Q1 a) $H(t, D) = -\sum_{I \in levels(t)} \left(P(t=I) \cdot \log_2(P(t=I)) \right)$

$\Rightarrow$ let $I \in$ ~~[25,000, 50,000]~~ $\{\overset{25\text{-}50k}{[25{,}000, 50{,}000]}, \overset{\leq 25k}{[-\infty, 25{,}000]}, \overset{>50k}{[50{,}000, \infty]}\}$

$$\Rightarrow H(t,D) \Rightarrow -\sum_{I} \left(P(t=I) \cdot \log_2(P(t=I)) \right)$$

$$= -\left[P(t \in [25{,}000, 50{,}000]) \cdot \log_2(P(t \in [25{,}000, 50{,}000]))\right]$$
$$+ \left[P(t \in (-\infty, 25000]) \cdot \log_2(P(t \in (-\infty, 25000]))\right]$$
$$+ \left[P(t \in (50000, \infty)) \cdot \log_2(P(t \in (50000, \infty)))\right]$$

$$= -\left(5/8 \cdot \log_2(5/8)\right) + \left(2/8 \cdot \log_2(2/8)\right) + \left(1/8 + \log_2(1/8)\right)$$

$= 1.299$ bits (target feature)

$$\Rightarrow H(\text{Occupation}) = -\left(2/8 \cdot \log(2/8)\right) + \left(3/8 \cdot \log(3/8)\right) + \left(2/8 \cdot \log(2/8)\right) + \left(1/8 \cdot \log(1/8)\right)$$
$$= 1.9056$$

$$\Rightarrow H(\text{Marital Status}) = -\left(3/8 \cdot \log(3/8)\right) + \left(4/8 \cdot \log(4/8)\right) + \left(1/8 \cdot \log(1/8)\right)$$
$$= 1.4056$$

$$\Rightarrow H(\text{Education}) = -\left(3/8 \cdot \log(3/8)\right) + \left(4/8 \cdot \log(4/8)\right) + \left(1/8 \cdot \log(1/8)\right)$$
$$= 1.4056$$

$$\Rightarrow H(\text{Age}) = -\left(1/8 \cdot \log(1/8)\right) + 7\left(1/8 \cdot \log(1/8)\right)$$
$$= 3$$

Q1 b) $\Rightarrow$ $Gini(Income, D) = 1 - \sum_{I} P(I)^2$, $I \in \{[25000, 50000], (-\infty, 25000], (50,000, \infty)\}$

$= 1 - \left((5/8)^2 + (2/8)^2 + (1/8)^2 \right)$

$= 0.5312$

$\Rightarrow$ $Gini(Occupation) = 1 - \left((2/8)^2 + (3/8)^2 + (2/8)^2 + (1/8)^2 \right)$

$= 0.71875$

$\Rightarrow$ $Gini(Marital\ Status) = 1 - \left((3/8)^2 + (4/8)^2 + (1/8)^2 \right)$

$= 0.59375$

$\Rightarrow$ $Gini(Education) = 1 - \left((3/8)^2 + (4/8)^2 + (1/8)^2 \right)$

$= 0.59375$

$\Rightarrow$ $Gini(Age) = 1 - \left(8(1/8)^2 \right)$

$= 0.875$

c)

Q1 d) ⇒ let IG be the function for information gain, rem be remainde

⇒ $IG(\text{Education}) = H(\text{Annual Income}) - rem(\text{Education})$

$rem(\text{Education}) = \sum \left|\frac{P(I)}{|P|}\right| \cdot H(t, D)$

$= \left|\frac{\text{doctorate}}{P}\right| \cdot H(t, D = \text{doctorate}) + \left|\frac{\text{highschool}}{P}\right| \cdot H(t, D = \text{highschool})$

$+ \left|\frac{\text{batchelor}}{P}\right| \cdot H(t, D = \text{batchelor})$

$= (|\tfrac{1}{8}| \cdot 0) + (|\tfrac{4}{8}| \cdot 1) + (|\tfrac{3}{8}| \cdot 0)$

$= 0.5$

$\therefore IG(\text{Education}) = 1.2988 - 0.5$

$= 0.7988$

⇒ $IG(\text{Marital Status}) = H(\text{Annual Income}) - rem(\text{Marital Status})$

$rem(\text{Marital Status}) = \left|\frac{\text{divorced}}{D}\right| \cdot H(t, D = \text{divorced}) + \left|\frac{\text{married}}{D}\right| \cdot H(t, D = \text{married})$

$+ \left|\frac{\text{single}}{D}\right| \cdot H(t, D = \text{single})$

$= (|\tfrac{1}{8}| \cdot 0) + (|\tfrac{4}{8}| \cdot 0.811) + (|\tfrac{3}{8}| \cdot 0.918)$

$= 0.7495$

$\therefore IG(\text{Marital Status}) = 1.2988 - 0.7495$

$= 0.5493$

⇒ $IG(\text{Occupation}) = H(\text{Annual Income}) - rem(\text{Occupation})$

$rem(\text{Occupation}) = \left|\frac{\text{transport}}{D}\right| \cdot H(t, D = \text{transport}) + \left|\frac{\text{proffesional}}{D}\right| \cdot H(t, D = \text{proffesional})$

