



# Aprendizagem - HomeWork 1

Pedro Curvo (ist1102716)

Salvador Torpes (ist1102474)

1º Semestre - 23/24

## 1 Dataset

Considering dataset D:

D	$y_1$	$y_2$	$y_3$	$y_4$	$y_{out}$
$x_1$	0.24	1	1	0	A
$x_2$	0.06	2	0	0	B
$x_3$	0.04	0	0	0	B
$x_4$	0.36	0	2	1	C
$x_5$	0.32	0	0	2	C
$x_6$	0.68	2	2	1	A
$x_7$	0.90	0	1	2	A
$x_8$	0.76	2	2	0	A
$x_9$	0.46	1	1	1	B
$x_{10}$	0.62	0	0	1	B
$x_{11}$	0.44	1	2	2	C
$x_{12}$	0.52	0	2	0	C

Tabela 1: Dataset D

## 2 Exercício 1.

De modo a corretamente completar a árvore de decisão, é necessário calcular o Information gain (IG) da variável de output  $y_{out}$  condicionada a cada uma das variáveis  $y_2$ ,  $y_3$  e  $y_4$ :

$$IG(y_{out}|y_2) = H(y_{out}) - H(y_{out}|y_2)$$

$$H(y_{out}) = \left( - \sum_{i=1}^3 p_{out_i} (\log_2 p_{out_i}) \right) = - \left( \frac{4}{12} \log_2 \left( \frac{4}{12} \right) + \frac{4}{12} \log_2 \left( \frac{4}{12} \right) + \frac{4}{12} \log_2 \left( \frac{4}{12} \right) \right) = 1.58496$$

$$H(y_{out}|y_2) = \sum_{i=0}^2 p_{y_2=i} H(y_{out}|y_2 = i)$$

D	$y_2$	$y_{out}$
$x_3$	0	B
$x_4$	0	C
$x_5$	0	C
$x_7$	0	A
$x_{10}$	0	B
$x_{12}$	0	C

Tabela 2: Dataset D com  $y_2 = 0$

Tabela apenas com os dados que verificam  $y_2 = 0$ :

$$H(y_{out}|y_2 = 0) = - \left( \frac{1}{6} \log_2 \left( \frac{1}{6} \right) + \frac{2}{6} \log_2 \left( \frac{2}{6} \right) + \frac{1}{2} \log_2 \left( \frac{1}{2} \right) \right) = 1.45915$$

Tabela apenas com os dados que verificam  $y_2 = 1$ :

D	$y_2$	$y_{out}$
$x_1$	1	A
$x_9$	1	B
$x_{11}$	1	C

Tabela 3: Dataset D com  $y_2 = 1$

$$H(y_{out}|y_2 = 1) = - \left( \frac{1}{3} \log_2 \left( \frac{1}{3} \right) + \frac{1}{3} \log_2 \left( \frac{1}{3} \right) + \frac{1}{3} \log_2 \left( \frac{1}{3} \right) \right) = 1.58496$$

Tabela apenas com os dados que verificam  $y_2 = 2$ :

D	$y_2$	$y_{out}$
$x_2$	2	B
$x_6$	2	A
$x_8$	2	A

Tabela 4: Dataset D com  $y_2 = 2$

$$H(y_{out}|y_2 = 2) = - \left( \frac{1}{3} \log_2 \left( \frac{1}{3} \right) + \frac{2}{3} \log_2 \left( \frac{2}{3} \right) \right) = 0.9183$$

Assim, podemos calcular a entropia de  $y_{out}$  condicionada a  $y_2$ :

$$\begin{aligned} H(y_{out}|y_2) &= \frac{6}{12} H(y_{out}|y_2 = 0) + \frac{3}{12} H(y_{out}|y_2 = 1) + \frac{3}{12} H(y_{out}|y_2 = 2) = \\ &= \frac{6}{12} \times 1.45915 + \frac{3}{12} \times 1.58496 + \frac{3}{12} \times 0.9183 = 1.33333 \end{aligned}$$