

Task 1)

$$a) N(\mu, \sigma^2) \Rightarrow P(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(u-\mu)^2}{2\sigma^2}}$$

$$H(u) = - \int_{-\infty}^{\infty} P(u) \ln P(u) du$$

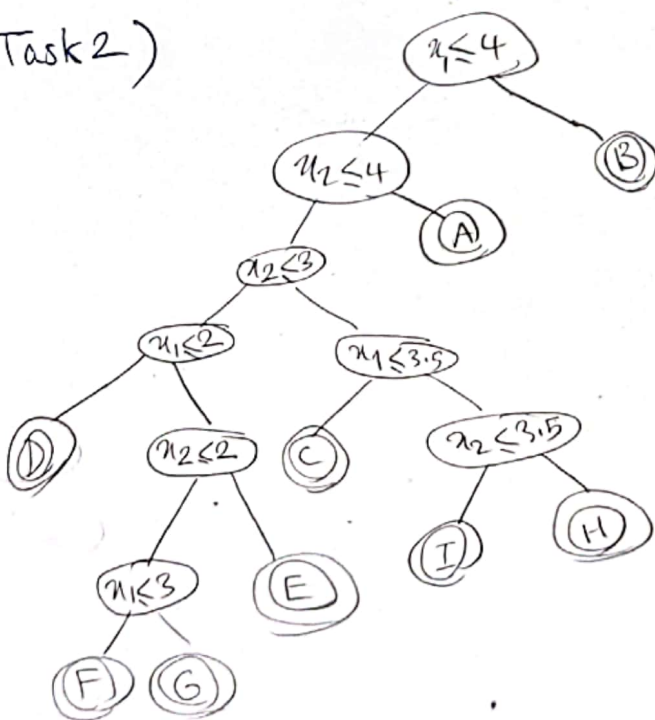
$$= \frac{1}{\sigma\sqrt{2\pi}} \left[ \int_{-\infty}^{\infty} \ln(\sigma\sqrt{2\pi}) e^{-\frac{(u-\mu)^2}{2\sigma^2}} du + \int_{-\infty}^{\infty} \frac{(u-\mu)^2}{2\sigma^2} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du \right]$$

$$= \ln(\sigma\sqrt{2\pi}) \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du}_{=1} + \frac{1}{2\sigma^2} \underbrace{\int_{-\infty}^{\infty} \frac{(u-\mu)^2}{\sigma\sqrt{2\pi}} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du}_{=\sigma^2}$$

$$= \frac{1}{2} \ln(2\pi\sigma^2)$$

$$b) H(u) < 0 \Rightarrow 2\pi\sigma^2 < 1 \Rightarrow 0 < \sigma < \frac{1}{\sqrt{2\pi}}$$

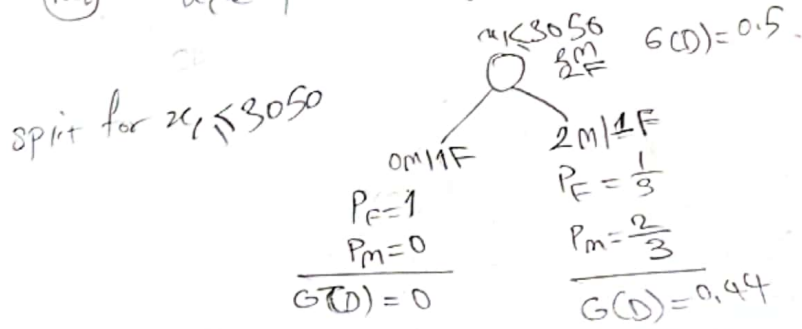
Task 2)



# Task 3)

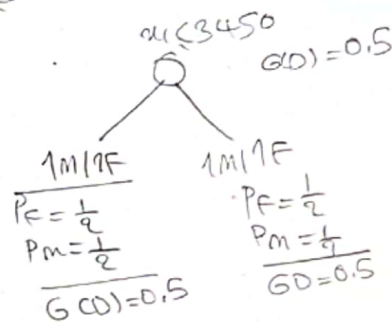
$$N_{min} = 1, G(D) = 1 - \sum_{t=1}^T p_t^2, GI(D, D_1, D_2) = G(D) - \sum_{r=1}^2 \frac{|D_r|}{|D|} G(D_r)$$

$$\text{root } u_1 \in \{2700, 3400, 3500, 5500\} \Rightarrow \text{splits} \in \{3050, 3450, 4500\}$$



