

# Homework I

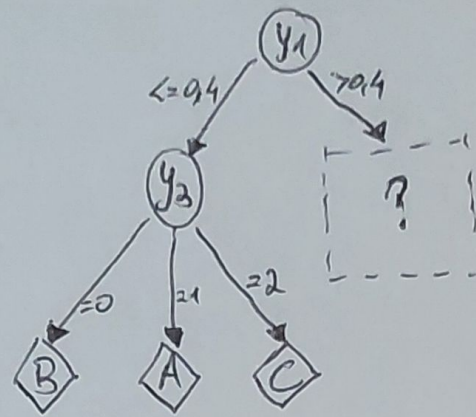
Kadalela Yury 110206  
Kadalela Rota 110355

1.  $IG(y_i) = H(y_{out}) - H(y_{out} | y_i)$

$$H(y) = - \sum_{i=1}^n p(y_i) \log_2 [p(y_i)]$$

•  $y_1 > 0.4$ :

D	$y_1$	$y_2$	$y_3$	$y_4$	$y_{out}$
$x_5$	0.45	1	0	2	C
$x_6$	0.52	0	0	0	B
$x_7$	0.58	2	1	2	C
$x_8$	0.62	1	0	1	A
$x_9$	0.71	1	2	1	A
$x_{10}$	0.83	1	2	1	B
$x_{11}$	0.90	2	1	2	B
$x_{12}$	0.95	2	2	2	C



•  $p(A) = \frac{2}{8} = \frac{1}{4}$

•  $p(B) = \frac{3}{8}$

•  $p(C) = \frac{3}{8}$

•  $H(y_{out}) = -p(A) \log_2 [p(A)] - p(B) \log_2 [p(B)] - p(C) \log_2 [p(C)] =$

$$= -\frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{3}{8} \log_2 \left(\frac{3}{8}\right) - \frac{3}{8} \log_2 \left(\frac{3}{8}\right) \approx 1.561$$

•  $H(y_{out} | y_2) = p(0) H(y_{out} | y_2=0) + p(1) H(y_{out} | y_2=1) + p(2) H(y_{out} | y_2=2) =$

$$= \frac{1}{8} \times 0 + \frac{1}{2} \left[ -\frac{1}{2} \log_2 \left(\frac{1}{2}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{1}{4} \log_2 \left(\frac{1}{4}\right) \right] + \frac{3}{8} \left[ -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{2}{3} \log_2 \left(\frac{2}{3}\right) \right] \approx 1.094$$

•  $H(y_{out} | y_3) = p(0) H(y_{out} | y_3=0) + p(1) H(y_{out} | y_3=1) + p(2) H(y_{out} | y_3=2) =$

$$= \frac{3}{8} \left[ -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) \right] + \frac{1}{4} \times 1 + \frac{3}{8} \left[ -\frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) \right] \approx 1.439$$

•  $H(y_{out} | y_4) = p(0) H(y_{out} | y_4=0) + p(1) H(y_{out} | y_4=1) + p(2) H(y_{out} | y_4=2) =$

$$= \frac{1}{8} \times 0 + \frac{3}{8} \left[ -\frac{2}{3} \log_2 \left(\frac{2}{3}\right) - \frac{1}{3} \log_2 \left(\frac{1}{3}\right) \right] + \frac{1}{2} \left[ -\frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{3}{4} \log_2 \left(\frac{3}{4}\right) \right] = 0.75$$

