HW - 1

Pen - and - paper

1. a) Entropia : $\epsilon(Class) = \left(\frac{2}{8}\log_2\frac{2}{8} + \frac{3}{8}\log_2\frac{3}{8} + \frac{3}{8}\log_2\frac{3}{8}\right) \simeq 1.5613$

$y_2 \rightarrow$ Split : $y_2 = 1 \wedge y \neq 1$

$\epsilon(Class\ |y_1 = 1) = -\frac{4}{8}\left(\frac{1}{4}\log_2\frac{1}{4} + \frac{1}{4}\log_2\frac{1}{4} + \frac{2}{4}\log_2\frac{2}{4}\right) = 0,75$

$\epsilon(class\ |y_1 \neq 1) = -\left(\frac{1}{8}\left(1\log_2 1\right) + \frac{3}{8}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)\right) \simeq 0,3444$

$IG(y_2) = 1.5613 - (0,3444 + 0,75) = 0,4669$

$y_4 \rightarrow$ Split : $y_4 = 2 \wedge y_4 \neq 2$

$\epsilon(class\ |y_4 = 2) = -\frac{4}{8}\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) \simeq 0,40569$

$\epsilon(class\ y_4 \neq 2) = -\left(\frac{1}{8}\left(1\log_2 1\right) + \frac{3}{8}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)\right) \simeq 0,3444$

$IG(y_4) = 1.5613 - (0,40569 + 0,3444) \simeq 0,8113$

Maior $IG : y_4$

Com $y_4 = 2$

	y_2	y_3	y_4	y_{out}
x_3	1	0	2	C
x_7	2	1	2	C
x_{11}	2	1	2	B
x_{12}	2	2	2	C

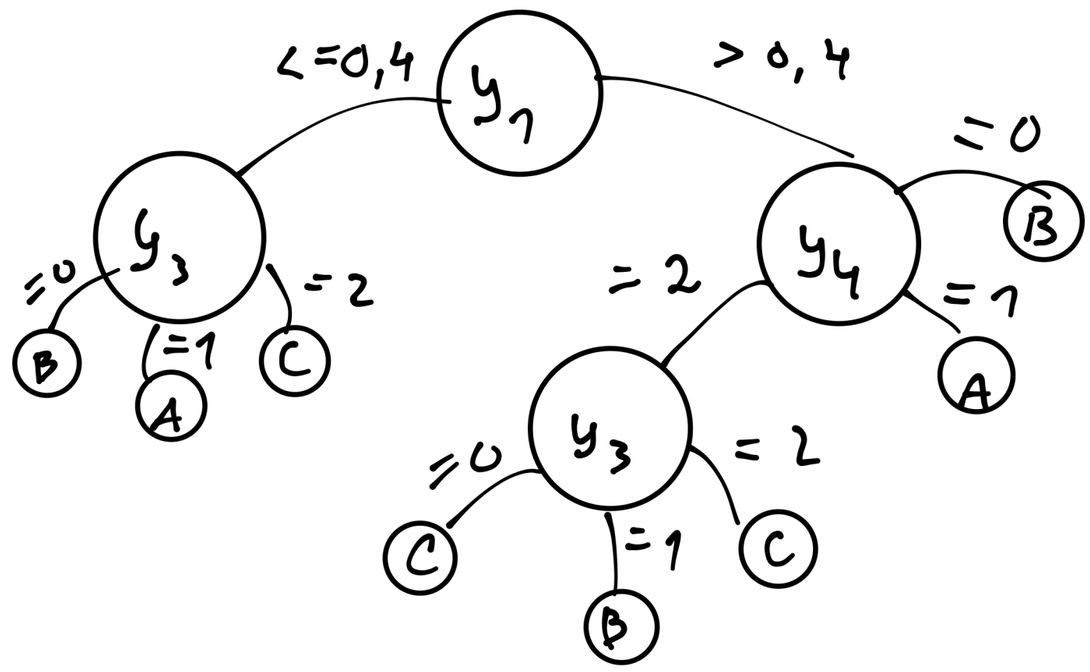
$\epsilon(Class\ |y_2) = \frac{1}{4}\left(1\log_2 1\right) + \frac{3}{4}\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right| = 0,6887$

$\epsilon(Class) = -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) \simeq 0,08113$

$\epsilon(Class\ |y_3) = \frac{1}{4}\left(1\log_2 1\right) + \frac{2}{4}\left(\frac{1}{2}\log_2\frac{1}{2} \times 2\right) + \left(1\log_2 1\right) \simeq 0,5$

$IG(y_2) \simeq -0,60757 \quad IG(y_3) \simeq -0,41887$

Árvore:



2.

