

# Decision Tree

37.

(Algoritmul ID3: aplicare)

prelucrare de Liviu Ciortuz, după  
CMU, 2020 fall, Aarti Singh, Recitation

Fie următorul set de date, relativ la când anume un copil ieșe afară să se joace:

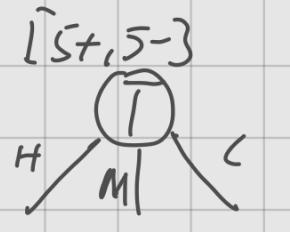
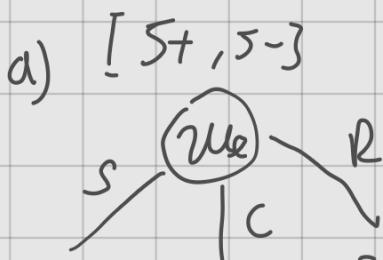
Day	Weather	Temperature	Humidity	Wind	Play
1	Sunny	Hot	High	Weak	No
2	Cloudy	Hot	High	Weak	Yes
3	Sunny	Mild	Normal	Strong	Yes
4	Cloudy	Mild	High	Strong	Yes
5	Rainy	Mild	High	Strong	No
6	Rainy	Cool	Normal	Strong	No
7	Rainy	Mild	High	Weak	Yes
8	Sunny	Hot	High	Strong	No
9	Cloudy	Hot	Normal	Weak	Yes
10	Rainy	Mild	High	Strong	No

În acest exercițiu vă cerem să elaborați arborele învățat de algoritmul ID3 pe acest set de date.

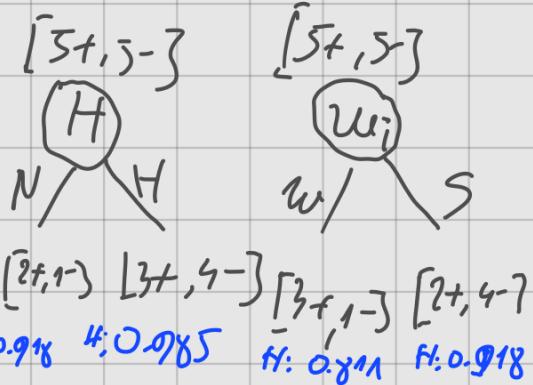
a. Stabiliti în mod riguros ce atribut va pune algoritmul ID3 în nodul rădăcină. Veți folosi următorul tabel, în care se dău entropiile mai multor variabile Bernoulli, determinate de valoarea parametrului  $p$ :

$p$	1/4	1/3	2/5	3/7
$H(p)$	0.8112	0.9182	0.9709	0.9852

b. Satisfaceti aceeași cerință, apelând acum la alte două metode: mai întâi raționamentul relațional („calitativ“) care a fost prezentat la Comentariul de la pag. 578 și apoi metoda expusă la problema 36 (vedeți în special Observația (2) de acolo).



$$0.9709$$



$$H[1+, 2-] = \frac{1}{3} \log_2 \frac{3}{2} + \frac{2}{3} \log_2 \frac{2}{1} = 0.918$$

$$H[2+, 1-] = \frac{2}{3} \log_2 \frac{2}{1} + \frac{1}{3} \log_2 \frac{1}{2} = 0.918$$

$$H[1+, 3-] = \frac{1}{3} \log_2 \frac{4}{3} + \frac{2}{3} \log_2 \frac{3}{2} = 0.811$$

$$H[3+, 4-] = \frac{3}{7} \log_2 \frac{7}{3} + \frac{4}{7} \log_2 \frac{3}{4} = 0.885$$

$$H[3+, 1-] = H[1+, 3-] = 0.811$$

$$H[2+, 4-] = \frac{1}{3} \log_2 \frac{3}{2} + \frac{2}{3} \log_2 \frac{2}{1} = 0.918$$

