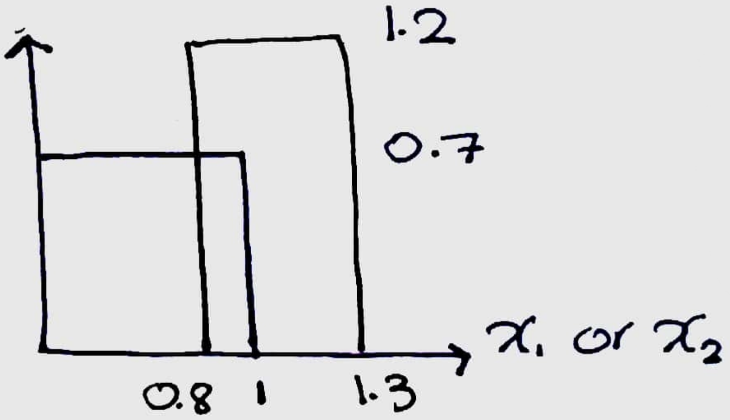
7.



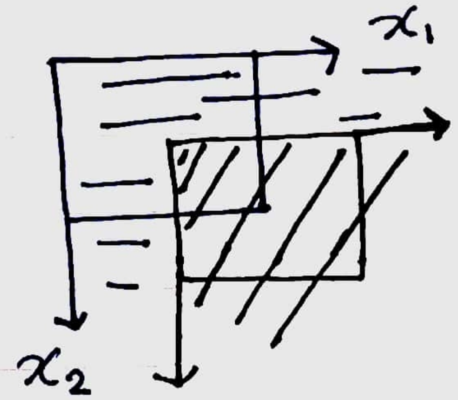
TOP View



Side view



→ Bayes classification rule



$x_1 \geq 0.8$ & $x_2 \geq 0.8 \rightarrow C2$
$x_1 < 0.8$ & $x_1 < 0.8 \rightarrow C1$

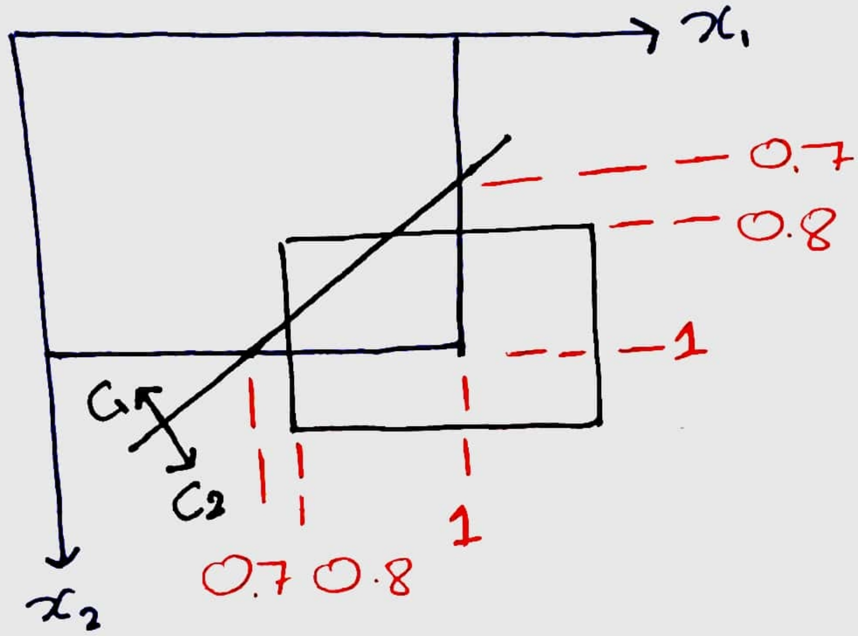
(higher
$P(C_i)P(\underline{x}|C_i)$
is assigned the region)

• when equal → arbitrarily break tie

Linear Classifier

$$f(\underline{X}) = x_1 + x_2 - 1.7 \begin{cases} < 0 \rightarrow C_1 \\ > 0 \rightarrow C_2 \end{cases}$$

$(x_1 + x_2 - 1.7 < 0)$
$\rightarrow x_2 < 1.7 - x_1$



P(error) = Volume under blue cuboid in C2 + Volume under black cuboid in C1

$$= 0.7 \times \underbrace{\frac{1}{2}\, 0.3 \times 0.3}_{\Delta}$$
$$+ \; 1.2 \times \underbrace{\frac{1}{2} \times 0.1 \times 0.1}_{\Delta} = 0.0375$$

NN Sheet 4

1)

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$$

- To get eigen values

$$\det\left(\begin{pmatrix} 2-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix}\right) = 0$$

$$(\lambda-2)(\lambda-3) - 1 = 0$$

$$\lambda^2 - 5\lambda + 5 = 0$$

- $\lambda_1 = 3.618$
  $\lambda_2 = 1.382$

- To get eigenvectors V_1, V_2

$$\begin{pmatrix} 2-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix} \underbrace{\begin{pmatrix} K_1 \\ K_2 \end{pmatrix}}_{\hookrightarrow V_1} = 0$$

i.e. need to find the matrix's null space

$\lambda_1 = 3.618$:

$$\begin{pmatrix} -1.618 & 1 \\ 1 & -0.618 \end{pmatrix} \xrightarrow{r_1 \times \frac{1}{1.618} + r_2} \begin{pmatrix} -1.618 & 1 \\ 0 & 0 \end{pmatrix}$$

Thus, $-1.618 K_1 + K_2 = 0 \rightarrow$ let $K_2 = t$ then
$K_1 = 0.618t$

$\rightarrow$ Null space is $\begin{pmatrix} 0.618 \\ 1 \end{pmatrix} t$

- any t works and we get eigenvector $\rightarrow t=1$

Likewise for λ_2,

$$\begin{pmatrix} 0.618 & 1 \\ 1 & 1.618 \end{pmatrix} \xrightarrow[+r_2]{r_1 \times -0.618^{-1}} \begin{pmatrix} 0.618 & 1 \\ 0 & 0 \end{pmatrix}$$

$$0.618 K_1 + K_2 = 0 \rightarrow K_2 = t \text{ then } K_1 = -1.618t$$

$$V_2 = \begin{pmatrix} -1.618 \\ 1 \end{pmatrix}$$

remeber to normalize the eigenvectors 1st

$$\underset{N\times 1}{Z} = \underset{N\times N}{U^T}\,\underset{N\times 1}{Y}$$

$$\begin{pmatrix} 0.618 & -1.618 \\ 1 & 1 \end{pmatrix}^T$$

if we want 2 features that are uncorrelated.

→ To select one keep $\begin{pmatrix} 0.618 \\ 1 \end{pmatrix}$ corresponding to $\lambda_1 = 3.618$

- This preserves $\frac{3.618}{3.618+1.382} = 72.36\%$ of the datasets variance

$$\underset{1\times 1}{Z} = \frac{(0.618 \quad 1)}{\sqrt{1+0.618^2}} \underset{2\times 1}{Y}$$

2) Have shown in the lecture that

$$\Sigma_Y = U \Lambda U^T$$

matrix of eigenvalues of Σ_Y

matrix of eigenvectors as columns of Σ_Y

Σ_Z as well.

$$\Sigma_Y = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0.2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1.25 & 0.75 & 0 \\ 0.75 & 1.25 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}$$

Thank you !! :)