TÉCNICO LISBOA

Aprendizagem 2022/23
Homework I - Group 66

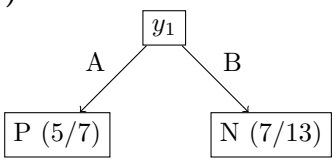
João Cardoso, 99251
José João Ferreira, 99259

I. Pen-and-paper

1)

		Real	
		P	N
Predicted	P	8	4
	N	3	5

2)



		Real	
		P	N
Predicted	P	5	2
	N	5	7

$$Precision = \frac{TP}{TP+FP} = \frac{5}{5+2} = \frac{5}{7}$$
$$Recall = \frac{TP}{TP+FN} = \frac{5}{5+5} = \frac{1}{2}$$

$$F_{\beta=1} = \frac{1}{\frac{1}{2Precision} + \frac{1}{2Recall}} = \frac{1}{\frac{7}{10} + \frac{2}{2}} = \frac{1}{\frac{17}{10}} = \frac{10}{17} \approx 0.588235$$

3) Some values of y_2, given $y_1 = A$, might be missing or corrupted. Another reason could be a very low information gain about y_{out} of the variable y_2, given $y_1 = A$.

4)

y_1	y_2	y_{out}
A		P
A		P
A		P
A		P
A		P
A		N
A		N
B	>2	P
B	>2	P
B	>2	P
B	>2	N
B	>2	N
B	>2	N
B	>2	N
B	>2	N
B	≤2	P
B	≤2	P
B	≤2	P
B	≤2	N
B	≤2	N

$$E(y) = -\sum_{v \in y} P(v) \log_2(P(v))$$

$$E(y_{out}) = -\frac{11}{20}\log_2(\frac{11}{20}) - \frac{9}{20}\log_2(\frac{9}{20}) \approx 0.992774$$

$$E(y_{out}|y_1) = \frac{7}{20}[-\frac{5}{7}\log_2(\frac{5}{7}) - \frac{2}{7}\log_2(\frac{2}{7})] + \frac{13}{20}[-\frac{7}{13}\log_2(\frac{7}{13}) - \frac{6}{13}\log_2(\frac{6}{13})] \approx 0.949315$$

$$\begin{aligned} IG(y_{out}|y_1) &= E(y_{out}) - E(y_{out}|y_1) \\ &= 0.992774 - 0.949315 \\ &= 0.043459 \end{aligned}$$

II. Programming

1)

2)

III. Appendix

This text uses a different font typeface. Paste your programming code here using Consolas 9pt or 10pt. Use highlighted or colored text to facilitate the analysis by your faculty hosts.