Calc. cîntigul de inf:

$$H(\text{Play}) = 1$$

$$IG(Play | W_e) = H(Play) - \underbrace{H(Play | W_e)}$$

$$\textcircled{a} \quad H(Play | W) = \frac{c+d}{a+b} H[S] + \frac{e+f}{a+b} H[C]$$

$$+ \frac{g+h}{a+b} H[R] = \frac{3}{10} \cdot 0.018 + \frac{3+0}{10} \cdot 0 + \frac{1+3}{10} \cdot 0.811 = \\ \approx 0.4002$$

$$\textcircled{b} \quad IG(Play | T) = H(Play) - H(Play | T)$$

$$= 1 - \left( \frac{2+2}{10} \cdot 1 + \frac{5}{10} \cdot H[M] + \dots \right) =$$

$$= 1 - \left( \frac{2}{5} + \frac{1}{2} \cdot 0.070 \right) \approx 0.115$$

$$\textcircled{c} \quad IG(Play | H) = H(Play) - H(Play | H) =$$

$$= 1 - \left( \frac{3}{10} \cdot 0.018 + \frac{7}{10} \cdot 0.085 \right) = 0.031$$

$$\textcircled{d} \quad IG(Play | W_i) = H(Play) - H(Play | W_i) =$$

$$= 1 - \left( \frac{4}{10} \cdot 0.011 + \frac{6}{10} \cdot 0.018 \right) =$$

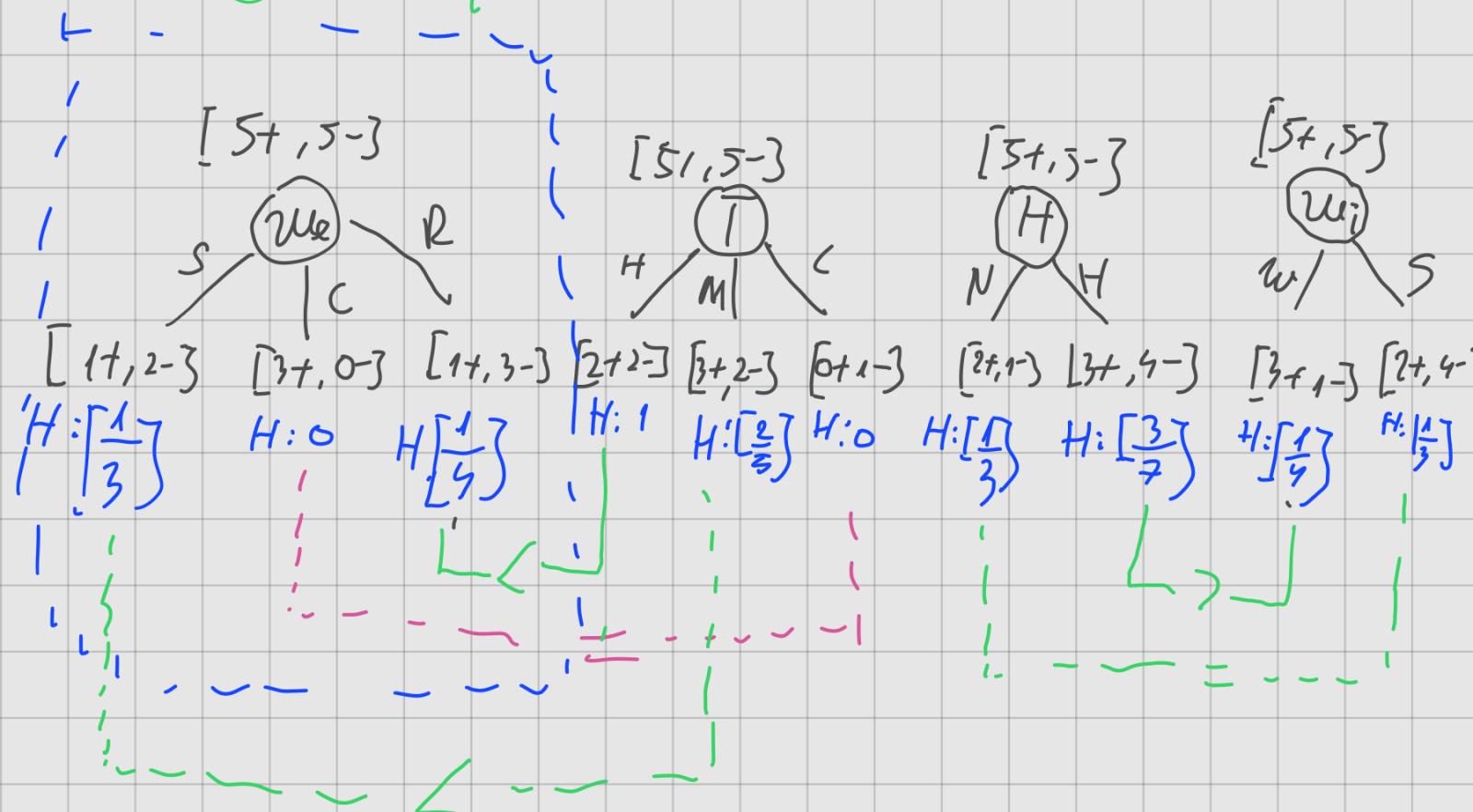
$$\approx 0.125$$

Cel mai bun mod rezolvării este We.

b)

Folosim alte 2 metode:

(1) Numerical calculation.



(2) Conform permisilor:

$$\frac{(c+d)^{c+d}}{c^c \cdot d^d} \cdot \frac{(e+f)^{e+f}}{e^e \cdot f^f}$$

$$\text{Ndl. } H_{0/W_e} = \frac{(1+2)^{1+2}}{1^1 \cdot 2^2} \cdot \frac{(1+3)^{1+3}}{1^1 \cdot 3^3} = \frac{\cancel{1}^2 \cdot \cancel{2}^3}{\cancel{1}^1 \cdot \cancel{3}^3} =$$

= 64

$$H_{0/IT} = \frac{(2+2)^{2+2}}{2^2 \cdot 2^2} \cdot \frac{(3+2)^{3+2}}{3^2 \cdot 2^2} = \frac{\cancel{2}^2 \cdot \cancel{2}^2}{\cancel{3}^3 \cdot \cancel{2}^2} \cdot \frac{\cancel{3}^3 \cdot \cancel{2}^2}{\cancel{2}^2 \cdot \cancel{3}^3} =$$

$$\frac{5^5}{3^2} = \frac{25 \cdot 25 \cdot 5}{9} = 347.2$$

$$H_{0/H} = \frac{(2+1)^{2+1}}{2^2 \cdot 1^1} \cdot \frac{(3+4)^{3+4}}{3^3 \cdot 4^4} = \frac{\cancel{2}^3 \cdot \cancel{2}^2}{\cancel{3}^2 \cdot \cancel{3}^3} = 804.24$$

$$H_{0/W_i} = \frac{(3+1)^{3+1}}{3^3 \cdot 1^1} \cdot \frac{(2+5)^{2+5}}{2^2 \cdot 4^4} = \frac{\cancel{3}^4 \cdot \cancel{1}^1}{\cancel{2}^3 \cdot \cancel{2}^2} \cdot \frac{\cancel{4}^4}{\cancel{4}^4} =$$

