PRITISH WADHWA
2019440

# SML ASSIGNMENT 1

A1) $P(w_1) = 1/4$ $\quad$ $P(w_2) = 3/4$

$P(x|w_1) = \mathcal{N}(2,1)$ $\quad$ $P(x|w_2) = \mathcal{N}(5,1)$

i) Zero - One loss $\quad$ $\lambda_{11} = 0$ $\quad$ $\lambda_{12} = 1$

now for decision boundary: $\quad$ $\lambda_{21} = 1$ $\quad$ $\lambda_{22} = 0$

$$(\lambda_{21} - \lambda_{11}) P(w_1|x) = (\lambda_{12} - \lambda_{22}) P(w_2|x)$$

$$P(w_i|x) = \frac{P(x|w_i)\, P(w_i)}{P(x)}$$

$$\therefore (1-0) \frac{P(x|w_1)\, P(w_1)}{\cancel{P(x)}} = (1-0) \frac{P(x|w_2)\, P(w_2)}{\cancel{P(x)}}$$

$$\Rightarrow \left(\frac{1}{\sqrt{2\pi(\sigma_1)^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\right)\left(\frac{1}{4}\right) = \left(\frac{1}{\sqrt{2\pi(\sigma_2)^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}\right)\left(\frac{3}{4}\right)$$

$$\Rightarrow \cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}(x-2)^2} \times \cancel{\frac{1}{4}} = \cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}(x-5)^2} \times \frac{3}{\cancel{4}}$$

$$\Rightarrow e^{-\frac{1}{2}(x-2)^2} = 3 e^{-\frac{1}{2}(x-5)^2}$$

taking log on both sides,

$$\ln e^{-\frac{1}{2}(x-2)^2} = \ln 3 e^{-\frac{1}{2}(x-5)^2}$$

$$\Rightarrow -\frac{1}{2}(x-2)^2 \ln e = \ln 3 - \frac{1}{2}(x-5)^2 \ln e$$

$$\Rightarrow \frac{1}{2}(\cancel{-x^2} - 4 + 4x) = \ln 3 - \frac{1}{2}(\cancel{x^2} - 25 + 10x)$$

$$\Rightarrow 6x = 21 - 2\ln 3$$

$$\Rightarrow x = \frac{21 - \ln 9}{6} = \frac{21 - 2.197}{6} = \frac{18.80}{6} = 3.133$$

$\therefore$ Decision boundary $= 3.133$

(ii) $\lambda_{11} = 0$ $\lambda_{12} = 2$
$\lambda_{21} = 3$ $\lambda_{22} = 0$

for decision boundary:

$$(\lambda_{21} - \lambda_{11}) P(x|\omega_1) P(\omega_1) = (\lambda_{12} - \lambda_{22}) P(x|\omega_2) P(\omega_2)$$

$$(3-0)\left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}\right)\left(\frac{1}{4}\right) = (2-0)\left(\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}\right)\left(\frac{3}{4}\right)$$

$$\Rightarrow \left(\cancel{\frac{3}{4}}\right)\left(\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}(x-2)^2}\right) = \left(\cancel{\frac{6}{4}}^{2}\right)\left(\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}(x-5)^2}\right)$$

$$\Rightarrow e^{-\frac{1}{2}(x-2)^2} = 2 e^{-\frac{1}{2}(x-5)^2}$$

taking log on both sides:

$$\Rightarrow \ln e^{-\frac{1}{2}(x-2)^2} = \ln 2e^{-\frac{1}{2}(x-5)^2}$$

$$\Rightarrow -\frac{1}{2}(x-2)^2 \ln e = \ln 2 - \frac{1}{2}(x-5)^2 \ln e$$

$$\Rightarrow -\cancel{x^2} - 4 + 4x = 2\ln 2 - \cancel{x^2} - 25 + 10x$$

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

$$\Rightarrow x = \frac{21 - 1.38}{6} = \frac{19.62}{6} = 3.27$$

$\therefore$ Decision boundary $= 3.27$

No zero-one loss wouldn't be preffered for a task like cancer prediction on a real world dataset because both false positives and false negatives are equally penalised. ~~This can cause~~ In real world, this is not always ~~true~~ ideal as for instance, predicting a person has cancer while he doesn't might not be as hazardous as predicting not cancer, while the person suffers from it.

ctd) $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ $\bar{U} = \begin{bmatrix} 5 \\ -5 \\ 6 \end{bmatrix}$ $A = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$ $B = 5$

Given $Y = A^T X + B$.

now $E[Y] = E[A^T X + B]$

$= E[A^T X] + E[B]$

$= E\left[ [2 \; -1 \; 2] \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \right] + E[5]$

$= E[2X_1 - X_2 + 2X_3] + E[5]$

$= E[2X_1] - E[X_2] + E[2X_3] + E[5]$

$= 2E[X_1] - E[X_2] + 2E[X_3] + E[5]$

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

$= 10 + 5 + 12 + 5$

$= 32$

$\therefore E[Y] = 32$ i.e. mean of $Y = A^T X + B$ is 32.