•  $IG(y_2) = H(y_{out}) - H(y_{out} | y_2) = 1.561 - 1.094 = 0.467$

$IG(y_4) > IG(y_2) > IG(y_3)$

•  $IG(y_3) = H(y_{out}) - H(y_{out} | y_3) = 1.561 - 1.439 = 0.122$

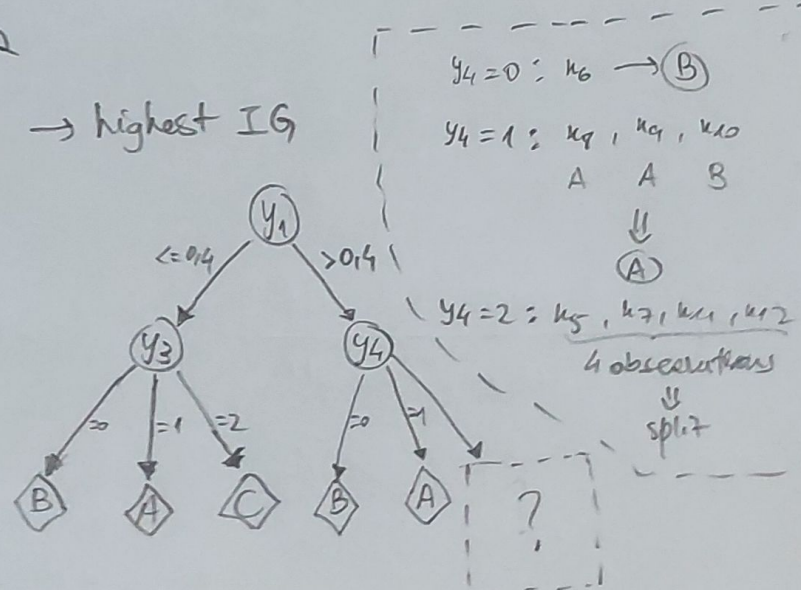
•  $IG(y_4) = H(y_{out}) - H(y_{out} | y_4) = 1.561 - 0.75 = 0.811 \rightarrow \text{highest } IG$

•  $y_1 > 0.4, y_4 = 2$ :

D	$y_1$	$y_2$	$y_3$	$y_4$	$y_{out}$
$x_5$	0.45	1	0	2	C
$x_7$	0.58	2	1	2	C
$x_{11}$	0.90	2	1	2	B
$x_{12}$	0.95	2	2	2	C

•  $p(B) = \frac{1}{4}$

•  $p(C) = \frac{3}{4}$





$$H(y_{out}) = -p(A) \log_2[p(A)] - p(B) \log_2[p(B)] - p(C) \log_2[p(C)] =$$

$$= 0 - \frac{1}{4} \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \log_2\left(\frac{3}{4}\right) \approx 0,811$$

$$H(y_{out} | y_2) = p(0) H(y_{out} | y_2=0) + p(1) H(y_{out} | y_2=1) + p(2) H(y_{out} | y_2=2) =$$

$$= 0 + \frac{1}{4} \times 0 + \frac{3}{4} \times \left[ -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right) \right] \approx 0,689$$

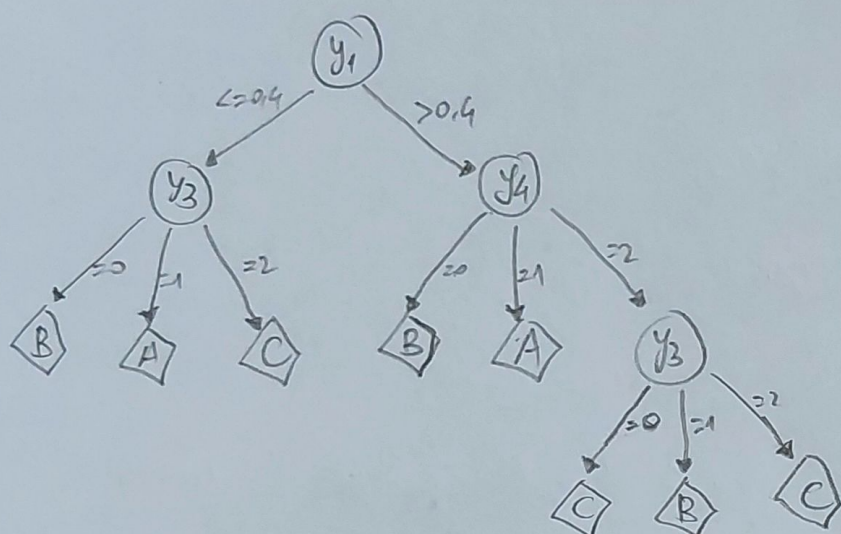
$$H(y_{out} | y_3) = p(0) H(y_{out} | y_3=0) + p(1) H(y_{out} | y_3=1) + p(2) H(y_{out} | y_3=2) =$$

