Niranjan Sundararajan
2020090

papergrid

Date: / /

SML - Assignment 1

Q1. Plotting done in Python.

i) For a zero-one loss,

$$\lambda = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

To minimize error,

Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$
and ω_2 otherwise (because $\lambda_{12} = \lambda_{21} = 1$ and $\lambda_{11} = \lambda_{22} = 0$).

$\Rightarrow P(\text{error}|x) = \min[P(\omega_1|x), P(\omega_2|x)]$.

$$\therefore P(\text{error}) = \int_{-\infty}^{x_0} P(\omega_2|x)\, P(x)\, dx + \int_{x_0}^{\infty} P(\omega_1|x)\, P(x)\, dx$$

where x_0 corresponds to the value of x for the Decision Boundary.

$\therefore$ For minimum value of $P(\text{error})$, x_0 would give us the correct decision boundary.

$$\therefore \frac{d\, P(\text{error})}{dx} = P(\omega_2|x_0)\, P(x_0) - P(\omega_1|x_0)\, P(x_0) = 0$$

$$\Rightarrow P(\omega_2|x_0)\, P(x_0) = P(\omega_1|x_0)\, P(x_0).$$

$$\Rightarrow P(x_0|\omega_2)\, P(\omega_2) = P(x_0|\omega_1)\, P(\omega_1)$$

$$\Rightarrow P(x_0|\omega_2) \cdot \frac{3}{4} = P(x_0|\omega_1)\, \frac{1}{4}$$

$$\Rightarrow 3\left(\frac{\cancel{1}}{\sqrt{2\cancel{\pi}\sigma_2^2}} \exp\left\{\frac{-1}{2} \frac{(x_0 - \mu_2)^2}{\cancel{\sigma_2^2}}\right\}\right) = \left(\frac{\cancel{1}}{\sqrt{2\cancel{\pi}\sigma_1^2}} \exp\left\{\frac{-1}{2} \frac{(x_0 - \mu_1)^2}{\cancel{\sigma_1^2}}\right\}\right).$$

As $\boxed{\sigma_1^2 = \sigma_2^2 = 1}$

P.T.O.

$$\therefore \ln 3 + \frac{-1}{2}(x_0 - 5)^2 = \frac{-1}{2}(x_0 - 2)^2$$

$$\Rightarrow \ln 3 = \frac{1}{2}\left[\cancel{x_0^2} + 25 - 10x_0 - \cancel{x_0^2} - 4 + 4x_0\right]$$

$$\ln 3 = \frac{1}{2}\left[21 - 6x_0\right]$$

$$\ln 9 = 21 - 6x_0.$$

$$6x_0 = 21 - \ln 9.$$

$$\boxed{x_0 = \frac{21 - \ln 9}{6} \simeq 3.134}$$

ii) ~~If we are minimizing the error, the DB would still continue to be~~ ~~$\boxed{x_0 = 3.134}$~~

∴ If we want to minimize the error, the the λ matrix comes into account.

$$\lambda = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}$$

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)\, P(\omega_j|x).$$

$$R(\alpha_1|x) = \cancel{\lambda_{11}}^{0} P(\omega_1|x) + \lambda_{12} P(\omega_2|x)$$

$$R(\alpha_2|x) = \lambda_{21} P(\omega_1|x) + \cancel{\lambda_{22}}^{0} P(\omega_2|x).$$

∴ The decision boundary for minimum error would be when

$$R(\alpha_1|x) = R(\alpha_2|x).$$ $x_0 \rightarrow$ DB.

$$\Rightarrow \lambda_{12} P(\omega_2|x_0) = \lambda_{21} P(\omega_1|x_0).$$

P.T.O.