A3) A) $P(x|w_i) = \left(\frac{1}{\pi b}\right)\left(\frac{1}{1 + \left(\frac{x - a_i}{b}\right)^2}\right)$ $i = 1, 2$

$P(w_1) = P(w_2) = 1/2$

$\lambda_{11} = 0$ $\lambda_{12} = 1$ $\lambda_{21} = 1$ $\lambda_{22} = 0$

for decision boundary

$(\lambda_{21} - \lambda_{11}) P(x|w_1) P(w_1) = (\lambda_{12} - \lambda_{22}) P(x|w_2) P(w_2)$

$$\Rightarrow (1-0)\left(\cancel{\frac{1}{\pi b}}\right)\left(\frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}\right)\cancel{\left(\frac{1}{2}\right)} = (1-0)\left(\cancel{\frac{1}{\pi b}}\right)\left(\frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}\right)\cancel{\left(\frac{1}{2}\right)}$$

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

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

$$\Rightarrow x^2+a_2^2-2xa_2 = x^2+a_1^2-2xa_1.$$

$$\Rightarrow 2x(a_1-a_2) = (a_1-a_2)(a_1+a_2)$$

assuming $a_1 \neq a_2$.

$$2x = a_1+a_2$$

$$\Rightarrow \boxed{x = \frac{a_1+a_2}{2}}$$

c) $a_1 = 3 \quad a_2 = 5 \quad b = 1.$

Decision boundary for above variables :

$$x = \frac{3+5}{2} = 4$$

for $x<4$ (say $x=3$)

$$P(w_1|x=3) = \frac{P(x=3|w_1)\,P(w_1)}{P(x=3)}$$

$$= \left(\frac{1}{\pi}\right)\left(\frac{1}{1+0^2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{P(x=3)}\right) = \frac{1}{2\pi\, P(x=3)}$$

$$P(w_2|x=3) = \frac{P(x=3|w_2)\,P(w_2)}{P(x=3)}$$

$$= \frac{1}{\pi}\left(\frac{1}{1+(-1)^2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{P(x=3)}\right) = \frac{1}{4\pi\, P(x=3)}$$

~~Since $P(w_2|x=3$~~

$\therefore$ for $x<4$, $P(w_2|x)$ is less than $P(w_1|x)$

for $x > 4$ (say $x = 5$)

$$P(\omega_1 | x=5) = \frac{P(x=5|\omega_1)\, P(\omega_1)}{P(x=5)}$$

$$= \left(\frac{1}{\pi}\right)\left(\frac{1}{1+(1)^2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{P(x=5)}\right) = \frac{1}{4\pi\, P(x=5)}$$

$$P(\omega_2 | x=5) = \frac{P(\cancel{x}=5|\omega_2)\, P(\omega_2)}{P(x=5)}$$

$$= \left(\frac{1}{\pi}\right)\left(\frac{1}{1+0^2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{P(x=5)}\right) = \frac{1}{2\pi\, P(x=5)}$$

$\therefore$ for $x > 4$, $P(\omega_1|x)$ is less than $P(\omega_2|x)$

Now overall error rate $=$ $P(\text{error})$

$$P(\text{error}) = \int_{-\infty}^{4} \min[P(\omega_1|x), P(\omega_2|x)]\, P(x)\, dx + \int_{4}^{\infty} \min[P(\omega_1|x), P(\omega_2|x)]\, P(x)\, dx$$

$$= \int_{-\infty}^{4} P(\omega_2|x)\, P(x)\, dx + \int_{4}^{\infty} P(\omega_1|x)\, P(x)\, dx.$$

$$= \int_{-\infty}^{4} \frac{P(x|\omega_2)\, P(\omega_2)}{\cancel{P(x)}} \cancel{P(x)}\, dx + \int_{4}^{\infty} \frac{P(x|\omega_1)\, P(\omega_1)}{\cancel{P(x)}} \cancel{P(x)}\, dx.$$

$$= \int_{-\infty}^{4} \left(\frac{1}{\pi b} \cdot \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}\right)\left(\frac{1}{2}\right) dx + \int_{4}^{\infty} \left(\frac{1}{\pi b} \cdot \frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}\right)\left(\frac{1}{2}\right) dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{4} \frac{1}{1+(x-5)^2}\, dx + \frac{1}{2\pi} \int_{4}^{\infty} \frac{1}{1+(x-3)^2}\, dx$$

$$= \left(\frac{1}{2\pi}\right)\left(\tan^{-1}(x-5)\Big|_{-\infty}^{4}\right) + \left(\frac{1}{2\pi}\right)\left(\tan^{-1}(x-3)\Big|_{4}^{\infty}\right)$$

