HW1

Q1a) This would not be possible as there is 2 variations of the same attribute which result in different classes

Different

A	B	C	Y
0	1	0	Yes
1	0	1	Yes
0	0	0	No
1	0	1	No
0	1	1	No
1	1	0	Yes

Q1b)

$$H(Y) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

$$H(Y) = -(3/6 \log_2 3/6) - (3/6 \log_2 3/6)$$

$$= -(.5(-1)) - (.5(1))$$

$$H(Y) = 1$$

$$H(Y|A=0) = -(1/3 \log_2 1/3) - (2/3 \log_2 2/3)$$

$$= -(.33 \times -1.585) - (.667 \times -.585)$$

$$H(Y|A=0) = 0.918$$

$$H(Y|A=1) = -(2/3 \log_2 2/3) - (1/3 \log_2 1/3)$$

$$H(Y|A=1) = 0.918$$

$$IG(A) = H(Y) - (\tfrac{3}{6} H(Y|A=0) + \tfrac{3}{6} H(Y|A=1))$$

$$= 1 - (\tfrac{3}{6} \times .918 + \tfrac{3}{6} \times .918)$$

$$= 0.082$$

$$H(Y|B=0) = -(\tfrac{1}{3}\log_2 \tfrac{1}{3}) - (\tfrac{2}{3}\log_2 \tfrac{2}{3})$$

$$= -(.33 \times -1.585) - (.667 \times -.585)$$

$$H(Y|B=0) = 0.918$$

$$H(Y|B=1) = -(\tfrac{2}{3}\log_2 \tfrac{2}{3}) - (\tfrac{1}{3}\log_2 \tfrac{1}{3})$$

$$H(Y|B=1) = 0.918$$

$$IG(B) = H(Y) - (\tfrac{3}{6} H(Y|B=0) + \tfrac{3}{6} H(Y|B=1))$$

$$= 1 - (\tfrac{3}{6} \times .918 + \tfrac{3}{6} \times .918)$$

$$= 0.082$$

$$H(Y|C=0) = -(1/3 \log_2 1/3) - (2/3 \log_2 2/3)$$

$$= -(.33 \times -1.585) - (.667 \times -.585)$$

$$H(Y|C=0) = 0.918$$

$$H(Y|C=1) = -(2/3 \log_2 2/3) - (1/3 \log_2 1/3)$$

$$H(Y|C=1) = 0.918$$

$$IG(C) = H(Y) - (3/6\, H(Y|C=0) + 3/6\, H(Y|C=1))$$

$$= 1 - (3/6 \times .918 + 3/6 \times .918)$$

$$= 0.082$$

All A, B, C Have the same information gain of 0.082

Q2)

$$H(Y) = -\left(\tfrac{4}{6}\log_2\tfrac{2}{6}\right) - \left(\tfrac{2}{6}\log_2\tfrac{2}{6}\right)$$

$$H(Y) = -(.0667 \times -0.585) - (0.333 \times -1.585)$$

$$H(Y) = 0.917$$

$H(Y \mid \text{Studed} = \text{Yes}) = 0$ All Passed

$$H(Y \mid \text{Studed} = \text{No}) = -\left(\tfrac{1}{3}\log_2\tfrac{1}{3}\right) - \left(\tfrac{2}{3}\log_2\tfrac{2}{3}\right)$$

$$= 0.918$$

$$IG(\text{Studied}) = H(Y) - \left(\tfrac{3}{6} H(Y \mid \text{Studied} = \text{false}) + H(Y \mid \text{Studd} = \text{T})\right)$$