This translates to,

$$\boxed{\frac{P(x_0|\omega_1)}{P(x_0|\omega_2)} = \frac{\lambda_{12}\ P(\omega_2)}{\lambda_{21}\ P(\omega_1)}}$$ (likehood ratio)

papergrid

Date: / /

$$2\ \frac{P(x_0|\omega_2)\cdot P(\omega_2)}{\cancel{P(x_0)}} = 3\ \frac{P(x_0|\omega_1)\cdot P(\omega_1)}{\cancel{P(x_0)}}$$

$$\Rightarrow 2\ P(x_0|\omega_2)\ \frac{\cancel{3}}{\cancel{4}} = \cancel{3}\cdot P(x_0|\omega_1)\ \frac{1}{\cancel{4}}$$

$$\Rightarrow 2\ P(x_0|\omega_2) = P(x_0|\omega_1)$$

$$\Rightarrow 2\left(\cancel{\frac{1}{\sqrt{2\pi(1)}}}\exp\left\{\frac{-1}{2}\frac{(x_0-5)^2}{1}\right\}\right)$$

$$= \left(\cancel{\frac{1}{\sqrt{2\pi}(1)}}\exp\left\{\frac{-1}{2}\frac{(x_0-2)^2}{1}\right\}\right)$$

$$\Rightarrow \ln 2 + \frac{-1}{2}(x_0-5)^2 = \frac{-1}{2}(x_0-2)^2$$

$$\Rightarrow \ln 2 = \frac{1}{2}\left[\cancel{x_0^2}+25-10x_0-\cancel{x_0^2}-4+4x_0\right]$$

$$\Rightarrow \ln 4 = 21-6x_0$$

$$\Rightarrow 6x_0 = 21-\ln 4$$

$$\boxed{x_0 = \frac{21-\ln 4}{6} \simeq 3.269}$$

- A zero-one data set would NOT be used for a real world cancer-prediction task.
  This is because, the risks associated in the conditional probabilities aren't the same for all predictions

Eg $\omega_1 \rightarrow$ has cancer
$\omega_2 \rightarrow$ does not have cancer.

$\therefore\ \lambda_{11}, \lambda_{22} = 0$ (Correct actions and predictions)

P.T.O.

Date: / /

However,

λ_{12} = Risk value associated with saying a person has cancer when s/he doesn't have cancer.

= c - (some positive const.).

λ_{21} = Risk value associated with saying a person doesn't have cancer when s/he has cancer.

= 10 c or 100 c or θc

(depending on the kind of cancer).

Here $\theta > 1$.

In a real world scenario, $\lambda_{21} \neq c$.

IMP [This is because, catching cancer early and treating it is necessary for its treatment.

P.T.O.

Q2. $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\mu = \begin{bmatrix} 5 \\ -5 \\ 6 \end{bmatrix}$ $A = [2 \ {-1} \ 2]^T$

$B = 5$

$$E(x) = \begin{bmatrix} E(x_1) \\ E(x_2) \\ E(x_3) \end{bmatrix} = \begin{bmatrix} +5 \\ -5 \\ 6 \end{bmatrix} = \mu$$

$$E(Y) = E(A^T x + B) = E\left([2 \ {-1} \ 2] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + B \right)$$

$$= E(2x_1 - x_2 + 2x_3 + 5)$$

$$= 2E(x_1) - E(x_2) + 2E(x_3) + 5$$

$$= (2 \times 5) - (-5) + 2(6) + 5$$

$$= 10 + 5 + 12 + 5$$

$$= \boxed{32}$$

This question does not use the co-variance matrix provided to us.

Q3. $P(x \mid \omega_i) = \dfrac{1}{\pi b} \cdot \dfrac{1}{1 + \left(\frac{x - a_i}{b}\right)^2}$, $i = 1, 2$

$P(\omega_1) = P(\omega_2) = 0.5$

A)

For minimum error rate,

$P(\omega_1 \mid x_0) = P(\omega_2 \mid x_0)$ [Equating posteriors, proved in Q1]

$x_0 \rightarrow DB$

$$\Rightarrow \frac{P(x_0 \mid \omega_1) P(\omega_1)}{\cancel{P(x_0)}} = \frac{P(x_0 \mid \omega_2) P(\omega_2)}{\cancel{P(x_0)}}$$

$$\Rightarrow \left(\frac{\cancel{1}}{\cancel{\pi b}} \cdot \frac{1}{1 + \left(\frac{x_0 - a_1}{b}\right)^2} \right) \times \cancel{\left(\frac{1}{2}\right)} = \frac{\cancel{1}}{\cancel{\pi b}} \cdot \frac{1}{1 + \left(\frac{x_0 - a_2}{b}\right)^2} \times \cancel{\left(\frac{1}{2}\right)}$$

$$\Rightarrow \cancel{1} + \left(\frac{x_0 - a_2}{\cancel{b}}\right)^2 = \cancel{1} + \left(\frac{x_0 - a_1}{\cancel{b}}\right)^2$$

