



Aprendizagem 2022

Lab 2: Decision Trees

Prof. Rui Henriques

Practical exercises

1. Consider the following testing targets, z , and the corresponding predictions, \hat{z} , by a decision tree:

$$z = [A A A B B B C C C C]$$

$$\hat{z} = [B B A C B A C A B C]$$

- a) Draw the confusion matrix

		<i>true</i>		
		A	B	C
<i>predicted</i>	A	1	1	1
	B	2	1	1
	C	0	1	2

- b) Compute the accuracy and sensitivity/recall per class

$$accuracy = 0.4, sensitivity_A = \frac{1}{3}, sensitivity_B = \frac{1}{3}, sensitivity_C = \frac{1}{2}$$

- c) Considering class C, identify precision and F1-measure

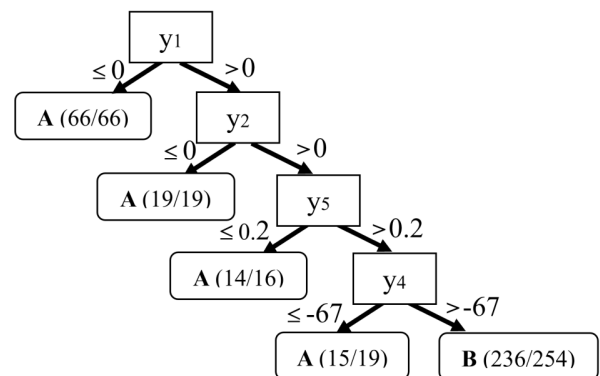
$$precision_C = \frac{2}{3}, F1_C = 0.57$$

- d) Identify the accuracy, sensitivity, and precision of the random classifier

$$accuracy_{random} = 0.(3), recall_{random}(A) = 0.(3), recall_{random}(B) = 0.(3), recall_{random}(C) = 0.(3)$$

$$precision_{random}(A) = 0.3, precision_{random}(B) = 0.3, precision_{random}(C) = 0.4$$

2. Consider a dataset composed by 374 records, described by 6 variables, and classified according to the decision tree below. Each leaf in the tree shows the label, number of classified records with the label, and total number of observations in the leaf. The positive class is the minority class.



- a) Compute the confusion matrix.

$$\#A = 66 + 19 + 14 + 15 + 18 = 132$$

$$\#B = 0 + 0 + 2 + 4 + 236 = 242$$

The minority class is the bad, hence is seen as positive.

		<i>Predicted</i>	
		P (B)	N (A)
<i>True</i>	P (B)	236	6
	N (A)	18	66 + 19 + 14 + 15 = 114

b) Compare the accuracy of the given tree versus a pruned tree with only two nodes.

Is there any evidence towards overfitting?

Considering training accuracy:

$$accuracy_{\text{depth}=4} = \frac{114 + 236}{114 + 236 + 6 + 18} = 0.936$$

$$accuracy_{\text{depth}=2} = \frac{(66 + 19) + (236 + 4 + 2)}{(66 + 19) + (236 + 4 + 2) + (14 + 15 + 18)} = 0.874$$

Without the testing accuracy, there is no sufficient evidence to assume the tree is prone to overfit input data.