Lecture 05 Cristiano Panighel 2058351

- Using the algorithm ID3, compute manually the decision tree corresponding to the task of realizing simple Boolean formulas (AND, OR, XOR) in n variables.

To compute the decision tree of each Boolean formula, I will use a specific case, i.e. with 3 variables but the proofs are easily extendable to cover the generall case as well.

AND

x_1	x_2	x_3	
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

$E(S) = -p_- \log(p_-) - p_+ \log(p_+) = -7/8 \cdot \log(7/8) - 1/8 \cdot \log(1/8) = 0.54$

Let's consider x_1

$E(S_{x_1=0}) = 4/4 \cdot \log(4/4) = 0$

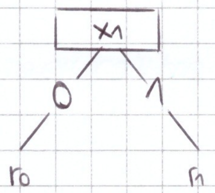
$E(S_{x_1=1}) = -1/4 \cdot \log(1/4) - 3/4 \cdot \log(3/4) = 0.81$

$G(S, x_1) = 0.54 - 1/2 \cdot (0.81) = 0.135$

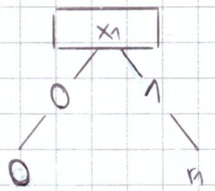
Same procedures and results are valid also for the other two attributes x_2 and x_3

$G(S, x_2) = G(S, x_3) = 0.135$

To choose the new attribute, I have to go with the one that maximizes the information gain G, but in this case, since all they have the same value, I choose randomly so I pick x_1.



- r_0: all the data samples with $x_1 = 0$ are labeled as 0, so there is no need to go on with computations and I obtain a leaf node.



- r_1: The new dataset decreases to those with $x_1 = 1$.

x_2	x_3	
0	0	0
0	1	0
1	0	0
1	1	1

$E(S) = -1/4 \log(1/4) - 3/4 \log(3/4) = 0.81$

Let's consider x_2

$E(S_{x_2=0}) = 0$

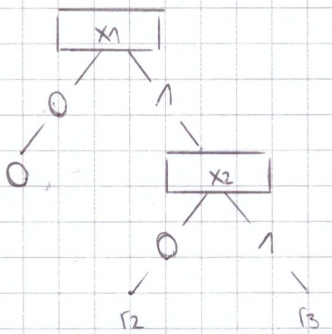
$E(S_{x_2=1}) = -1/2 \log(1/2) - 1/2 \log(1/2) = 1$

$G(S, x_2) = 0.81 - 1/2 (1) = 0.31$

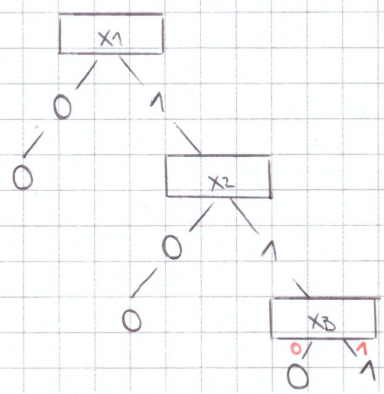
Same procedures and results also for x_3

$G(S, x_3) = 0.31$

Again I choose randomly between x_2 and x_3 since they both have the same value of information gain G, and I choose x_2



- r_2: same reasoning as r_0
- r_3: the last attribute is x_3 and it can have exactly two values.



This is the final DT for the Boolean formula AND

OR

x_1	x_2	x_3	
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

$E(S) = -p_- \cdot \log(p_-) - p_+ \cdot \log(p_+) = -1/8 \cdot \log(1/8) - 7/8 \log(7/8) = 0.54$

Let's consider x_1

$E(S_{x_1=0}) = -1/4 \cdot \log(1/4) - 3/4 \cdot \log(3/4) = 0.81$

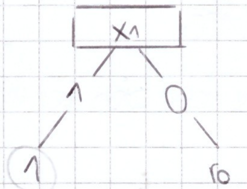
$E(S_{x_1=1}) = 0$

$G(S, x_1) = 0.54 - 1/2 \cdot (0.81) = 0.135$

Same results also for the other two attributes x_2 and x_3

$G(S, x_2) = G(S, x_3) = 0.135$

Since all they have the same value of information gain, I choose x_1 randomly.