$$\Rightarrow x_0 - a_2 = \pm x_0 - a_1$$

Here, $x_0 - a_2 \neq x_0 - a_1$ (Trivial case)

$$\therefore x_0 - a_2 = -x_0 + a_1$$

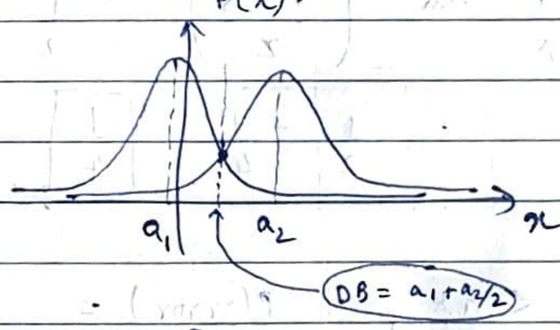
$$2x_0 = a_1 + a_2$$

$$\boxed{x_0 = \frac{a_1 + a_2}{2}}$$

B). Done in python

c). For overall error rate,
let us assume $\underline{a_1 < a_2}$

$\therefore$ The graphs would look something like this.



$$P(\text{error}|x) = \min\left[P(\omega_1|x), P(\omega_2|x)\right]$$

$$\therefore P(\text{error}) = \int_{-\infty}^{a_1+a_2/2} P(\omega_2|x)\, P(x)\, dx + \int_{a_1+a_2/2}^{\infty} P(\omega_1|x) P(x) dx$$

$$= \int_{-\infty}^{a_1+a_2/2} P(x|\omega_2) P(\omega_2)\, dx + \int_{a_1+a_2/2}^{\infty} P(x|\omega_1)\, P(\omega_1)\, dx$$

$$= \int_{-\infty}^{a_1+a_2/2} \frac{1}{\pi b} \cdot \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2} \left(\frac{1}{2}\right) dx + \int_{\frac{a_1+a_2}{2}}^{\infty} \frac{1}{\pi b} \frac{1}{1+\left(\frac{x-a_1}{b}\right)^2} \left(\frac{1}{2}\right) dx$$

$$= \frac{1}{2\pi \cancel{b}} \cancel{b} \left[\tan^{-1}\left(\frac{x-a_2}{b}\right)\right]_{-\infty}^{\frac{a_1+a_2}{2}} + \frac{1}{2\pi \cancel{b}} \cancel{b} \left[\tan^{-1}\left(\frac{x-a_1}{b}\right)\right]_{\frac{a_1+a_2}{2}}^{\infty}$$

P.T.O.

$$\frac{1}{2\pi}\tan^{-1}\left(\frac{\frac{a_1+a_2}{2}-a_2}{b}\right)-\frac{1}{2\pi}\tan^{-1}\left(-\infty\right)$$

$$+\frac{1}{2\pi}\tan^{-1}\left(+\infty\right)-\frac{1}{2\pi}\tan^{-1}\left(\frac{\frac{a_1+a_2}{2}-a_1}{b}\right)$$

$$=\frac{1}{2\pi}\tan^{-1}\left(\frac{a_1-a_2}{2b}\right)+\frac{1}{2\cancel{\pi}}\left(\frac{+\cancel{\pi}}{2}\right)$$

$$+\frac{1}{2\cancel{\pi}}\left(\frac{\cancel{\pi}}{2}\right)-\frac{1}{2\pi}\tan^{-1}\left(\frac{a_2-a_1}{2b}\right)$$

$$=\frac{1}{\pi}\tan^{-1}\left(\frac{a_1-a_2}{2b}\right)+\frac{1}{4}+\frac{1}{4}$$

$$\underline{P(error)}=\boxed{\frac{1}{\pi}\tan^{-1}\left(\frac{a_1-a_2}{2b}\right)+\frac{1}{2}}$$

for $a_1=3$, $a_2=5$, $b=1$

$$P(error)=\frac{1}{\pi}\tan^{-1}\left(\frac{\overset{-1}{\cancel{3-5}}}{\cancel{2}}\right)+\frac{1}{2}=\frac{1}{\cancel{\pi}}\cdot\left(\frac{-\cancel{\pi}}{4}\right)+\frac{1}{2}$$

$$=\frac{1}{2}-\frac{1}{4}=\boxed{\frac{1}{4}}$$

If $a_2<a_1$, $P(error)=\dfrac{1}{\pi}\tan^{-1}\left(\dfrac{a_2-a_1}{2b}\right)+\dfrac{1}{2}$

