Pedro Nunes   109368
Andrei Barb   109762

1)

$$IG(y_i) = E(\text{class}) - E(\text{class} \mid y_i)$$

$$E(\text{class}) = -\sum P(\text{class} = C) \log_2 P(\text{class} = C)$$

$$E(\text{class} \mid y_i) = -\sum P(y_i = k) E(\text{class} \mid y_i = k)$$

$$IG(y_2 \mid y_1 > 0,4) = E(\text{class} \mid y_1 > 0,4) - E(\text{class} \mid y_2, y_1 > 0,4)$$

$$E(\text{class} \mid y_1 > 0,4) = -\left(\frac{2}{8} \log_2 \frac{2}{8} + 2 \times \left(\frac{3}{8} \log_2 \frac{3}{8} \right) \right) = 1,56$$

$$E(\text{class} \mid y_2, y_1 > 0,4) = \frac{1}{8}\left(-1 \log_2 1 \right) + \frac{4}{8}\left(-\left(\frac{2}{4} \log_2 \frac{2}{4} + \right.\right.$$

$$\left.\left. 2 \times \left(\frac{1}{4} \log_2 \frac{1}{4} \right) \right) \right) + \frac{3}{8}\left(-\left(\frac{2}{3} \log_2 \frac{2}{3} + \right.\right.$$

$$\left.\left. \frac{1}{3} \log_2 \frac{1}{3} \right) \right) = \frac{1}{8} \times 0 + \frac{4}{8} \times 1,5 + \frac{3}{8} \times 0,918$$
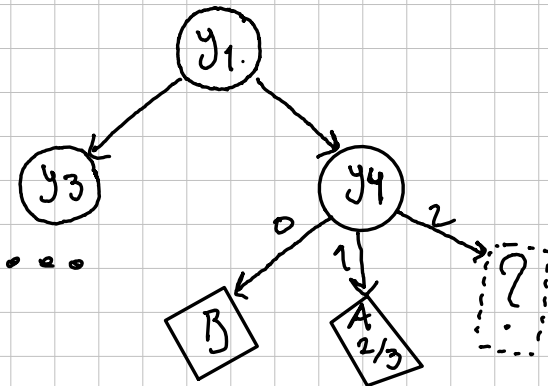
$$= 1,09$$

$$IG(y_2 \mid y_1 > 0,4) = 1,56 - 1,09 = 0,47$$

$$E(\text{class} \mid y_3, y_1 > 0.4) = \frac{3}{8}\left(-\left(3 \times \left(\frac{1}{3}\log_2\frac{1}{3}\right)\right)\right) + \frac{2}{8}\left(-\left(2 \times \left(\frac{1}{2}\log_2\frac{1}{2}\right)\right)\right)$$

$$+ \frac{3}{8}\left(-\left(3 \times \left(\frac{1}{3}\log_2\frac{1}{3}\right)\right)\right)$$

$$= \frac{3}{8} \times 1.585 + \frac{2}{8} \times 1 + \frac{3}{8} \times 1.585$$

$$= 1.44$$

$$IG(y_3 \mid y_1 > 0.4) = 1.56 - 1.44 = 0.12$$

$$E(\text{class} \mid y_4, y_1 > 0.4) = \frac{4}{8}\left(-\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right)\right)$$

$$+ \frac{1}{8}\left(-\left(1\log_2 1\right)\right) + \frac{3}{8}\left(-\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)\right)$$

$$= \frac{4}{8} \times 0.81 + \frac{1}{8} \times 0 + \frac{3}{8} \times 0.918 = 0.75$$

$$IG(y_4 \mid y_1 > 0.4) = 1.56 - 0.75 = 0.81$$

$$IG(y_2 \mid y_1 > 0,4, \; y_4 = 2) = E(class \mid y_1 > 0,4, y_4 = 2)$$
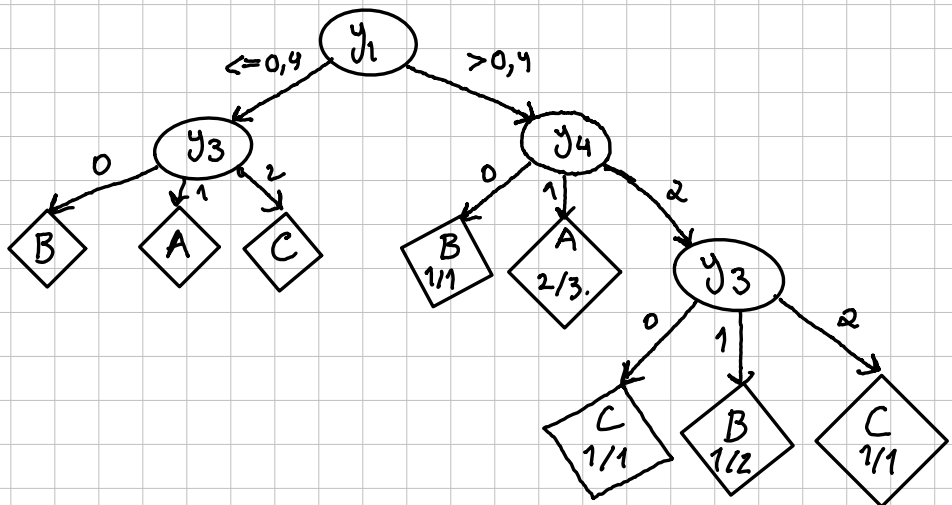$$- E(class \mid y_2, \; y_1 > 0,4, \; y_4 = 2)$$

$$E(class \mid y_1 > 0,4, \; y_4 = 2) = -\left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4}\right) = 0,81$$