$$= \frac{1}{4} \times 0 + \frac{1}{2} \times 1 + \frac{1}{4} \times 0 = \frac{1}{2} = 0,5$$

$$IG(y_2) = H(y_{out}) - H(y_{out} | y_2) = 0,811 - 0,689 = 0,122$$

$$IG(y_3) > IG(y_2)$$

$$IG(y_3) = H(y_{out}) - H(y_{out} | y_3) = 0,811 - 0,5 = 0,311 \rightarrow \text{highest } IG$$



$$y_3 = 0: x_5 \rightarrow \textcircled{C}$$

$$y_3 = 1: x_7, x_{11}$$

$$C \quad B \Rightarrow \textcircled{B} \quad \text{by ascending alphabetical}$$

$$y_3 = 2: x_{12} \rightarrow \textcircled{C}$$

2.	D	$y_{out}(\text{true})$	$y_{out}(\text{predict})$	
	$x_1$	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
	$x_8$	A	A	✓
	$x_9$	A	A	✓
	$x_{10}$	B	A	x
	$x_{11}$	B	B	✓
	$x_{12}$	C	C	✓

Training Confusion Matrix

	TRUE			
	A	B	C	
Predict	A	4	1	0
	B	0	3	1
	C	0	0	3

$$3. F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Recall} = \frac{TP}{P} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

Ⓐ  $TP = 4$   
 $TN = 3 + 1 + 3 = 7$   
 $FP = 1$   
 $FN = 0$

Ⓑ  $TP = 3$   
 $TN = 4 + 3 = 7$   
 $FP = 1$   
 $FN = 1$

Ⓒ  $TP = 3$   
 $TN = 4 + 1 + 3 = 8$   
 $FP = 0$   
 $FN = 1$



P - precision R - recall

(A)  $P = \frac{TP}{TP+FP} = \frac{4}{4+1} = \frac{4}{5}$

$R = \frac{TP}{TP+FN} = \frac{4}{4} = 1$

$F_1 = \frac{2 \cdot \frac{4}{5} \cdot 1}{\frac{4}{5} + 1} = \frac{\frac{8}{5}}{\frac{9}{5}} = \frac{8}{9}$

(B)  $P = \frac{TP}{TP+FP} = \frac{3}{3+1} = \frac{3}{4}$

$R = \frac{TP}{TP+FN} = \frac{3}{3+1} = \frac{3}{4}$

$F_1 = \frac{2 \cdot \frac{3}{4} \cdot \frac{3}{4}}{\frac{3}{4} + \frac{3}{4}} = \frac{\frac{18}{16}}{\frac{6}{4}} = \frac{18 \cdot 4}{16 \cdot 6} = \frac{3}{4}$

(C)  $P = \frac{TP}{TP+FP} = \frac{3}{3} = 1$

$R = \frac{TP}{TP+FN} = \frac{3}{3+1} = \frac{3}{4}$

$F_1 = \frac{2 \cdot 1 \cdot \frac{3}{4}}{1 + \frac{3}{4}} = \frac{\frac{6}{4}}{\frac{7}{4}} = \frac{6}{7}$

class (B) has the lowest training  $F_1$  score.  $\left( \frac{3}{4} < \frac{6}{7} < \frac{8}{9} \right)$

4.