$+ \left|\frac{\text{agriculture}}{D}\right| \cdot H(t, D = \text{agriculture}) + \left|\frac{\text{armed forces}}{D}\right| \cdot H(t, D = \text{armed forces})$

$= (|\tfrac{2}{8}| \cdot 0) + (|\tfrac{3}{8}| \cdot 0.918) + (|\tfrac{2}{8}| \cdot 1) + (|\tfrac{1}{8}| \cdot 0)$

$= 0.594$

$\therefore IG(\text{Occupation}) = 1.2988 - 0.594$

$= 0.7048$

Q1 e) $\Rightarrow GR(\text{Education}) = \dfrac{IG(\text{Education})}{-\sum(P(d=1)) \cdot \log(P(d=1))}$

$$= \frac{0.7988}{-\left(\frac{3}{8} \cdot \log\left(\frac{3}{8}\right) + \frac{4}{8} \cdot \log\left(\frac{4}{8}\right) + \frac{1}{8} \cdot \log\left(\frac{1}{8}\right)\right)}$$

$$= 0.5683$$

$\Rightarrow GR(\text{Marital Status}) = \dfrac{IG(\text{Marital Status})}{-\sum(P(d=1)) \cdot \log(P(d=1))}$

$$= \frac{0.7495}{-\left(\frac{3}{8} \cdot \log\left(\frac{3}{8}\right) + \frac{4}{8} \cdot \log\left(\frac{4}{8}\right) + \frac{1}{8} \cdot \log\left(\frac{1}{8}\right)\right)}$$

$$= 0.3908$$

$\Rightarrow GR(\text{Occupation}) = \dfrac{IG(\text{Occupation})}{-\sum(P(d=1)) \cdot \log(P(d=1))}$

$$= \frac{0.7048}{-\left(\frac{2}{8} \cdot \log\left(\frac{2}{8}\right) + \frac{3}{8} \cdot \log\left(\frac{3}{8}\right) + \frac{1}{8} \cdot \log\left(\frac{1}{8}\right)\right)}$$

$$= 0.3699$$

Q1 f) $\Rightarrow IG(\text{Education}) = Gini(\text{Education}) - rem(\text{Education})$

$= 0.59375 - 0.5$

$= 0.09375$

$\Rightarrow IG(\text{Marital Status}) = Gini(\text{Marital Status}) - rem(\text{Marital Status})$

$= 0.59375 - 0.7495$

$= -0.15575$

$\Rightarrow IG(\text{Occupation}) = Gini(\text{Occupation}) - rem(\text{Occupation})$

$= 0.71875 - 0.5940$

$= 0.12475$

Q2 a) $\Rightarrow$ let R_0 = raining today, R_1 = raining yesterday, S_0 = sunny today, S_1 = sunny yesterday

$\Rightarrow P(R_0|R_1) = 0.70$; $P(S_0|S_1) = 0.4$

$\Rightarrow P(R_0|S_1) = 1 - P(S_0|S_1)$
$= 0.60$

$\Rightarrow P(S_0|R_1) = 1 - P(R_0|R_1)$
$= 0.30$

$\Rightarrow P(S_0) = P(S_0|R_1) \cdot P(R_1) + P(S_0|S_1) \cdot P(R_1)$
$= 0.35$

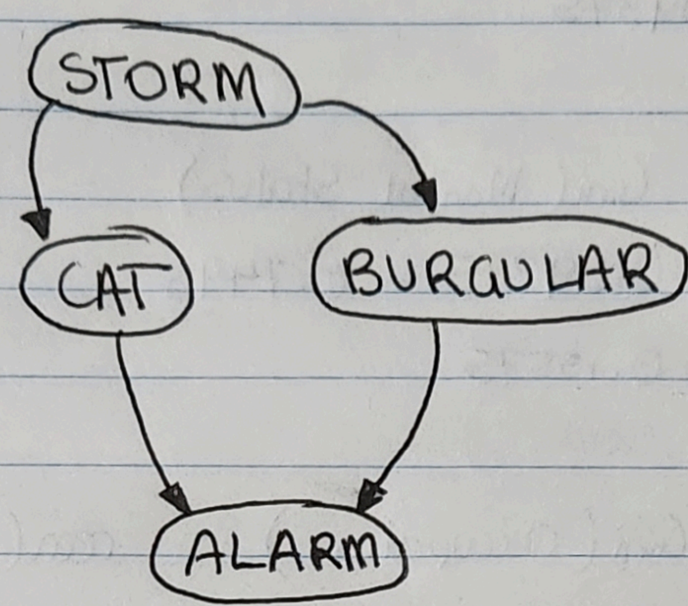
$\Rightarrow P(R_1) = P(R_1|S_0) \cdot P(S_0) + P(R_0|R_1) \cdot P(R_1)$
$(0.5) = P(R_1|S_0) \cdot (0.35) + (0.70)(0.50)$

$\therefore P(R_1|S_0) = 0.4285$

$\therefore$ there's a 42.8% chance of it raining yesterday if it was sunny today

Q3 a,b)

	P(S)
T	0.3077
F	0.6923



S	P(C=T|S)
T	0.75
F	0.25

S	P(B=T|S)
T	0.25
F	0.35

B	C	P(A=T|BC)
T	T	1.00
T	F	0.66
F	T	0.25
F	F	0.20

c) from the tables and CPT, when Burgular and Cat are true, Alarm is always true (100 % probability)