Problem-3: Bayesian Classification (15 points)

i) Classifying the new test example using Naive Bayes method,

The sample to classify is,

21	Gender = M	Car Type = Family	Shirt Size = Large	?

From the given table,

$P(\text{Class} = C0) = 10/20 = 0.5$ | $P(\text{Class} = C1) = 10/20 = 0.5$

For the attribute Gender,

$P(\text{Gender} = M \mid \text{Class} = C0) = 6/10$

$P(\text{Gender} = M \mid \text{Class} = C1) = 4/10$

$P(\text{Gender} = F \mid \text{Class} = C0) = 4/10$

$P(\text{Gender} = F \mid \text{Class} = C1) = 6/10$

For the Car Type attribute,

$P(\text{Car Type} = \text{Family} \mid \text{Class} = C0) = 1/10$

$P(\text{Car Type} = \text{Family} \mid \text{Class} = C1) = 3/10$

$P(\text{Car Type} = \text{Sports} \mid \text{Class} = C0) = 8/10$

$P(\text{Car Type} = \text{Sports} \mid \text{Class} = C1) = 0/10$

$P(\text{Car Type} = \text{Luxury} \mid \text{Class} = C0) = 1/10$

$P(\text{Car Type} = \text{Luxury} \mid \text{Class} = C1) = 7/10$

For Shirt Size attribute,

$P(\text{Shirt Size} = \text{Small} \mid \text{Class} = C0) = 3/10$

$P(\text{Shirt Size} = \text{Small} \mid \text{Class} = C1) = 2/10$

$P(\text{Shirt Size} = \text{Medium} \mid \text{Class} = C0) = 3/10$

$P(\text{Shirt Size} = \text{Medium} \mid \text{Class} = C1) = 4/10$

$P(\text{Shirt Size} = \text{Large} \mid \text{Class} = C0) = 2/10$

$P(\text{Shirt Size} = \text{Large} \mid \text{Class} = C1) = 2/10$

$P(\text{Shirt Size} = \text{Extra Large} \mid \text{Class} = C0) = 2/10$

$P(\text{Shirt Size} = \text{Extra Large} \mid \text{Class} = C1) = 2/10$

$$V_{NB} = \underset{v_j \in \{yes, no\}}{\text{argmax}} \; P(v_j) \prod_i P(a_i \mid v_j) \quad \text{——} \; (1)$$

A test record,

$X = (\text{Gender} = M, \text{Car Type} = \text{Family}, \text{Shirt Size} = \text{Large})$

~~P(X | Class = C0) × P(Gen~~

From eqn (1),

$$V_{NB} = \underset{v_j \in \{\cancel{yes}, \cancel{no}\} \; C0, C1}{\text{argmax}} \; P(v_j) \cdot P(\text{Gender} = M \mid v_j) \times P(\text{Car Type} = \text{Family} \mid v_j) \times P(\text{Shirt Size} = \text{Large} \mid v_j)$$

$$V_{NB}(C0) = P(\text{Class} = C0) \times P(\text{Gender} = M \mid C0) \times P(\text{Car Type} = \text{Family} \mid C0) \times P(\text{Shirt Size} = \text{Large} \mid C0)$$

$$= 0.5 \times 6/10 \times 1/10 \times 2/10$$

$\langle V_{NB}(C0) = 0.006 \rangle$

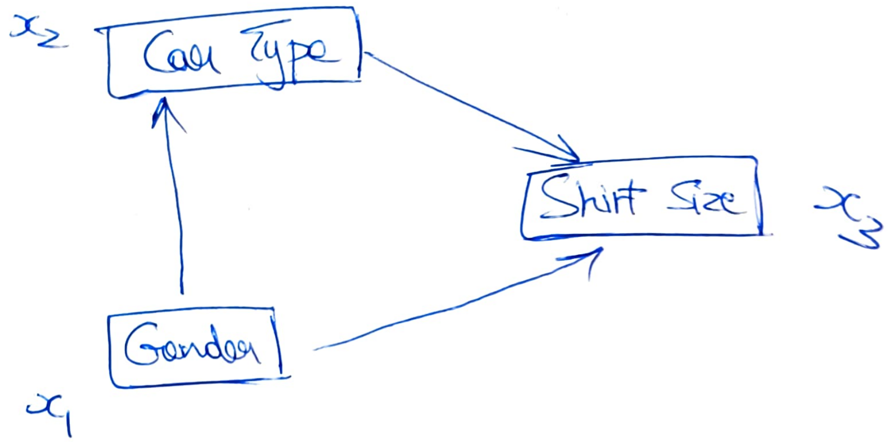
$$V_{NB}(C1) = P(\text{Class} = C1) \times P(\text{Gender} = M \mid C1) \times P(\text{Car Type} = \text{Family} \mid C1) \times P(\text{Shirt Size} = \text{Large} \mid C1)$$

$$= 0.5 \times 4/10 \times 3/10 \times 2/10$$

$\langle V_{NB}(C1) = 0.012 \rangle$

Since, $V_{NB}(c_1) > V_{NB}(c_2)$, the new test example will be classified as $\boxed{Class = c_1}$.

2) Classifying the test sample using Bayesian Network approach,



let Gender be x_1,

Car Type be x_2,

Shirt Size be x_3.

So, we need to predict the label for C given a vector of 3 attributes, this can be formulated as -

$$P(C|x_1, x_2, x_3) = \frac{P(x_1, x_2, x_3|C)P(C)}{P(x_1, x_2, x_3)}$$

Ignoring the denominator,

$$= P(x_1, x_2, x_3/C).P(C)$$

$$= \underline{P(x_1|C).P(x_2|C, x_1).P(x_3|C, x_1, x_2).P(C)}$$

~~Ef~~ Therefore, for ~~C~~ $C = C_0$,

$\Rightarrow P(x_1 = M,\ x_2 = \text{Family},\ x_3 = \text{Large} \mid C_0)\, P(C_0)$

$= 6/10 \times 1/6 \times 0/1 \times 10/20$

$= \cancel{0.05}\ 0$

For $C = C_1$,

$\Rightarrow P(x_1 = M,\ x_2 = \text{Family},\ x_3 = \text{Large} \mid C_1) \cdot P(C_1)$

$= 4/10 \times 3/4 \times 1/3 \times 10/20$

$= \underline{0.05}$

Therefore, we found out that the maximum value of $P(x_1, x_2, x_3 \mid C) \cdot P(C)$ is for class $= C_1$. Hence, the Bayesian network classifies the new sample as $\boxed{\text{class} = C_1}$