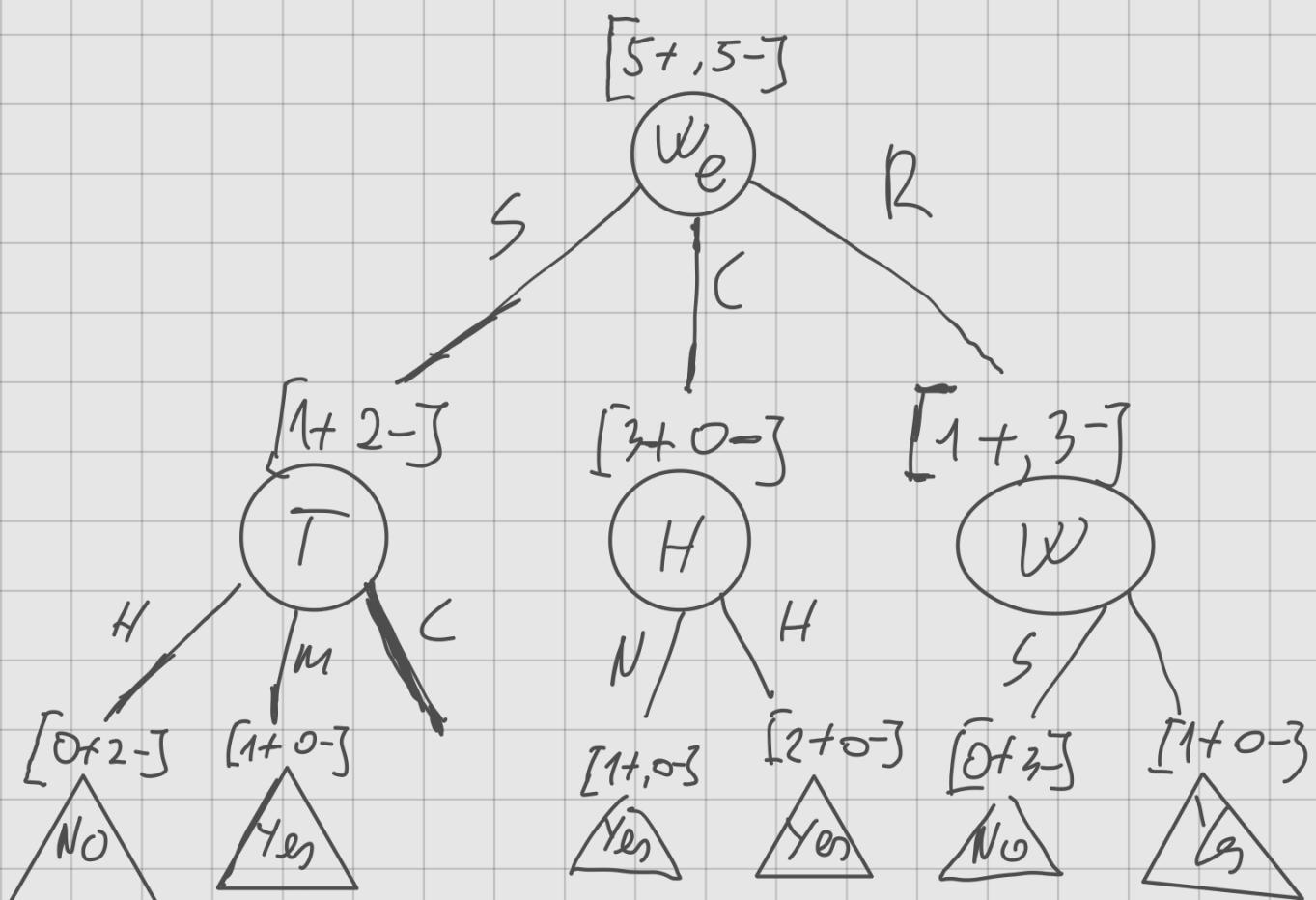
$$= 559.872$$

Oleas . Min comp. că rădăcina se  
dă ce cel mai mic produs .

c)

c. Stabiliti ce atribute pot fi puse in nodurile de test de pe nivelul 2 (adică, în descendenții nodului rădăcină), iar apoi completați arborele ID3.<sup>580</sup> Este oare arborele [elaborat de algoritmul] ID3 unic? Care este eroarea pe setul de date de antrenare?

Să stim că cel mai bun nod rădăcină este We.



Dacă lărgim nodul T nu trimitem S, pentru inputul „cool”, nu vom putea clasifica corect, decarece .

nu avem date în setul de antrenare care să primească

input  $V_e = \text{Sunny}$  și  $T = \text{cold}$ . Aceasta ar fi  
eroră în setul de antrenare.

- În acestă miercură nu se întâmplă dacă vom avea noapte

Typea noastră  $V_e = \text{Cloudy}$ .

Nu este unic, deoarece în același miercură

① multe schimbuli nodurile descendente

Intre ele.

38.

