

1. Consider the following dataset:

	y_1	y_2	y_3	z
x_1	a	a	a	+
x_2	c	b	c	+
x_3	c	a	c	+
x_4	b	a	a	-
x_5	a	b	c	-
x_6	b	b	c	-

Plot the learned decision tree using information gain (Shannon entropy). Show your calculations.

2. Show if a decision tree can learn the AND, OR and XOR logical functions. If so, plot the corresponding decision boundaries.

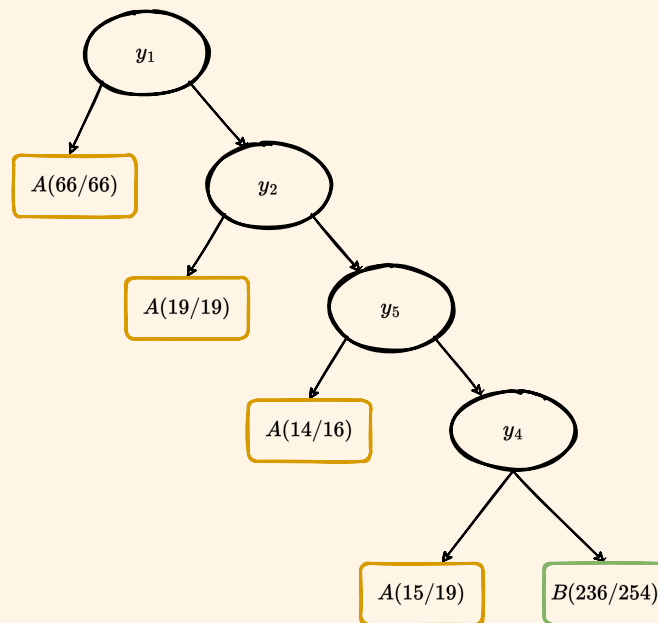
3. Consider the following testing targets, z , and the corresponding predictions, \hat{z} , by a decision tree:

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

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

- (a) Draw the confusion matrix.
- (b) Compute the accuracy and recall (sensitivity) for each class.
- (c) Regarding class C, identify its precision and F-measure.
- (d) identify the accuracy, sensitivity and precision of a random classifier.

4. Consider a dataset composed by 374 records, described by 6 variables, classified according to the following decision tree:



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.
- (b) Compare the accuracy of the given tree versus a pruned tree, with only two nodes. Is there any evidence towards overfitting?
- (c) Are decision trees learned from high-dimensional data susceptible to underfitting? Why does an ensemble of DTs minimize this problem?