Pred. \ True	A	B	C
A	4	1	0
B	0	3	1
C	0	0	3

	R	P
x_1	A	A
x_2	B	B
x_3	C	C
x_4	A	A
x_5	C	C
x_6	B	B
x_7	C	B
x_8	A	A
x_9	A	A
x_{10}	B	A
x_{11}	B	B
x_{12}	C	C

3. $F_1 = 2 \times \frac{precision \times recoil}{precision + recoil}$ $\quad recoil = \frac{TP}{TP + FN}$ $\quad precision = \frac{TP}{TP + FP}$

$A \rightarrow$ recoil : $TP_A = 4$ $\quad FN = 0 \rightarrow 1$

precision : $TP_A = 4$ $\quad FP = 1 \rightarrow 0,8$

$F_{1_A} = \frac{2 \times 0,8 \times 1}{0,8 + 1} \simeq 0,889$

$B \rightarrow$ recoil : $TP_B = 3$ $\quad FN = 1 \rightarrow 0,75$

precision : $TP_B = 3$ $\quad FP = 1 \rightarrow 0,75$

$F_{1_B} = \frac{2 \times 0,75 \times 0,75}{2 \times 0,75} = 0,75$

$C \rightarrow$ recoil : $TP_C = 3$ $\quad FN = 1 \rightarrow 0,75$

precision: $TP_C = 3$ $\quad FP = 0 \rightarrow 1$

$F_{1_C} = \frac{2 \times 1 \times 0,75}{1,75} \simeq 0,857$

Lower F_1 score : B

4.

$p(0 \leq y_1 < 0{,}2 \mid y_{out} = A) = 2/5$

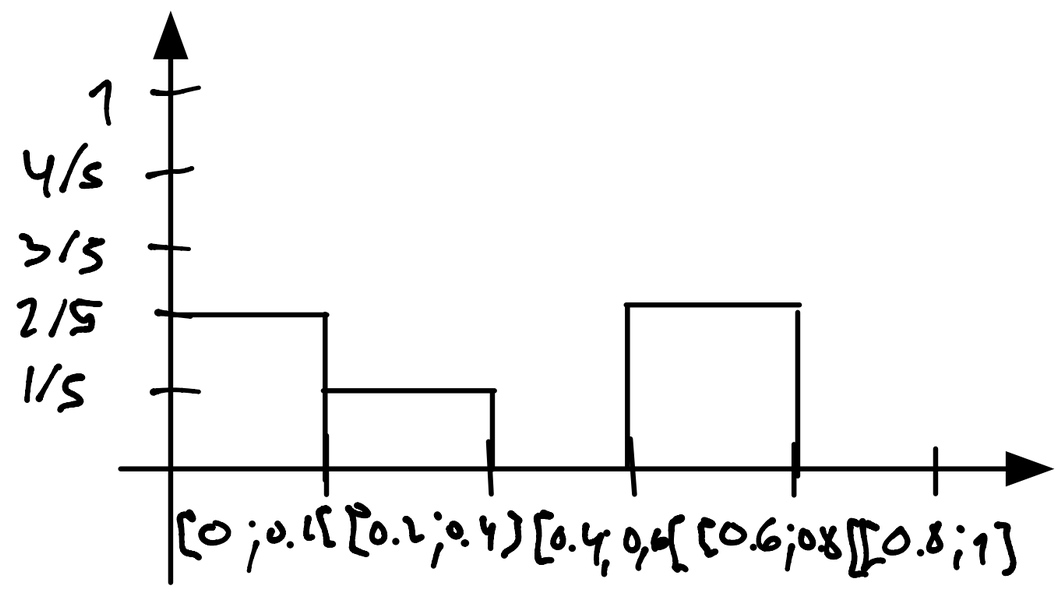
$p(0{,}2 \leq y_1 < 0{,}4 \mid y_{out} = A) = 1/5$

$p(0{,}4 \leq y_1 < 0{,}6 \mid y_{out} = A) = 0$

$p(0{,}6 \leq y_1 < 0{,}8 \mid y_{out} = A) = 2/5$

$p(0{,}8 \leq y_1 \leq 1 \mid y_{out} = A) = 0$

A



$p(0 \leq y_1 < 0{,}2 \mid y_{out} = B) = 1/4$

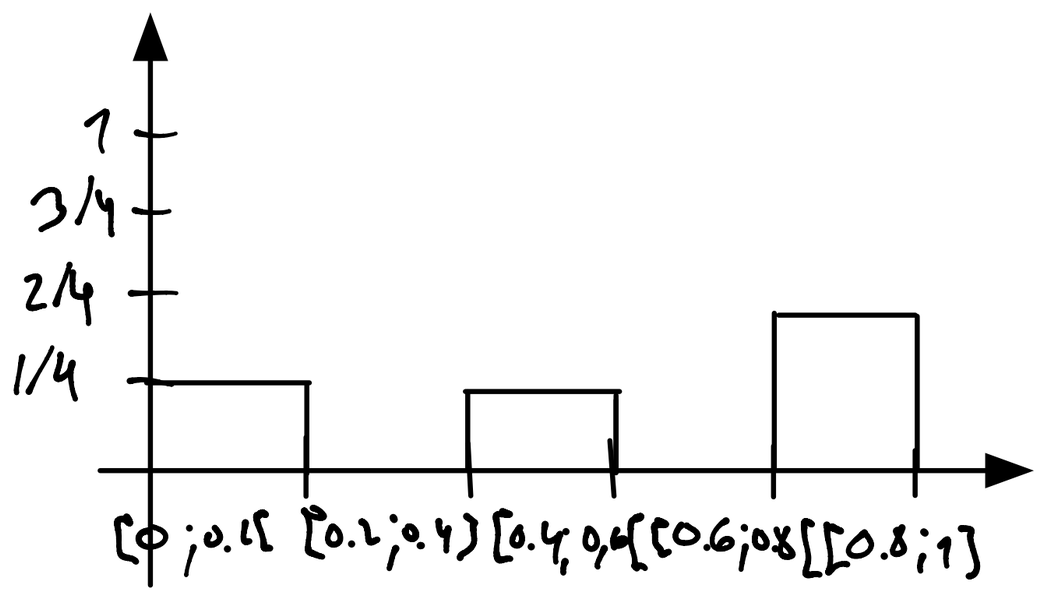
$p(0{,}2 \leq y_1 < 0{,}4 \mid y_{out} = B) = 0$

