

HW - 1

Pen - and - paper

$$1. a) \text{ Entropy : } E(\text{class}) = \left(\frac{2}{8} \log_2 \frac{2}{8} + \frac{3}{8} \log_2 \frac{3}{8} + \frac{3}{8} \log_2 \frac{3}{8} \right) \approx 1.5613$$

$$y_2 \rightarrow \text{Split : } y_2 = 1 \text{ \& } y_2 \neq 1$$

$$E(\text{class} | y_2 = 1) = -\frac{4}{8} \left(\frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right) = 0.75$$

$$E(\text{class} | y_2 \neq 1) = -\left(\frac{1}{8} (1 \log_2 1) + \frac{3}{8} \left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right) \approx 0.3444$$

$$IG(y_2) = 1.5613 - (0.3444 + 0.75) = 0.4669$$

$$y_4 \rightarrow \text{Split : } y_4 = 2 \text{ \& } y_4 \neq 2$$

$$E(\text{class} | y_4 = 2) = -\frac{4}{8} \left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) \approx 0.40569$$

$$E(\text{class} | y_4 \neq 2) = -\left(\frac{1}{8} (1 \log_2 1) + \frac{3}{8} \left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \right) \approx 0.3444$$

$$IG(y_4) = 1.5613 - (0.40569 + 0.3444) \approx 0.8113$$

Major IG : y_4

Case $y_4 = 2$

	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

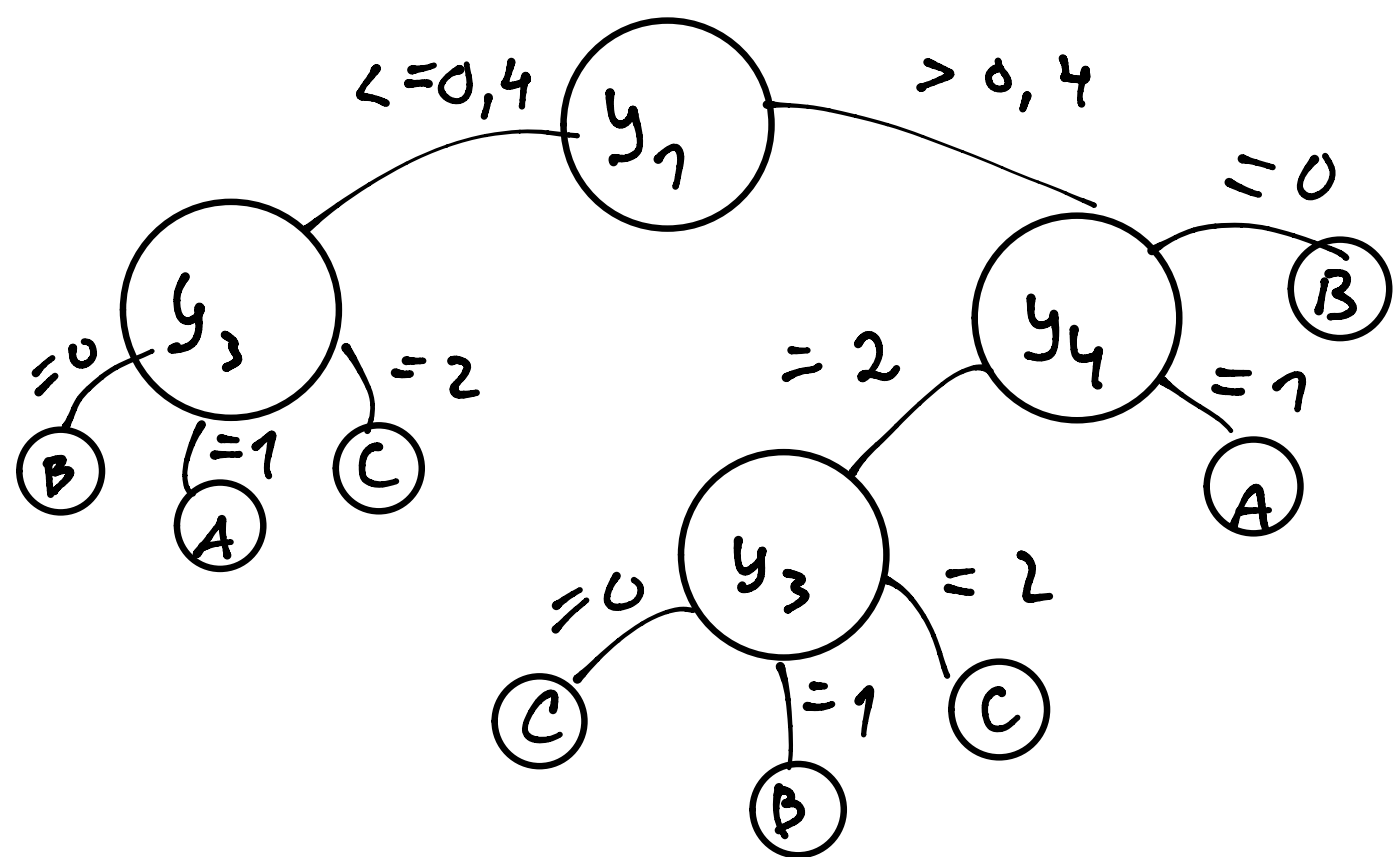
$$E(\text{class} | y_2) = \frac{1}{4} (1 \log_2 1) + \frac{3}{4} \left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) = 0.6887$$