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

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

$\therefore$ Overall error rate $= 1/4 = 0.25$

Ans4) (a) $x = \begin{bmatrix} a \\ b \end{bmatrix}$  $a \rightarrow$ Bernoulli$(\theta)$
$b \rightarrow$ Gaussian$(m, \sigma)$

from given covariance matrix,

$cov(a, b) = 0$.
This is true for independent Random variables (but not always)

Assuming $a$ & $b$ are independent,
we get pdf of $x$ $= p(x) = \cancel{pdf}\ p(a, b)$
$= p(a) \cdot p(b)$

now $p(a) = (\theta^{a})(1-\theta)^{1-a}$ , $a \in \{0, 1\}$
please note that $\theta$ can not be equal to 0 or 1, else it will become indeterminate form $(0^0)$

$$p(b) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{b-m}{\sigma}\right)^2}$$

$$\therefore\ p(x) = p\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \left((\theta)^{a}(1-\theta)^{1-a}\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{b-m}{\sigma}\right)^2}\right)$$

(b) Now N iid samples are taken

for 1 $x = \begin{bmatrix} a \\ b \end{bmatrix}$  $p(x) = (\theta^{a}(1-\theta)^{1-a})\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{b-m}{\sigma}\right)^2}\right)$

for N such $x_i = \begin{bmatrix} a_i \\ b_i \end{bmatrix}$  $i \in \{1, 2, ---, N\}$

& $X = [x_1\ x_2\ x_3 - - - - -\ x_N]$

$$q(X) = p(x_1)\,p(x_2) - - - -\ p(x_n)$$
$$= \prod_{i=1}^{n} p(x_i) \qquad i \in \{1, 2, ---, N\}$$

$$= \prod_{i=1}^{n} \left((\theta)^{a_i}(1-\theta)^{1-a_i}\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{b_i-m}{\sigma}\right)^2}\right)$$

taking log on both sides

$$\ln q(X) = \ln\left(\prod_{i=1}^{n} p(x_i)\right)$$

$$= \sum_{i=1}^{n} \ln p(x_i) = \sum_{i=1}^{n} \ln\left((\theta)^{a_i}(1-\theta)^{1-a_i}\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{b_i-m}{\sigma}\right)^2}\right)\right)$$

$$\ln q(x) = \sum_{i=1}^{n}\left(\ln \theta^{a_i} + \ln(1-\theta)^{1-a_i} + \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \ln e^{-\frac{1}{2}\left(\frac{b_i-m}{\sigma}\right)^2}\right)$$

Differentiating both sides w.r.t. $\theta$.

$$\frac{q'(x)}{q(x)} = \sum_{i=1}^{n} \frac{a_i}{\theta} + \sum_{i=1}^{n}\left(\frac{1-a_i}{\theta-1}\right)$$

(rest terms were constant w.r.t $\theta$ thus got ~~can~~ zeroed out)

$$q'(x) = (q(x))\left(\sum_{i=1}^{n} \frac{\cancel{a_i\theta} - a_i + \theta - \cancel{a_i\theta}}{(\theta)(\theta-1)}\right)$$

$$= (q(x))\left(\sum_{i=1}^{n} \frac{\theta - a_i}{(\theta)(\theta-1)}\right) = 0.$$

now 2 cases : (i) $q(x) = 0$ (ii) $\sum_{i=1}^{n} \frac{\theta - a_i}{(\theta)(\theta-1)} = 0$

(i) $q(x) = 0$

$$\Rightarrow (\theta)^{a_i}(1-\theta)^{1-a_i} \times \mathcal{N}(m,\sigma) = 0.$$

now $\mathcal{N}(m,\sigma) \neq 0$ (property)

& since $\theta$ can not be equal to either 0 or 1 because in that case always one class would be predicted & ~~Only no gain~~ the ~~g~~ RV would be simply dependent on normal distribution. ~~Also the equation~~ And the point about indeterminate form still remains

thus $q(x) \neq 0$

$\therefore$ equation (ii) will hold.

Scanned with CamSca

(ii) $\sum_{i=1}^{n} \frac{\theta - a_i}{(\theta)(1-\theta)} = 0.$

$\Rightarrow \sum_{i=1}^{n} (\theta - a_i) = 0.$

$\Rightarrow n\theta - \sum_{i=1}^{n} a_i = 0$

$\Rightarrow \theta = \frac{\sum_{i=1}^{n} a_i}{n} \Rightarrow \frac{k}{n}$, $k \rightarrow$ no. of times $a_i$ was one. $0 \leq k \leq n$

Now if I take the double derivative of ~~g~~ $q(x)$, it would always come out to be -ve irrespective of the input. This means, there is only 1 solution for the maxima & that is at $q'(x) = 0$. Which proves that for $\theta = \frac{\sum_{i=1}^{n} a_i}{n} = \frac{k}{n}$, the joint probability will be maximized.