Answer of question 18.2 :

1)

$$H(Y|X_1) = \frac{-3}{4}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) - \frac{1}{4}\left(0\log 0 + 1\log 1\right)$$

$$= 0,687$$

$$H(Y|X_2) = \frac{-1}{2}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right)\frac{-1}{2}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right)$$

$$= 1$$

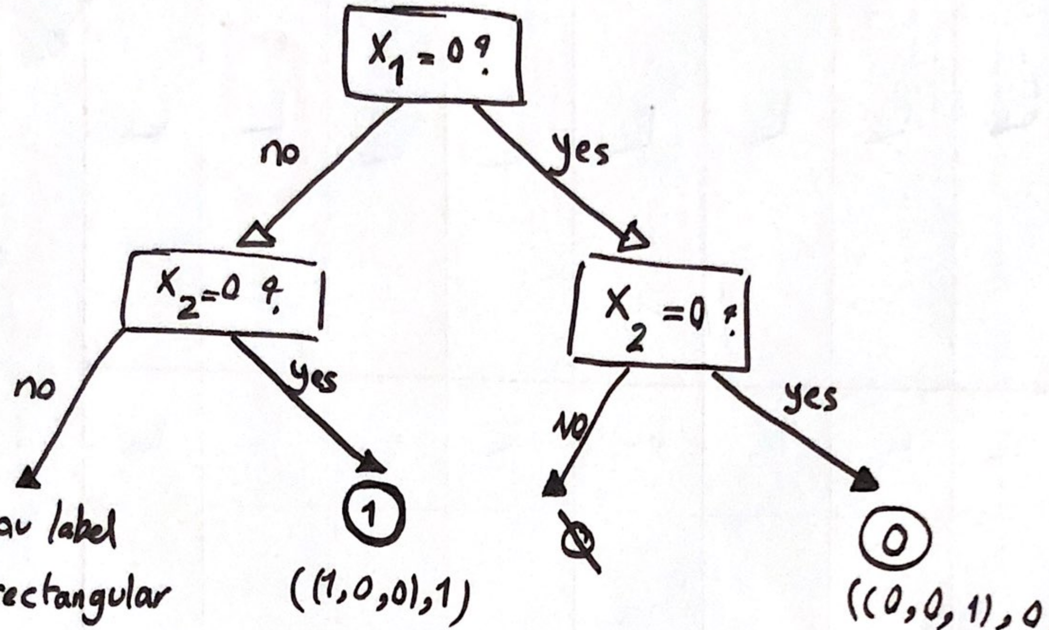
$$H(Y|X_3) = \frac{-1}{2}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right)\frac{-1}{2}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right)$$

$$= 1$$

We should choose the variable that has the highest IG or lowest $H(Y|X_i)$. As you know:

$$\boxed{\underset{j}{\operatorname{argmax}}\ I(X_j : Y) = \underset{j}{\operatorname{argmin}}\ H(Y|X_j)}$$

$\Rightarrow$ So, we'll pick X_1 as the root.



if you label this rectangular region → 1, then you will misclassify ((1,1,0), 0) and if you label it → 0, you'll misclassify ((1,1,0),1).

⟹ With the above tree with depth 2 and X_1 as the root, you'll missclassify 1 data point out of 4, and if you gather more data, you'll know that your error would at least be $\frac{1}{4}$.

2)