$p(0{,}4 \leq y_1 < 0{,}6 \mid y_{out} = B) = 1/4$

$p(0{,}6 \leq y_1 < 0{,}8 \mid y_{out} = B) = 0$

$p(0{,}8 \leq y_1 \leq 1 \mid y_{out} = B) = 2/4$

B



$p(0 \leq y_1 < 0{,}2 \mid y_{out} = C) = 0$

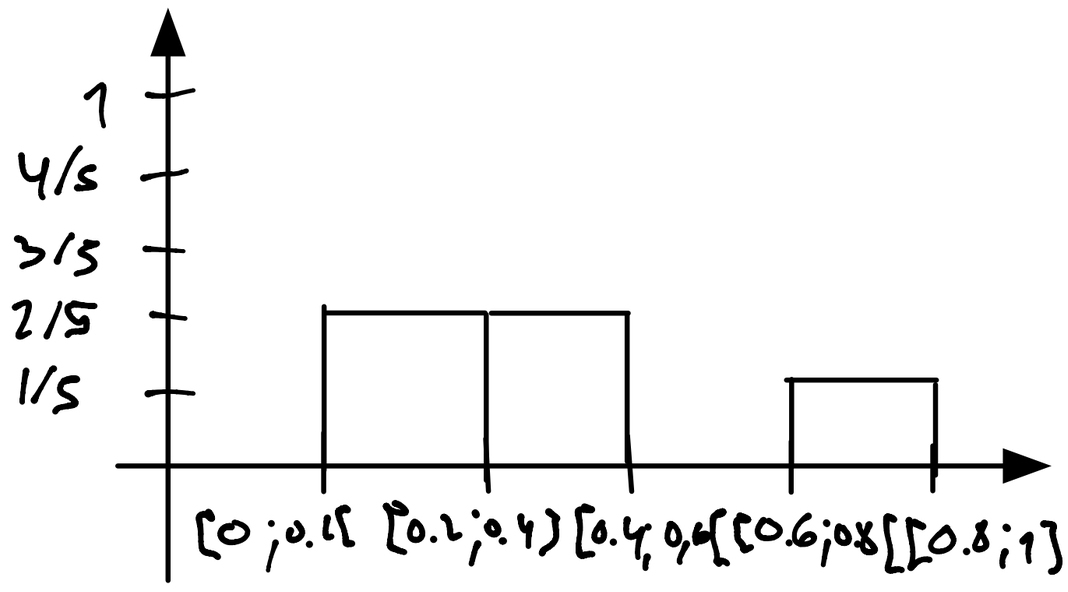
$p(0{,}2 \leq y_1 < 0{,}4 \mid y_{out} = C) = 2/5$

$p(0{,}4 \leq y_1 < 0{,}6 \mid y_{out} = C) = 2/5$

$p(0{,}6 \leq y_1 < 0{,}8 \mid y_{out} = C) = 0$

$p(0{,}8 \leq y_1 \leq 1 \mid y_{out} = C) = 1/5$

C



5. $Q_1 \rightarrow$ Mediana : $\frac{0,25 + 0,33}{2} = 0,29$

$Q_3 \rightarrow$ Mediana : $\frac{0,71 + 0,83}{2} = 0,77$

IQR $\rightarrow 0,77 - 0.29 = 0,48$

Intervalo = $[-0,43 ; 1,49]$

Para y_1 não há outliers

$y_2 \Rightarrow$ frequência (0) : $3/12 \simeq 25\%$
frequência (1) : $5/12 \simeq 41,67\%$.
frequência (2) : $4/12 \simeq 33,33\%$

$y_3 \Rightarrow$ frequência (0) : $4/12 \simeq 33,33\%$.
frequência (1) : $4/12 \simeq 33,33\%$.
frequência (2) : $4/12 \simeq 33,33\%$

$y_4 \Rightarrow$ frequência (0) : $3/12 \simeq 25\%$.
frequência (1) : $5/12 \simeq 41,67\%$.
frequência (2) : $4/12 \simeq 33,33\%$