D	$y_1$	$y_{out}$
$x_{13}$	0,02	A
$x_1$	0,12	A
$x_2$	0,18	B
$x_3$	0,25	C
$x_{14}$	0,27	C
$x_4$	0,33	A
$x_5$	0,45	C
$x_6$	0,52	B
$x_7$	0,58	C
$x_8$	0,62	A
$x_9$	0,71	A
$x_{10}$	0,83	B
$x_{11}$	0,90	B
$x_{12}$	0,95	C

5 equally bins in  $[0,1]$ :

$\hookrightarrow [0; 0,2] : 3$

$\hookrightarrow ]0,2; 0,4] : 3$

$\hookrightarrow ]0,4; 0,6] : 3$

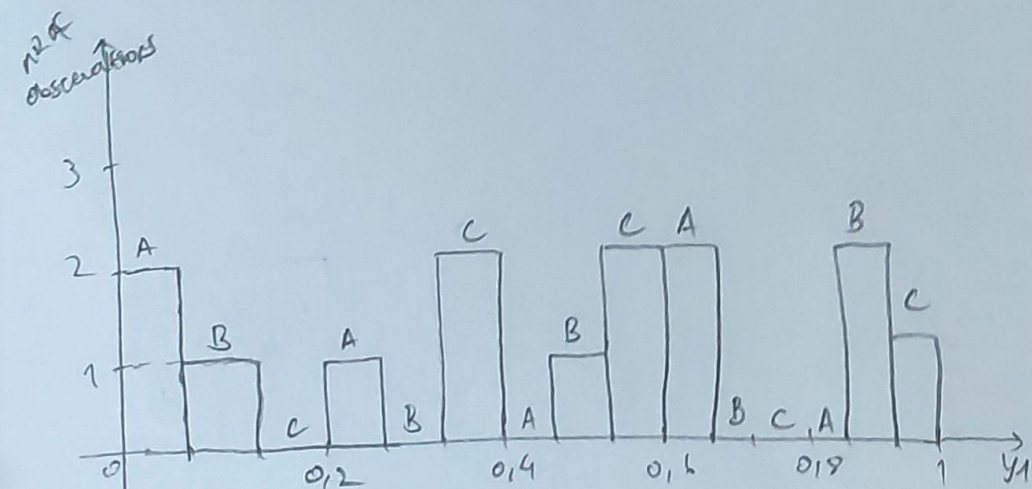
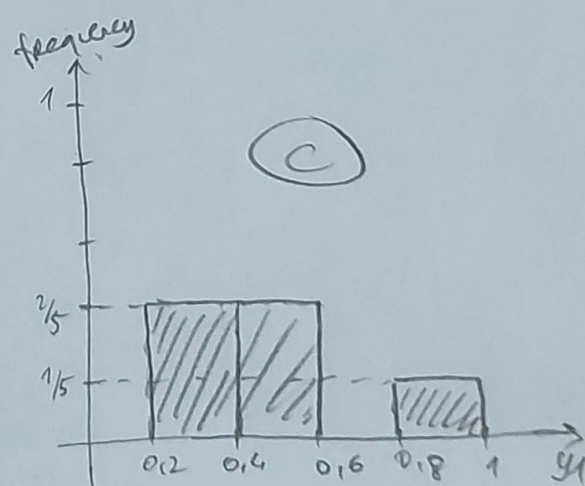
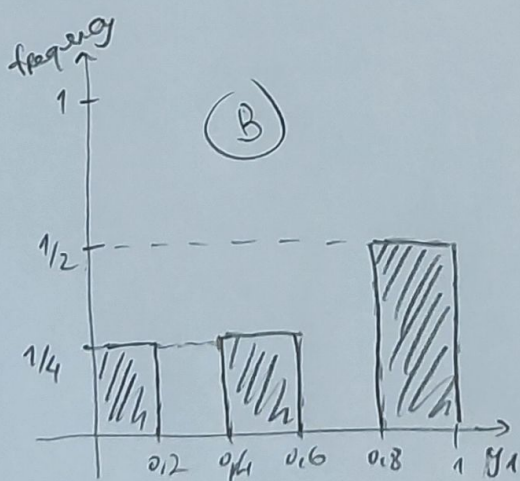
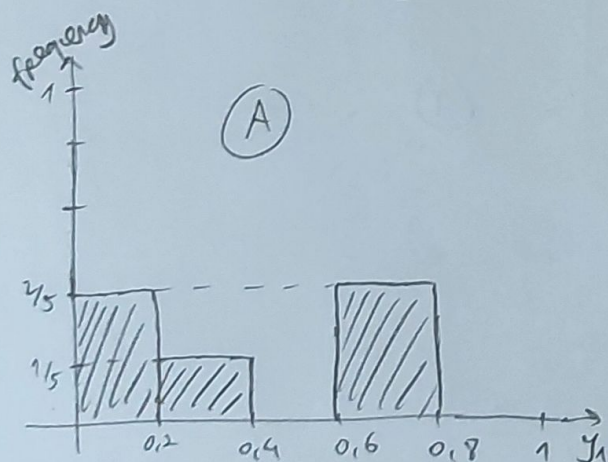
$\hookrightarrow ]0,6; 0,8] : 2$