$$E(\text{class}) = -\left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) \approx 0.8113$$

$$E(\text{class} | y_3) = \frac{1}{4} (1 \log_2 1) + \frac{2}{4} \left(\frac{1}{2} \log_2 \frac{1}{2} \times 2 \right) + (1 \log_2 1) \approx 0.5$$

$$IG(y_2) \approx -0.60757 \quad IG(y_3) \approx -0.41887$$

Arvore:



2.

Pred.

	True		
	A	B	C
A	4	2	0
B	0	1	0
C	0	1	4

	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

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

$$A \rightarrow \text{recall} : TP_A = 4 \quad FN = 0 \rightarrow 1$$

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

$$F_{1A} = \frac{2 \times 0,8 \times 1}{0,8 + 1} \approx 0,889$$

$$B \rightarrow \text{recall} : TP_B = 3 \quad FN = 1 \rightarrow 0,75$$

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

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

$$C \rightarrow \text{recall} : TP_C = 3 \quad FN = 1 \rightarrow 0,75$$

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

$$F_{1C} = \frac{2 \times 1 \times 0,75}{1,75} \approx 0,857$$

Lower F_1 score: B

4.

A

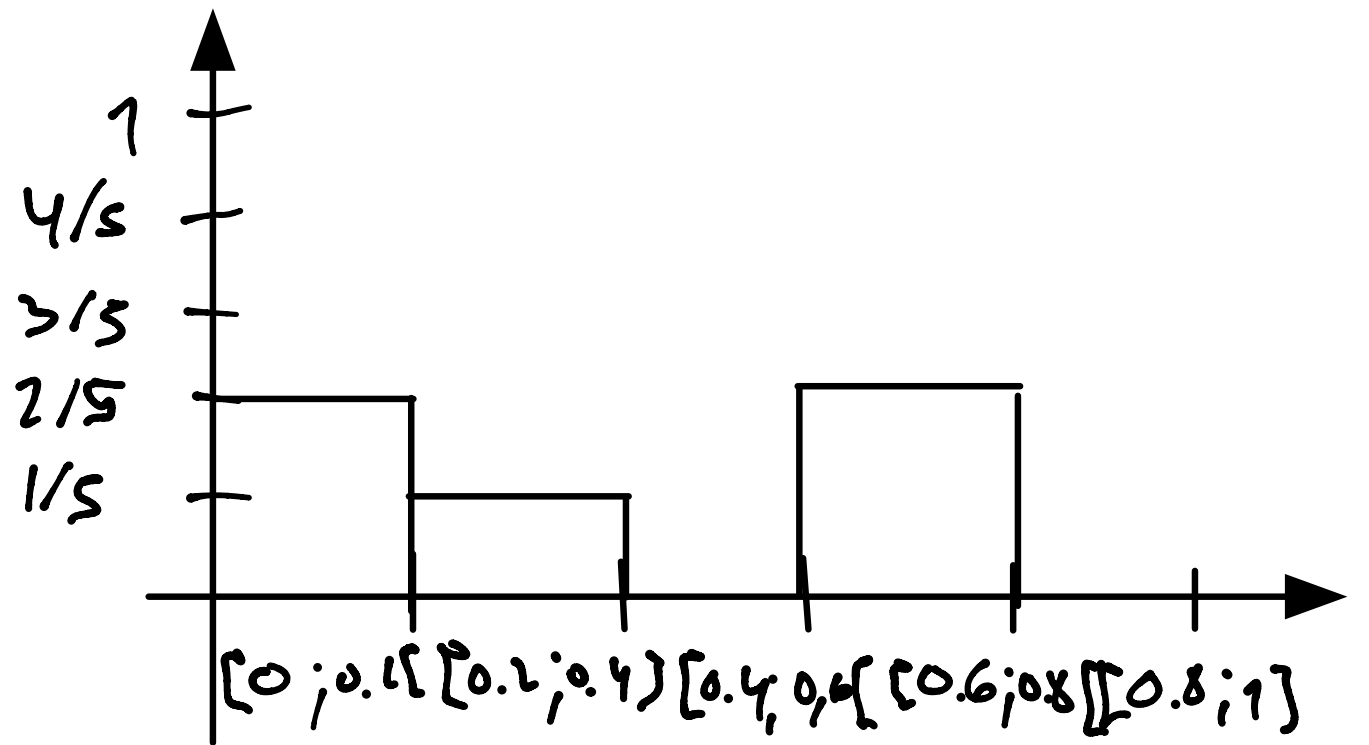
$$p(0 \leq y_1 < 0.2 \mid y_{out} = A) = 2/5$$