(Calcularea unor entropii;  
aplicarea algoritmului ID3)

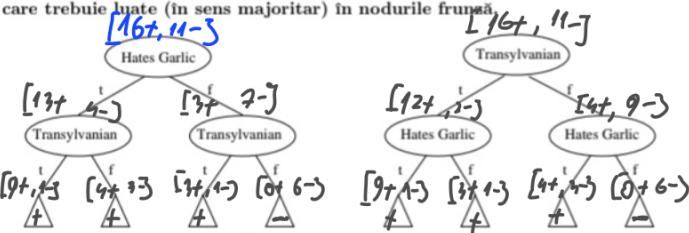
• CMU, 2014 spring, Seyoung Kim, HW2, pr. 1.4.1-5

Centrul [de Medicină] pentru Controlul și Prevenția Maladiilor a fost sesizat în legătură cu o creștere surprinzătoare a aparițiilor de vampiri. Acest centru a colectat date preliminare referitoare la anumite *caracteristici*, atât pentru vampiri [deja] cunoscuți cât și pentru non-vampiri, și acum ar dori să construiască un arbore de decizie ca să-i ajute pe cetăteni să identifice noi vampiri. Datele culese sunt prezentate în tabelul următor:

$V$ (Vampire)	$G$ (Hates Garlic)	$T$ (Transylvanian)	Nr. apariții
+	t	t	9
+	t	f	4
+	f	t	3
+	f	f	0
-	t	t	1
-	t	f	3
-	f	t	1
-	f	f	6

Fiecare linie indică ce caracteristici au fost „observate”, și de câte ori a fost observată fiecare combinație de caracteristici. De exemplu, combinația (+, t, t) a fost observată de 9 ori, pe când combinația (+, f, f) nu-a fost observată niciodată. (Simbolii 't' și 'f' au fost folosiți în locul lui True și False, pentru a evita confuzia cu atributul  $T_e$ .)

a. Completăți arborii de mai jos cu informații (count-uri) referitoare la partionarea datelor (sub formă  $[+n, -m]$ ) în fiecare nod, precum și cu deciziile care trebuie luate (în sens majoritar) în nodurile frunză.



<sup>580</sup> Atenție! Puteți elabora raționamente perfecte fără să faceți calcule laborioase!

Explorând cu atenție datele de antrenament, veți vedea că puteți să identificați atrbute cu putere de predicție / „discriminare” maximă [a valorilor variabilei de ieșire]. Cât este entropia condițională medie a acestor atrbute [în raport cu variabila de ieșire]?

5  
27, note erorii.

b. Calculați entropia condițională medie  $H(V|G)$ . (Toate calculele intermedii trebuie făcute cu o precizie de cel puțin 4 zecimale, pentru a ne asigura că răspunsul final are primele 3 zecimale corecte.)

c. Calculați entropia condițională medie  $H(V|T)$ . (Din nou, toate calculele intermedii trebuie făcute cu o precizie de cel puțin 4 zecimale.)

d. Care dintre cei doi arbori de decizie de mai sus reprezintă rezultatul învățării realizate de algoritm ID3 pe aceste date?

e. Adevărat sau Fals: Arborele produs de către ID3 va clasifica o persoană căreia-i displace usturoiul (engl., hates garlic) dar nu este transilvănean ca fiind vampir.

*Indicație:* Este posibil să aveți nevoie de următoarele valori pentru entropia ( $H(p)$ ) unei variabile aleatoare Bernoulli de parametru  $p$ :  $H(1/7) = 0.5916$ ,  $H(4/17) = 0.7871$ ,  $H(3/10) = 0.8812$ ,  $H(4/13) = 0.8904$ ,  $H(11/27) = 0.9751$ .