$$E(class \mid y_2, y_1 > 0,4, y_4 = 2) = \frac{1}{4}\left(-\left(1 \log_2 1\right)\right) + \frac{3}{4}\left(-\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right)\right)$$
$$= \frac{1}{4} \times 0 + \frac{3}{4} \times 0,918 = 0,69$$

$$IG(y_2 \mid y_1 > 0,4, \; y_4 = 2) = 0,81 - 0,69 = 0,12$$

$$E(class \mid y_3, y_1 > 0,4, y_4 = 2) = \frac{1}{4}\left(-\left(1 \log_2 1\right)\right) + \frac{2}{4}\left(-\left(2 \times \left(\frac{1}{2} \log_2 \frac{1}{2}\right)\right)\right)$$
$$+ \frac{1}{4}\left(-1 \log_2 1\right) = \frac{2}{4} \times 1 = 0,5$$

$$IG(y_3 \mid y_1 > 0,4, \; y_4 = 2) = 0,81 - 0,5 = 0,31$$

2)

Predicted		A	B	C
	A	4	0	0
	B	1	3	1
	C	0	0	3

3)

class A:

$$\text{precisiom} = \frac{TP}{TP+FP} = \frac{4}{4 + 1 + 0} = \frac{4}{5}$$

$$\text{recal} = \frac{TP}{TP+FN} = \frac{4}{4} = 1$$

$$F_1 = 2 \times \frac{\text{precisiom} \times \text{recal}}{\text{precisiom} + \text{recal}} = 2 \times \frac{4/5 \times 1}{4/5 + 1} = 0,(8)$$

Class B:

$$\text{precisiom} = \frac{3}{3+0+0} = 1$$

$$\text{recal} = \frac{3}{1+3+1} = 3/5$$

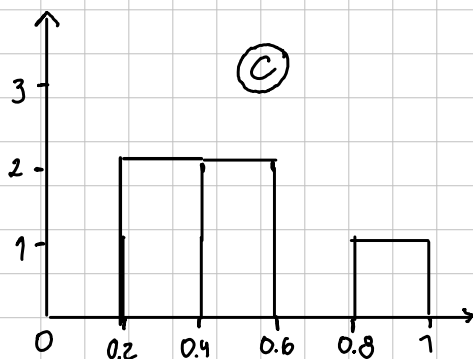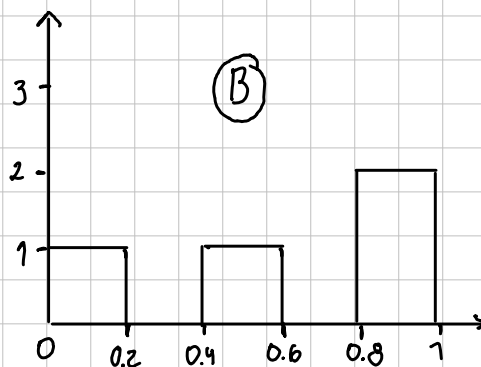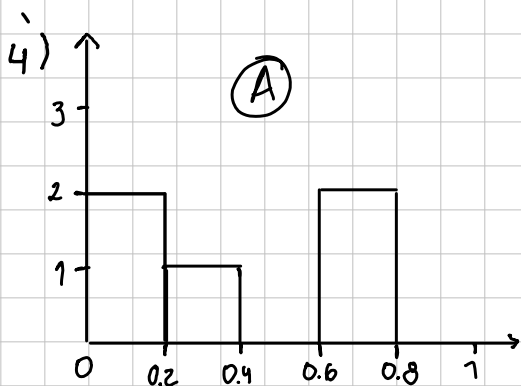$$F_1 = 2 \times \frac{1 \times 3/5}{1 + 3/5} = 0,75$$

The class with the lowest F1-score is B

Class C:

$$\text{precisiom} = \frac{3}{3+1+0} = 3/4$$

$$\text{recal} = \frac{3}{3+0+0} = 1$$

$$F_1 = 2 \times \frac{1 \times 3/4}{1 + 3/4} \approx 0,857$$

4)

Bim 1 $[0, 0,2[$: A = 0,4 B = 0,25 C = 0 $\therefore$ A
Bim 2 $[0,2 ; 0,4[$: A = 0,2 B = 0 C = 0,4 $\therefore$ C
Bim 3 $[0,4 ; 0,6[$: A = 0 B = 0,25 C = 0,4 $\therefore$ C
Bim 4 $[0,6 ; 0,8[$: A = 0,4 B = 0 C = 0 $\therefore$ A
Bim 5 $[0,8 ; 1[$: A = 0 B = 0,5 C = 0,2 $\therefore$ B

$\begin{cases} y_1 < 0,2 \Rightarrow A \\ 0,2 \leq y_1 < 0,6 \Rightarrow C \\ 0,6 \leq y_1 < 0,8 \Rightarrow A \\ 0,8 \leq y_1 < 1 \Rightarrow B \end{cases}$

5) Bounds = $[Q_1 - 1.5 \times IQR \ , \ Q_3 + 1.5 \times IQR]$
   $IQR = Q_3 - Q_1$

| | Q_1 | | | | Q_2 | | | | Q_3 | | |

0,12 0,18 0,25 | 0,33 0,45 0,52 | 0,58 0,62 0,71 | 0,83 0,90 0,95

$Q_1 = \dfrac{0.25 + 0.33}{2} = 0.29$

$Q_3 = \dfrac{0.71 + 0.83}{2} = 0.77$

$IQR = 0.77 - 0.29 = 0.48$

Bounds $= [0.29 - 1.5 \times 0.48 \ ; \ 0.77 + 1.5 \times 0.48] = [-0.43 ; 1.49]$

All y, values are im $[0.12, 0.95]$, wich lies inside $[-0.43 ; 1.49]$.
Therefore, there are mo outliers.