$$= 0.917 - \left(\tfrac{3}{6} \times .918 + \tfrac{3}{6} \times 0\right)$$

$$= 0.458$$

$H(Y \mid GPA = High) = 0$ All pass

$H(Y \mid GPA = Medium) = -(1/2 \log_2 1/2) - (1/2 \log_2 1/2)$

$= 1$

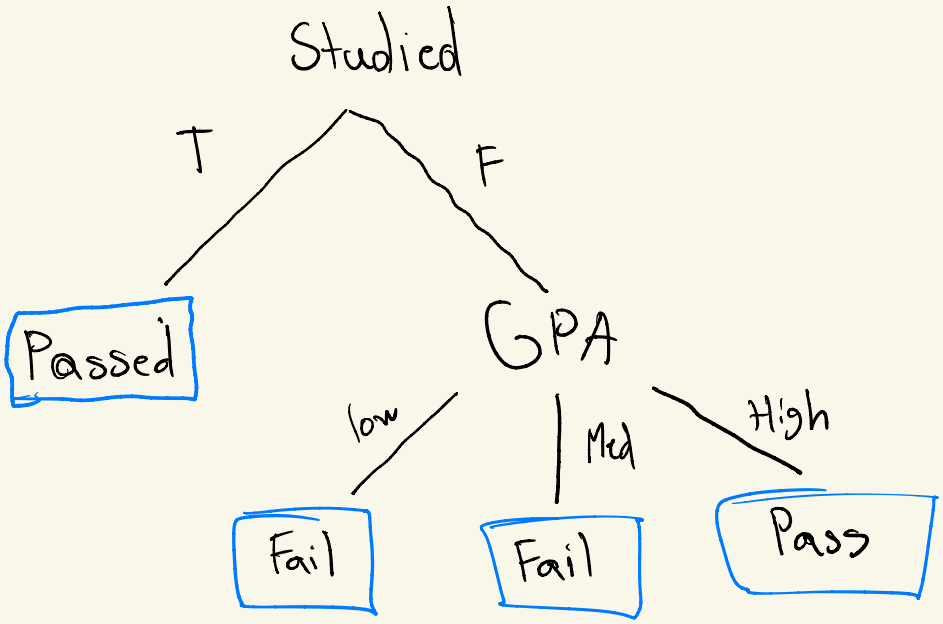
$H(Y \mid GPA = low) = -(1/2 \log_2 1/2) - (1/2 \log_2 1/2)$

$= 1$

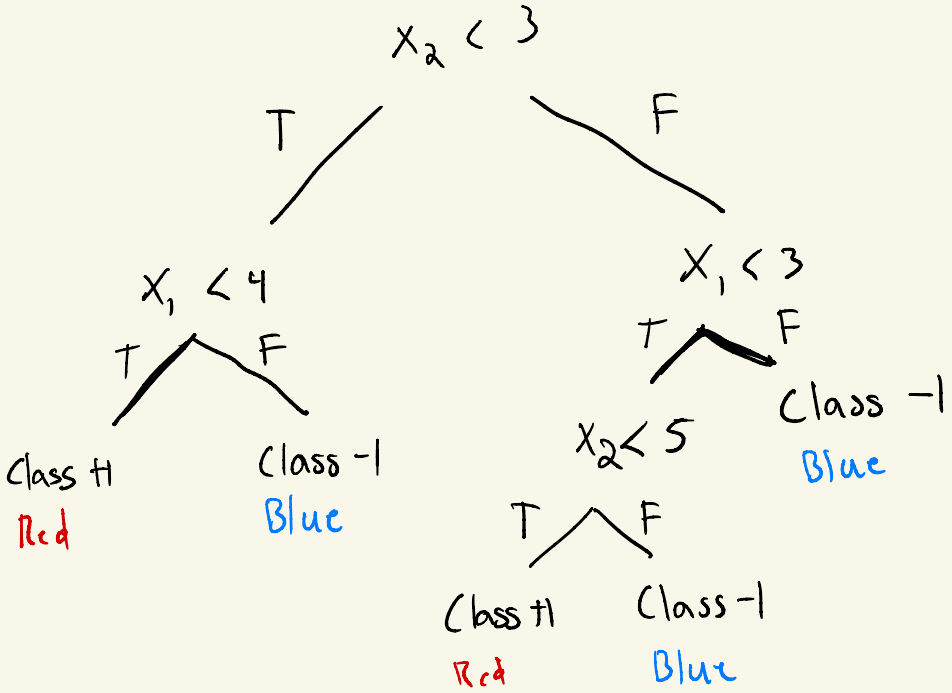
$IG(GPA) = 0.917 - (2/6 \times 1 + 2/6 \times 1 + 2/6 \times 0)$

$= 0.917 - 0.667$

$= 0.25$


Studied
T
F
Passed
GPA
low
Med
High
Fail
Fail
Pass

23)


$X_2 < 3$
T
F
$X_1 < 4$
T
F
Class +1
Red
Class -1
Blue
$X_1 < 3$
T
F
Class -1
Blue
$X_2 < 5$
T
F
Class +1
Red
Class -1
Blue