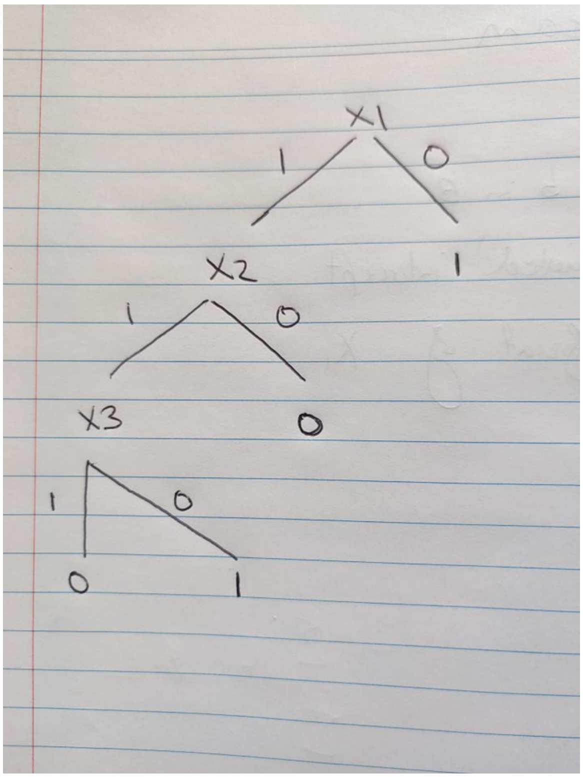
Question 1

a



i.From the working here 0.1, we can see that X_1 has the largest gain when splitting the data and therefore should be the first branch. X_2 and X_3 both have 0 gain. For the next split at $X_{1,1}$, both X_2 and X_3 produce the same gain which is $gain(X_{1,1}, X_2) = gain(X_{1,1}, X_3) = 0.39317$ which is why it doesn't matter which one is selected. I chose X_2 to be the next split and then finished the tree.

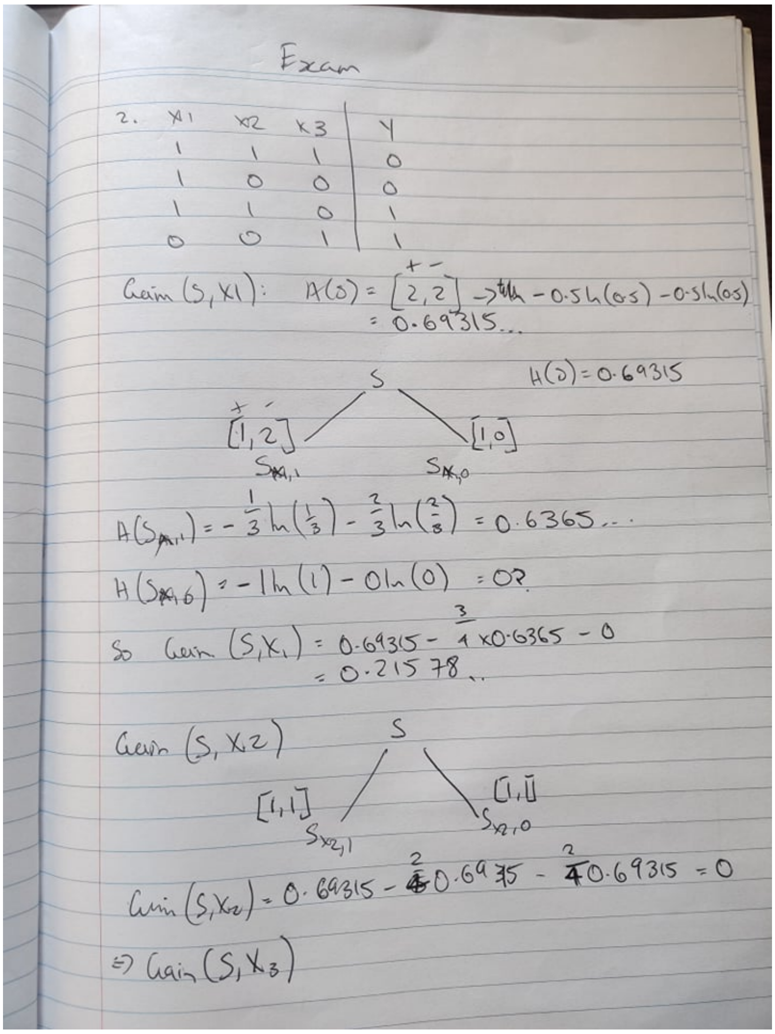
The training error for this tree is 0 since our tree uses every feature and there are no observations where the exact same values for each feature is given but it maps to a separate output.

ii. There is no depth 2 decision tree that has a lower training error than the one found with ID3. The training error is already the lowest it can be at 0. This tells us that ID3 can fit any dataset with no conflicting classifications and therefore has very high complexity. Without specifying the ID3 algorithm to stop early and with a non-conflicting dataset, ID3 can always achieve 0 training error.

b

Appendix

0.1 q2a



Exam

2.

X1	X2	X3	Y
1	1	1	0
1	0	0	0
1	1	0	1
0	0	1	1

Gain(S, X1): $H(S) = \overset{+\ -}{[2,2]} \to -0.5\ln(0.5) - 0.5\ln(0.5)$
$= 0.69315\ldots$

$H(S) = 0.69315$

S → $\overset{+\ -}{[1,2]}$ $S_{x_1,1}$; $[1,0]$ $S_{x_1,0}$

$H(S_{x_1,1}) = -\frac{1}{3}\ln\left(\frac{1}{3}\right) - \frac{2}{3}\ln\left(\frac{2}{3}\right) = 0.6365\ldots$

$H(S_{x_1,0}) = -1\ln(1) - 0\ln(0) = 0?$

So $\text{Gain}(S, X_1) = 0.69315 - \frac{3}{4} \times 0.6365 - 0$
$= 0.21578\ldots$

Gain(S, X2)

S → $[1,1]$ $S_{x_2,1}$; $[1,1]$ $S_{x_2,0}$

$\text{Gain}(S, X_2) = 0.69315 - \frac{2}{4}0.69315 - \frac{2}{4}0.69315 = 0$

$\Rightarrow \text{Gain}(S, X_3)$

$$H(X_{1,1}) = -\frac{1}{3}\ln\left(\frac{1}{3}\right) - \frac{2}{3}\ln\left(\frac{2}{3}\right)$$
$$= 0.6365\ldots$$

$\text{Gain}(X_{1,1}, X_2)$: $H(X_{1,1}) = 0.6365$

$[1,1]$ $H(X_{2,1})$ $[0,1]$ $H(X_{2,0})$

$$\text{Gain}(X_{1,1}, X_2) = 0.6365 - \frac{2}{3} \times 0.6365 - 0$$
$$= 0.39317\ldots$$

$H(X_{1,1}) = 0.6365$

$[0,1]$ $H(X_{3,1})$ $[1,1]$ $H(X_{3,0})$

$$\text{Gain}(X_{1,1}, X_3) = 0.6365 - 0 - \frac{2}{3} \times 0.6365$$
$$= 0.39317\ldots$$

So it doesn't matter which one we pick

Pick X_2.