4)

(a)

Prior class probabilities of "Yes" class

$= \frac{\text{no. of Yes classes}}{\text{total no. of Yes \& No}} = \frac{3}{10}$

Prior class probability of "No"

$= 7/10$

(b) To find: $P(x|w_i)$ using MLE for continuous features.

here, Taxable Income is a continuous feature.

To find: MLE estimate

to compute sample mean & sample variance for both classes.

↳ Sample mean for class Yes $= \frac{1}{n} \sum$ Taxable income for Yes

$$\mu_{yes} = \frac{88+90+85}{3} = 87.67$$

↳ Sample variance for class "Yes" $1/n\,(x-\mu_{yes})^2$

$$\sigma^2_{yes} = (88-87.67)^2 + (90-87.67)^2 + (85-87.67)^2 / 3$$

$$= \frac{12.66}{3} = 4.222$$

∴ for class Yes, pdf is given by

$$P(x|Yes) = N(\mu_{yes}, \sigma^2_{yes})$$

x = taxable income

$$= N(87.67, 4.222)$$

Sample mean for class "No"

$$= \frac{122+77+106+210+72+117+60}{7}$$

$$\mu_{No} = 109.14$$

Sample variance for class "No"

$$\sigma^2_{No} = (122-109.14)^2 + (77-109.14)^2 + (106-109.14)^2 + (210-109.14)^2 + (72-109.14)^2 + (117-109.14)^2 + (60-109.14)^2$$

$$\sigma^2_{No} = \frac{\cancel{[illegible]}}{\cancel{7}} = \frac{15236.85}{7}$$

$$= 2176.69$$

$\therefore$ pdf for $p(x|NO)$

$$= P(x|\omega_i = NO) = N(109.14, 2176.69)$$

(c) discrete features: Refund and Marital Status.

to calculate pmf as:

$$P_{X|\omega_i}(x_j|\omega_i) = P(X = x_j|\omega_i) = \frac{n_{ji}}{n_i}$$

n_{ji} = no. of instances in class ω_i for j feature.

n_i = no. of instances in class ω_i

For Refund feature: let j = Refund, let Yes ~~[illegible]~~ class = $\omega_1 = i = 1$

$\therefore$ ~~[illegible]~~

for Refund = Yes & evade = Yes

$n_{ji} = 0$

$n_i = 3$

$\therefore$ ~~B~~ $P_{X|\omega_i}(x = \text{Refund} = \cancel{no}^{\,yes} | \omega_i = yes) \boxed{= 0}$

for,
refund = yes evade (w_i) = no

$P_{x|w_i = no}$ (yes, no) = $3/7$

refund = no evade = yes

$P_{x|w_i}$ (no, no) = $\cancel{3}/\cancel{3}3$ = 1

refund = no evade = no

$P_{x|w_i}$ (no, no) = $4/7$

For feature : marital status :

$P_{x|w_i}$ (Single, yes) = $2/3$

$P_{\cancel{x}|w_i}$ (Single, no) = $2/7$

$P_{x|w_i}$ (married, w_i = yes) = $0/3$ $\boxed{=0}$

$P_{x|w_i}$ (married, w_i = no) = $4/7$

$P_{x|w_i}$ (divorced, yes) = $1/3$

$P_{x|w_i}$ (divorced, no) = $1/7$

With the above estimation, we see that few have pmf = 0, this can be problematic, since all the features that belong to feature refund = yes will get classified as evade = yes & all the features with married will get classified as evade = no. in the test ~~test~~ set

(d)

Laplace correction helps avoid zero probabilities by adding 1 to each count and normalising by the total count + no. of possible levels l.

Ex: for feature : Refund

$P_{X|\omega_i}(\text{refund} = \text{yes} \mid \text{evade} = \text{yes}) = \frac{(n_{ij}+1)}{n_i + l} = \frac{0+1}{3+2} = 1/5$

$P_{X|\omega_i}(\text{refund} = \text{yes} \mid \text{evade} = \text{no}) = 4/_{7+2} = 4/9$

$P_{X|\omega_i}(\text{refund} = \text{no} \mid \text{evade} = \text{yes}) = \frac{3+1}{3+2} = 4/5 = 4/5$

$P_{X|\omega_i}(\text{refund} = \text{no} \mid \text{evade} = \text{no}) = \frac{4+1}{7+2} = \frac{5}{9}$

For feature: Marital Status

$P_{X|\omega_i}(\text{married} \mid \text{yes}) = \frac{1}{3+3} = \frac{1}{6}$

$P_{X|\omega_i}(\text{married} \mid \text{no}) = \frac{5}{9}$

$P_{X|\omega_i}(\text{Single} \mid \omega_i = \text{yes}) = 3/6 = 1/2$

$P_{X|\omega_i}(\text{Single} \mid \omega_i = \text{no}) = 3/9 = 1/3$

$P_{X|\omega_i}(\text{divorced} \mid \omega_i = \text{yes}) = 2/6 = 1/3$

$P_{X|\omega_i}(\text{divorced} \mid \omega_i = \text{no}) = 2/9$

This can be a valid estimate, Since we have not encountered any '0' probabilities.

(e) Minimum error Rate Decision Rule using Laplace correction.

$$P(\omega_i | x_j) = \frac{P(x|\omega_i)\, P(\omega_i)}{P(x)} \quad \text{—(i)}$$

Laplace correction: $P(x_j|\omega_i) = \dfrac{n_{ji}+1}{n_i + l}$

$$P(x|\omega_i) = \prod_{j=1}^{K} \frac{n_{ji}+1}{n_i+l}$$

j = features . here, 2 features
i = class.

this can be simplified to, taking log on (i)

$$\log P(\omega_i|x) = \log P(\omega_i) + \sum_{j=1}^{K} \frac{n_{ji}+1}{n_i+l}$$

∴ decision rule, becomes

$$\omega = \arg\max_i \left\{ \log P(\omega_i) + \sum_{j=1}^{K} \frac{n_{ji}+1}{n_i+l} \right\}$$