b)

$$\text{Calc } H(V|G) = \frac{12+3}{16+11} H(V|G=t) + \frac{3+7}{16+11} H(V|G=f) =$$

$$= \frac{12+3}{16+11} H\left[\frac{12+4-3}{16+11} H\left[\frac{3+7-5}{11}\right] = \dots\right] + \dots H\left[\frac{3}{10}\right] =$$

$$= \frac{17}{27} \cdot 0.7871 + \frac{10}{27} \cdot 0.8812 = 0.5955 + 0.3263 = \\ = 0.9218$$

$$\frac{\log_a x}{\log_b x} = \frac{\log_a}{\log_b}$$

c)  $H(V|T) = \frac{12+2}{27} H(V|T=t) + \frac{15+9}{27} H(V|T=f) =$

$$= \frac{12+2}{27} H\left[\frac{12+2-3}{15+9} + \frac{9+9}{27} H\left[\frac{9+7-5}{15}\right] = \frac{15}{27} H\left[\frac{15}{27}\right] + \right]$$

$$+ \frac{13}{27} H\left[\frac{4}{13}\right] = \frac{15}{27} \cdot 0.5916 + \frac{13}{27} \cdot 0.8904 = .$$

d) ambele arbori produc același rezultat.  
ambele clasifică greșit pentru instanțele din tabel:

V	G	T	Nr.
-	t	f	3
-	f	T	1

}

2 erori.

e) Adevărat. Aceea și astfel de persoane își devință.

(“Decision stump” produs de ID3:  
raționament calitativ pe un exemplu simplu)

• CMU, 2010 fall, Ziv Bar-Joseph, midterm exam, pr. 5.a

Vrem să construim un arbore de decizie care să ne ajute să prezicem întârziările avioanelor. Timp de câteva luni am colectat informații, iar un rezumat al acestora este prezentat în tabelul următor:

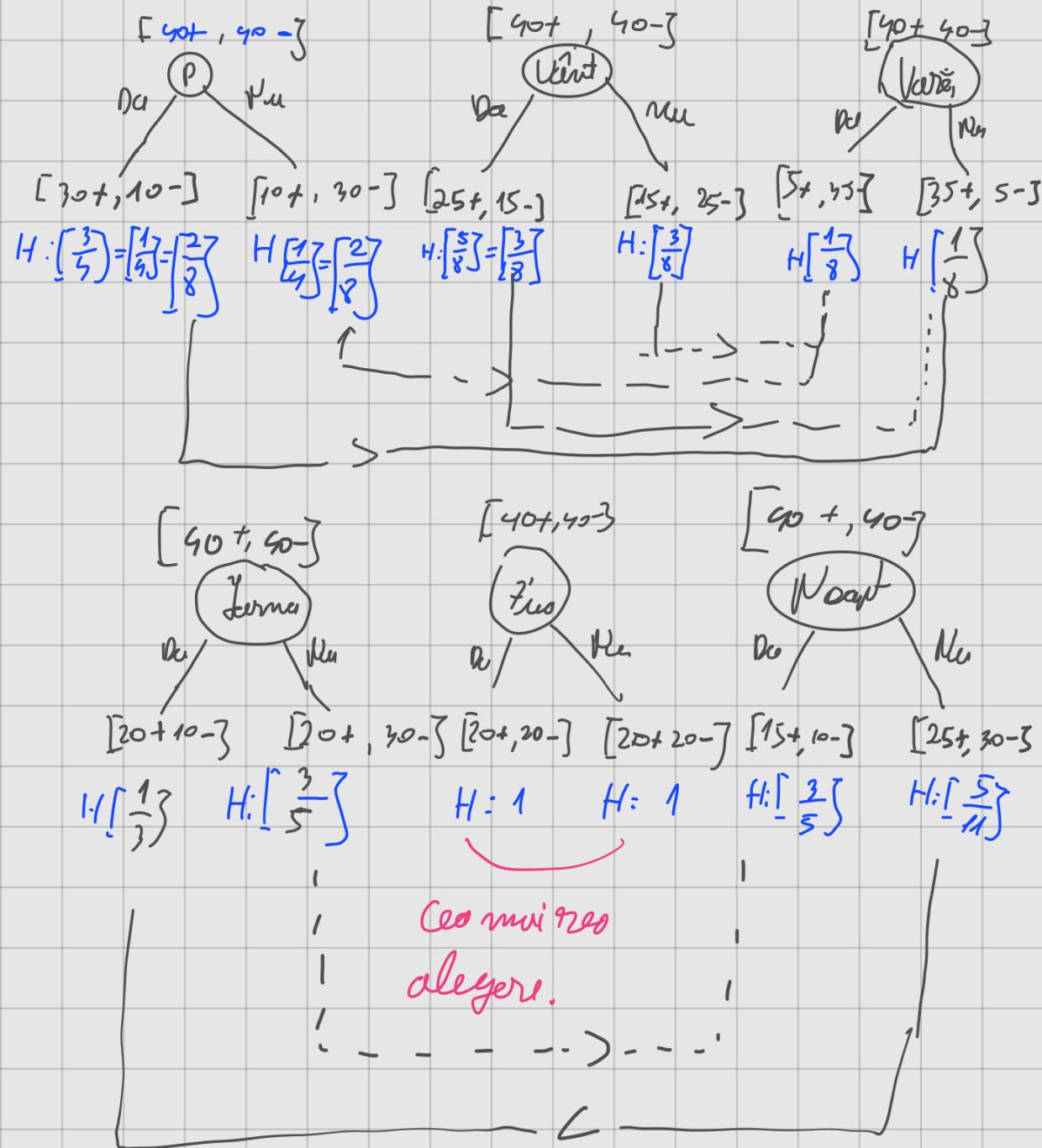
Atribut	Valoare = Da +		Valoare = Nu -	
	#Zboruri amânante	neamânante	#Zboruri amânante	neamânante
Ploaie	30	10	10	30
Vânt	25	15	15	25
Vara	5	35	35	5
Iarna	20	10	20	30
Ziua	20	20	20	20
Noaptea	15	10	25	30