all the instances with $x_1=1$ have label 1

- r_0: The new dataset decreases to those with $x_1 = 0$

x_2	x_3	
0	0	0
0	1	1
1	0	1
1	1	1

$E(S) = -1/4 \log(1/4) - 3/4 \log(3/4) = 0.81$

Let's consider x_2

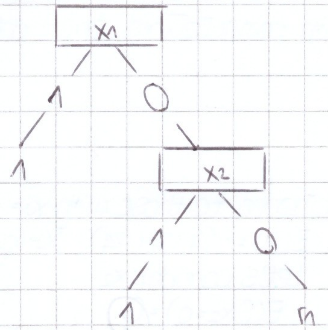
$E(S_{x_2=0}) = -1/2 \log(1/2) - 1/2 \log(1/2) = 1$

$E(S_{x_2=1}) = -2/2 \log(2/2) = 0$

$G(S, x_2) = 0.81 - 1/2(1) = 0.31$

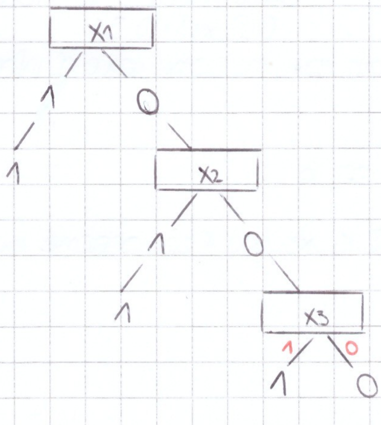
Same results also for x_3, $G(S, x_3) = 0.31$

Again I choose randomly between x_2 and x_3 since they both have the same value of information gain G, and I choose x_2.



all the instances with $x_2=1$ have label 1

- r_1: the remaining attribute is x_3 and it can have exactly two values.



This is the final DT for the Boolean formula OR

XOR

x_1	x_2	x_3	
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

$E(S) = -p_- \cdot \log(p_-) - p_+ \cdot \log(p_+) = -1/2 \log(1/2) - 1/2 \log(1/2) = 1$

Let's consider x_1

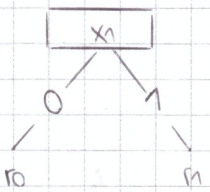
$E(S_{x_1=0}) = -1/2 \cdot \log(1/2) - 1/2 \cdot \log(1/2) = 1$

$E(S_{x_1=1}) = -1/2 \cdot \log(1/2) - 1/2 \cdot \log(1/2) = 1$

$G(S, x_1) = 1 - 1/2 \cdot (1) - 1/2 \cdot (1) = 0$

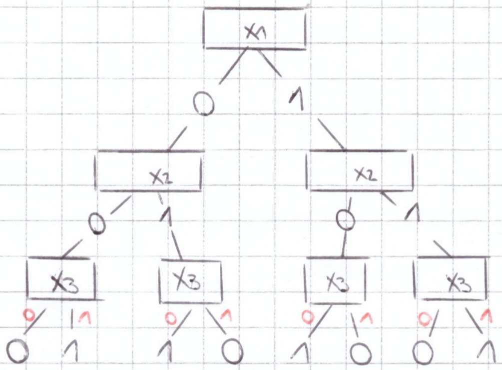
Same results for x_2 and x_3, $G(S,x_2) = G(S,x_3) = 0$

Since all they have the same of information gain, I choose x_1 randomly



- r_0 : If we try to decrease the dataset to those samples with $x_1 = 0$, the information gain for both x_2 and x_3 is again 0. $G(S,x_2) = G(S,x_3) = 0$
- r_1 : same reasoning also works for r_1.

In this situation, since for all the remaining attributes x_2 and x_3, the information gain is 0, so I need to build branches until data are all covered.



This is the final decision tree for the Boolean formula XOR. In this case, the DT cannot help us classify the samples efficiently. The reason is that it works based on information Gain which is always 0.