$\hookrightarrow ]0,8; 1] : 3$

$n = 14$

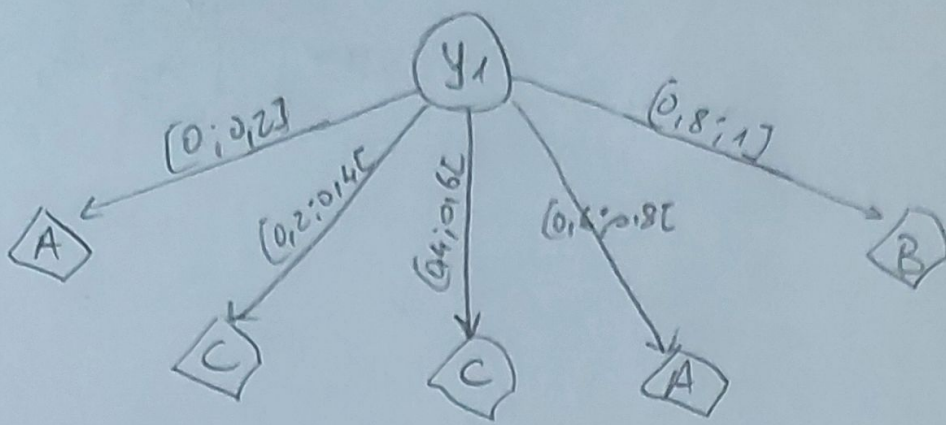
A (5)	B (4)	C (5)
$\frac{2}{5}$	$\frac{1}{4}$	0
$\frac{1}{5}$	0	$\frac{2}{5}$
0	$\frac{1}{4}$	$\frac{2}{5}$
$\frac{2}{5}$	0	0
0	$\frac{1}{2}$	$\frac{1}{5}$

(frequency of the classes in each bin)



classes per bin:  
[A; C; C; A; B]





5.  $y_1: 0,12 \quad 0,18 \quad 0,25 \mid 0,33 \quad 0,45 \quad 0,52 \mid 0,58 \quad 0,62 \quad 0,71 \mid 0,83 \quad 0,90 \quad 0,95$

$$Q_2 = \frac{0,52 + 0,58}{2} = 0,55$$

$$Q_1 = \frac{0,25 + 0,33}{2} = 0,29$$

$$IQR = Q_3 - Q_1 = 0,77 - 0,29 = 0,48$$

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

$$\text{Bounds} = [Q_1 - 1,5 \times IQR; Q_3 + 1,5 \times IQR] =$$

$$= [0,29 - 1,5 \times 0,48; 0,77 + 1,5 \times 0,48] =$$

$$= [-0,43; 1,49] \longrightarrow \text{No outliers.}$$