$\therefore$ P(error) for <u>any value of a_1 ; a_2</u>

$$=\boxed{\frac{1}{2}-\frac{1}{\pi}\tan^{-1}\left(\frac{|a_2-a_1|}{b}\right)}$$

P.T.O.

papergrid

Date: / /

Q4. a. pdf of x = $[a \;\; b]$

$a \rightarrow$ Bernoulli RV $\quad p(a=1) = \theta, \quad p(a=0) = (1-\theta)$

$b \rightarrow$ Gaussian RV $\quad p(b) = N(m, \sigma^2)$

$$\text{cov}[x] = \begin{bmatrix} \theta(1-\theta) & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

$\Rightarrow$ a and b are independent events

because $\boxed{\sigma_{12}^2 = \sigma_{21}^2 = 0}$.

If 2 RVs have $\sigma_{12}^2 / \sigma_{21}^2$ as 0, it is very likely that they are independent.

$\therefore$ $\boxed{\text{pdf of } x = \text{pdf of } (a, b) = \text{pdf}(a) \times \text{pdf}(b)}$

$$\therefore \quad p(x) = \begin{cases} \theta \; N(m, \sigma^2)(b) & \text{if } a = 1 \\ (1-\theta) \; N(m, \sigma^2)(b) & \text{if } a = 0 \end{cases}$$

In continuous terms,

$$* \; N(m, \sigma^2)(b) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2}\left(\frac{b-m}{\sigma}\right)^2\right\}$$

$$\boxed{p(x) = \theta^a (1-\theta)^{1-a} \times N(m, \sigma^2)(b)}$$

here, $a \in \{0, 1\}$, $b \in (-\infty, \infty)$

b. Let the N iid samples be :-

$p(x_1), p(x_2), \ldots p(x_N)$

$q(x)$ = Joint prob. of these samples

$= p(x_1, x_2, \ldots x_N)$

$= p(x_1) \cdot p(x_2) \cdots p(x_N)$

$\because$ IID samples are independent.

PTO

papergrid

Date: / /

$$\therefore q(x) = \prod_{i=1}^{N} \theta^{a_i} (1-\theta)^{1-a_i} \times N(m, \sigma^2)(b_i)$$

$$\Rightarrow \ln(q(x)) = \sum_{i=1}^{N} \ln\left(\theta^{a_i} (1-\theta)^{1-a_i} \times N(m,\sigma^2)(b_i)\right)$$

$$\Rightarrow F(\theta) = \sum_{i=1}^{N} a_i \ln\theta + \sum_{i=1}^{N} (1-a_i) \ln(1-\theta)$$
$$+ \sum_{i=1}^{N} \ln\left(N(m,\sigma^2)(b_i)\right)$$

For maximum value of θ, calculating $\underline{\frac{\partial}{\partial\theta} F(\theta) = 0}$

$$\Rightarrow \frac{\partial}{\partial\theta}(F(\theta)) = \frac{\sum_{i=1}^{N} a_i}{\theta} + \frac{\sum_{i=1}^{N} (1-a_i)(-1)}{(1-\theta)} + 0 = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{N} a_i}{\theta} = \frac{\sum_{i=1}^{N} (1-a_i)}{1-\theta}$$

$$\Rightarrow \sum_{i=1}^{N} a_i = \theta\left(\sum_{i=1}^{N} a_i + \sum_{i=1}^{N} (1-a_i)\right)$$

$$\Rightarrow \sum_{i=1}^{N} a_i = \theta\left(\cancel{\sum_{i=1}^{N} a_i} + N - \cancel{\sum_{i=1}^{N} a_i}\right)$$

$$\Rightarrow \boxed{\theta = \frac{\sum_{i=1}^{N} a_i}{N}}$$

$\therefore$ When $\theta = \frac{\sum_{i=1}^{N} a_i}{N}$, the value of $F(\theta)$ is max.

$\therefore \ln(q(x))$ is max $\Rightarrow \boxed{q(x) \text{ is max.}}$