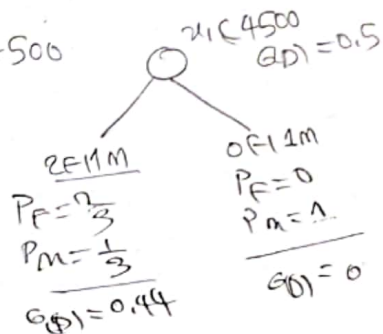
$$GI = 0.5 - \left( \frac{1}{4}(0) + \frac{3}{4}(0.44) \right) = 0.17$$

split for  $x_1 \leq 3450$



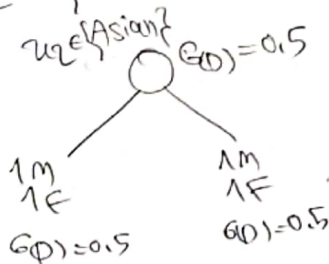
$$GI = 0.5 - \left( \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} \right) = 0$$

split for  $x_1 \leq 4500$



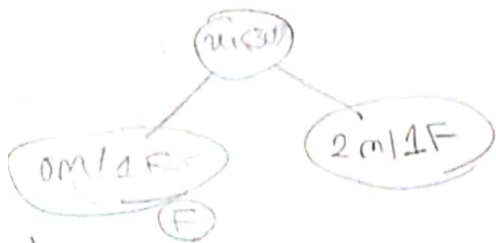
$$GI = 0.5 - \left( \frac{1}{4}(0) + \frac{3}{4}(0.44) \right) = 0.17$$

$u_2 \in \{\text{Asian}, \text{African}\} \Rightarrow \text{splits} \in \{\{\text{African}\} \text{ vs } \{\text{Asian}\}\}$



$$GI = 0$$

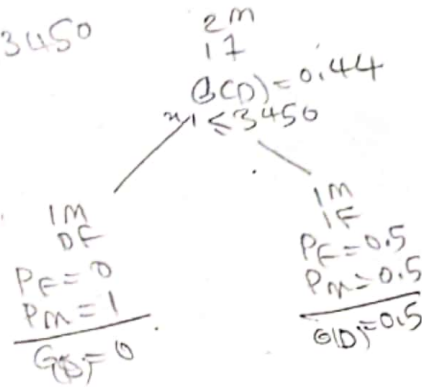
Hence here the best GI is for  $x_1 \leq 3050$  or  $x_1 \leq 4500$   
we choose the  $x_1 \leq 3050$  for our first split.



hence we only have one class here (F), we note this leaf F.

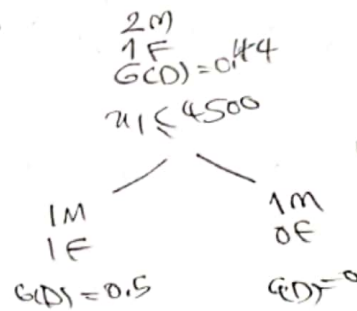
$$u_1 \in \{3400, 3500, 5500\} \Rightarrow \text{splits} \in \{3450, 4500\}$$

for the split  $u_1 \leq 3450$



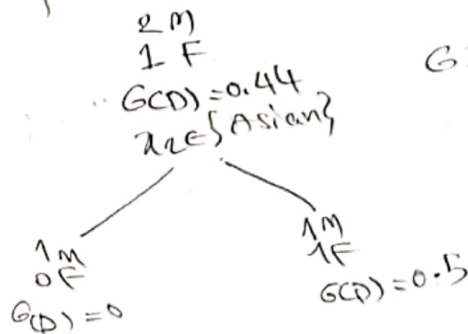
$$GI = 0.44 - \left( \frac{1}{3} \times 0 + \frac{2}{3} \times \frac{1}{2} \right) = 0.11$$

