P02 Decision Trees

Luis Sa-Couto[1] and Andreas Wichert[2]

INESC-ID, Instituto Superior Tecnico, Universidade de Lisboa
{luis.sa.couto,andreas.wichert}@tecnico.ulisboa.pt

1 Decision Trees

1) Consider the following data table:

F_1	F_2	F_3	O
a	a	a	$+$
c	b	c	$+$
c	a	c	$+$
b	a	a	$-$
a	b	c	$-$
b	b	c	$-$

a) Determine the whole decision tree using ID3 (**information gain**), taking "O" as the target. Show all steps.

2) Consider the following data table:

F_1	F_2	F_3	O
d	a	b	m
c	a	b	n
c	a	a	y
d	a	a	y
c	b	a	f
c	b	b	f

a) Compute the first attribute to be tested using ID3.

b) Complete the tree started in the previous question. There is no need to perform all computations. What do you need to take into account?

3) Show if a decision tree can learn the following logical functions and if so plot the corresponding decision boundaries.

a) AND

b) OR

c) XOR

2 Thinking Questions

Can you think why we want a good heuristic when building the tree? Why are usually smaller trees preferred?