

(Adevărat sau Fals?)

Învățăm un arbore de decizie folosind algoritmul ID3 standard, fără pruning. Atributele de intrare  $(X_1, X_2, \dots, X_m)$  sunt categoricale, iar atributul de ieșire  $(Y)$  este de asemenea categoric.

Marcați cu A (adevărat) sau F (fals) fiecare din afirmațiile de mai jos și dați în fiecare caz o explicație succintă însoțită eventual de un exemplu sau contra-exemplu minimalist(!)

- Dacă  $IG(Y, X_i) = 0$ , atunci atributul  $X_i$  nu vă apăsă în arborele de decizie.
- Adâncimea maximă a arborelui de decizie este de cel mult  $m$ .  
Notă: Dacă arborele este format doar din nodul rădăcină, ceea ce corespunde cazului în care toate exemplele de antrenament sunt identic clasificate, atunci se consideră că adâncimea arborelui este 0.
- Dacă sunt  $R$  exemple de antrenament, atunci adâncimea maximă a arborelui de decizie este de cel mult  $1 + \log_2 R$ .

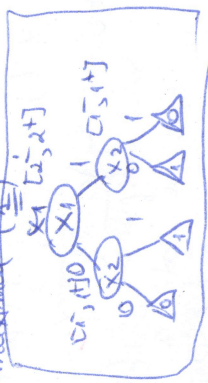
d. Dacă sunt  $R$  exemple de antrenament, iar unul dintre atributele de intrare are  $R$  valori distincte și ia valori  $v_1, \dots, v_R$  în mod injectiv (pe mulțimea formată de exemplele de antrenament), atunci arborele de decizie va avea adâncimea 0 sau 1.

a) Fals. Exemplu este arborele i3 realizat pt.  $\neg$  (Fals XOR).

exemplu:

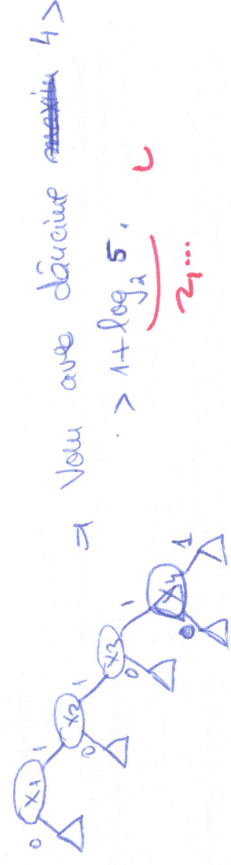
$X_1$	$X_2$	$Y$
0	0	0
0	1	1
1	0	1
1	1	0

La început, vom alege pe  $X_1$ , pe  $X_2$  în rădăcină, iar-ul fiind 0, dar mai apoi, pe fiecare dintre subarborii,  $K$ -ul este maximal ( $\frac{1}{2}$ ).

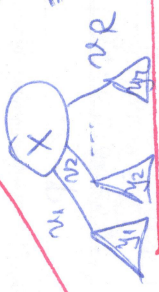


b) Adevărat. Deoarece atributele  $X_1, X_2, \dots, X_m$  sunt categorice, atunci pe ORICE drum de la rădăcină până la o frunză, vom putea întâlni cel mult o dată atributul  $X_i$  într-un nod,  $\forall i = 1, \dots, m$ .  $\Rightarrow$  orice drum de la rădăcină la o frunză are lungime maximă

-u  $\Rightarrow$  Adâncimea arborelui este maximă  $m$ .  
c) Fals. Considerăm cazul când  $R \gg m$ .  
de ex, dacă avem 5 exemple, vom putea avea arborele:



d) Adevărat. Fie  $X$ -atributul vizat.  
Dacă TOATE exemplele au aceeași etichetă  $y$ , atunci arborele de decizie va fi:  $\Delta_y$  ( $\Rightarrow$  adâncime 0).  
În caz contrar, i3 va alege atributul  $X$  în rădăcină (va avea i3 maxim  $\Rightarrow \log_2 R$ ). Arborele va arăta:  $\Rightarrow$  adâncime 1.



$\Rightarrow$  partitii simple  
deci, arborele de decizie are adâncime 0 sau 1.  $\Rightarrow$  LG maxim

NU heaparat

Poate să mai existe un arbore în LG maxim.  
Lună arde  $\Rightarrow H=0$  pt fiecare nodul rădăcină (nu adinec, dacă toate ex. au ac-si etichetă)  
fiec partitie nu are int. decât de tip  $\Rightarrow$  arbore  $\Rightarrow$  bbb este de adinec  
(într-un heap se atribu. este injectiv!)