for the split  $u_1 \leq 4500$



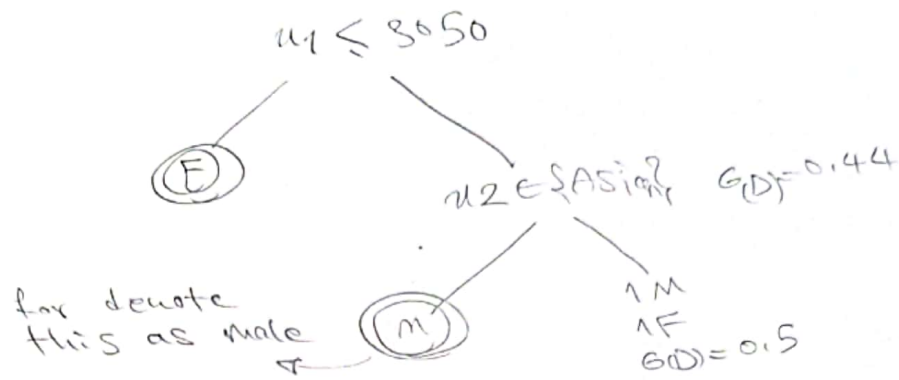
$$GI = 0.11$$

for the split  $u_2 \in \{\text{Asian}\}$

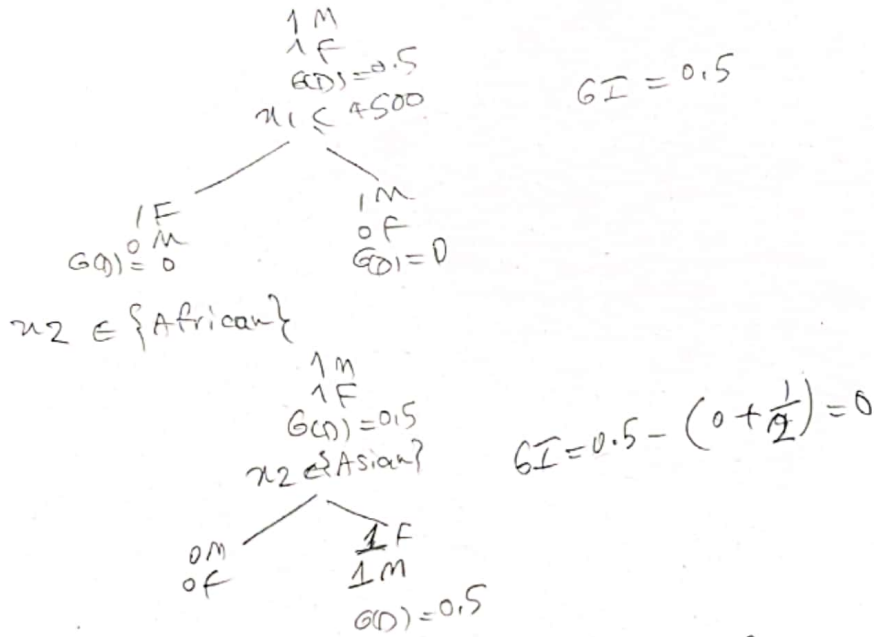


$$GI = 0.11$$

we here have the same GI with all splits; So we select  $u_2 \in \{\text{Asian}\}$  this time.

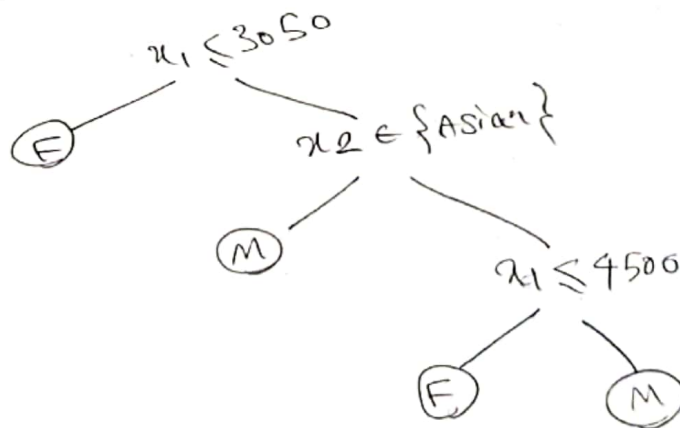


$x_1 \in \{3500, 5500\} \Rightarrow \text{split } x_1 \leq 4500$



Hence we select the better  $G_I$  (0.5) and the split  $x_1 \leq 4500$  is selected at the end, we stop at the leaves cause we only have a single label.

The final tree:



The minimal Depth would be for this tree:



Because we used the Gini-index as the criteria, and the fact that the algorithm is greedy, which means it selects the best split (aiming for pure nodes) at each step, it would miss the highlighted pink node in the tree above that is actually not pure and the best choice when splitting from the start.