$$p(0.2 \leq y_1 < 0.4 \mid y_{out} = A) = 1/5$$

$$p(0.4 \leq y_1 < 0.6 \mid y_{out} = A) = 0$$

$$p(0.6 \leq y_1 < 0.8 \mid y_{out} = A) = 2/5$$

$$p(0.8 \leq y_1 \leq 1 \mid y_{out} = A) = 0$$



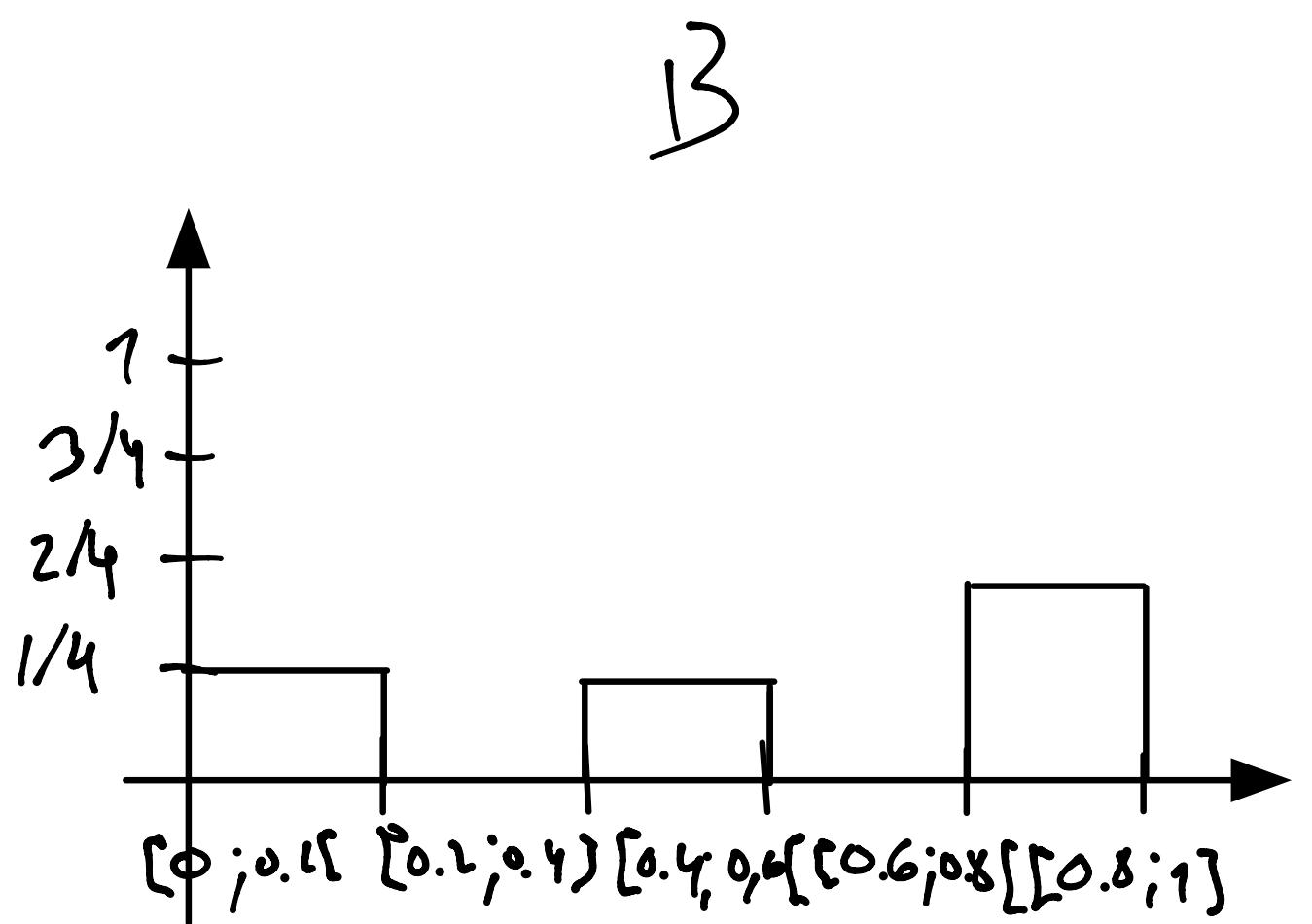
$$p(0 \leq y_1 < 0.2 \mid y_{out} = B) = 1/4$$

$$p(0.2 \leq y_1 < 0.4 \mid y_{out} = B) = 0$$

$$p(0.4 \leq y_1 < 0.6 \mid y_{out} = B) = 1/4$$

$$p(0.6 \leq y_1 < 0.8 \mid y_{out} = B) = 0$$

$$p(0.8 \leq y_1 \leq 1 \mid y_{out} = B) = 2/4$$



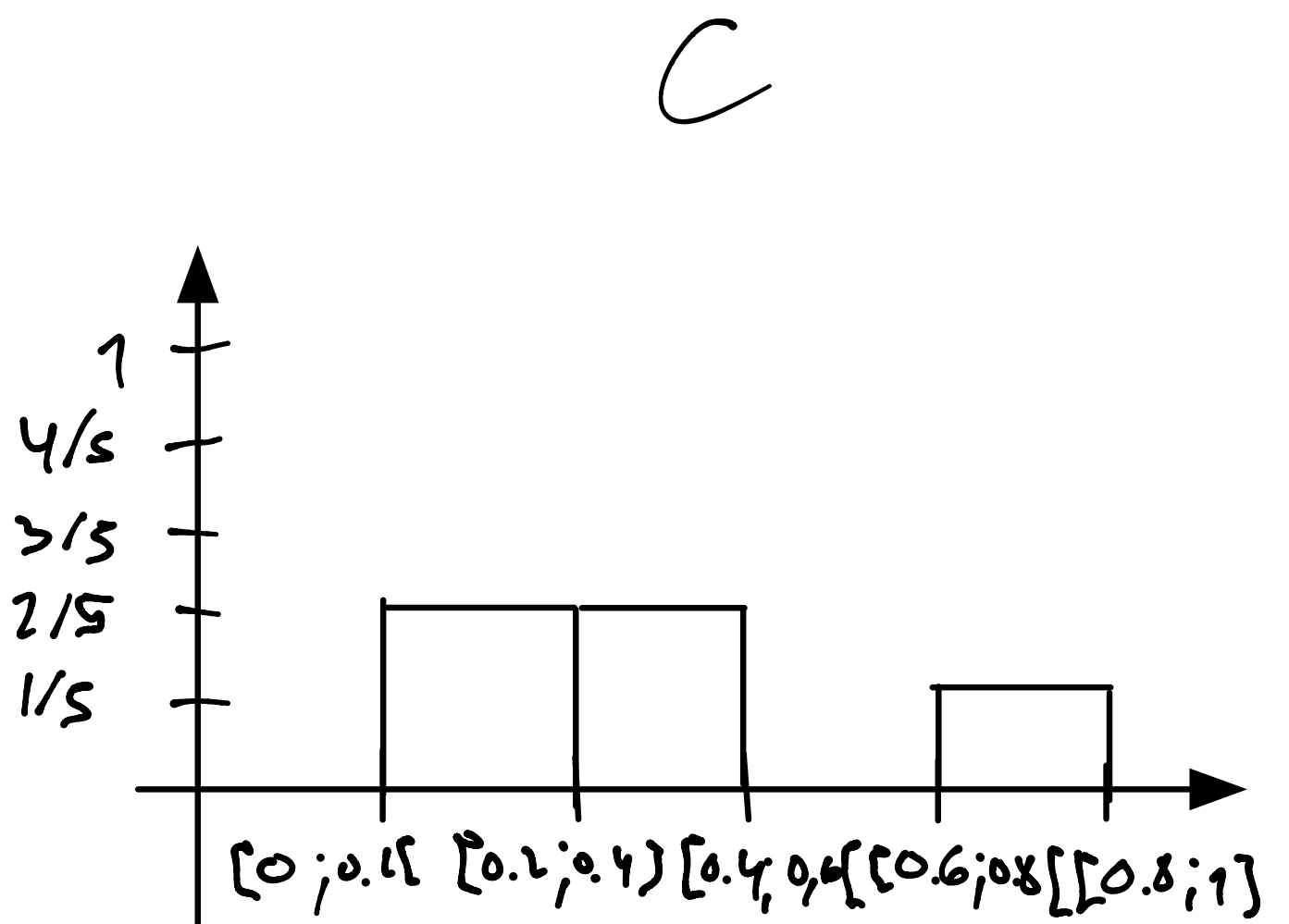
$$p(0 \leq y_1 < 0.2 \mid y_{out} = C) = 0$$

$$p(0.2 \leq y_1 < 0.4 \mid y_{out} = C) = 2/5$$

$$p(0.4 \leq y_1 < 0.6 \mid y_{out} = C) = 2/5$$

$$p(0.6 \leq y_1 < 0.8 \mid y_{out} = C) = 0$$

$$p(0.8 \leq y_1 \leq 1 \mid y_{out} = C) = 1/5$$



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

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

$$IQR \rightarrow 0,77 - 0,29 = 0,48$$

$$\text{Intervalo} = [-0,43; 1,49]$$

Para y_1 não há outliers

$$y_2 \rightarrow \text{frequência}(0) : 3/12 \approx 25\%$$

$$\text{frequência}(1) : 5/12 \approx 41,67\%$$

$$\text{frequência}(2) : 4/12 \approx 33,33\%$$

$$y_3 \rightarrow \text{frequência}(0) : 4/12 \approx 33,33\%$$

$$\text{frequência}(1) : 4/12 \approx 33,33\%$$

$$\text{frequência}(2) : 4/12 \approx 33,33\%$$

$$y_4 \rightarrow \text{frequência}(0) : 3/12 \approx 25\%$$

$$\text{frequência}(1) : 5/12 \approx 41,67\%$$

$$\text{frequência}(2) : 4/12 \approx 33,33\%$$