a. Pe baza acestui tabel precizați ce atribut ar trebui să fie pus în rădăcina arborelui de decizie, folosind criteriul câștigului de informație. Justificați riguros; nu este însă necesar să elaborați în detaliu toate calculele.

b. Pe baza aceluiași tabel, precizați care dintre atrbute ar trebui să apară pe al doilea nivel (nivelul de sub rădăcină) al arborelui de decizie.

Cereafest

d)



$$H_{\text{O/Noante}} = \frac{1}{80} \log_2 \frac{25^{25}}{15^{15} \cdot 10^{10}} \cdot \frac{55^{55}}{25^{25} \cdot 30^{30}} = 0,9867$$

$$H_{\text{O/Iarmi}} = \frac{1}{80} \log_2 \left( \frac{30}{10^{10} \cdot 20^{20}} \cdot \frac{50}{20^{20} \cdot 30^{30}} \right) = 0,055$$

$$H_{\text{O/Kere}} = \frac{1}{80} \log_2 \left( \frac{40}{5^{5} \cdot 35^{35}} \cdot \frac{40}{5^{5} \cdot 35^{35}} \right) = 0,5132$$

$$H_{\text{O/Kere}} < H_{\text{O/Iarmi}} < H_{\text{O/Noante}} \Rightarrow$$

$\Rightarrow$  Aleger Kere ca rezolvare

Dacă prim răspuns este certificat  $\Rightarrow$  rezolvare cu așa.

„Kere” are cea mai mică entropie  $\equiv$  cel mai mare conținut de informație.

b) Jabelul nu oferă destulă informație, ceea ce înseamnă că nu putem calcula cu certitudine maximul de inf. și determina contribuția de pe mirelul 2.

40.

(Algoritmul ID3: aplicare; analiza „optimalității“ arborelui ID3)

• \* CMU, 2005 spring, C. Guestrin, T. Mitchell, midterm, pr. 4

Agenția spațială NASA dorește să distingă între martieni (M) și pământeni (H) folosind următoarele caracteristici:  $\text{Green} \in \{N, Y\}$ ,  $\text{Legs} \in \{2, 3\}$ ,  $\text{Height} \in \{S, T\}$ ,  $\text{Smelly} \in \{N, Y\}$ .

581

## ARBORI de DECIZIE

## Probleme propuse

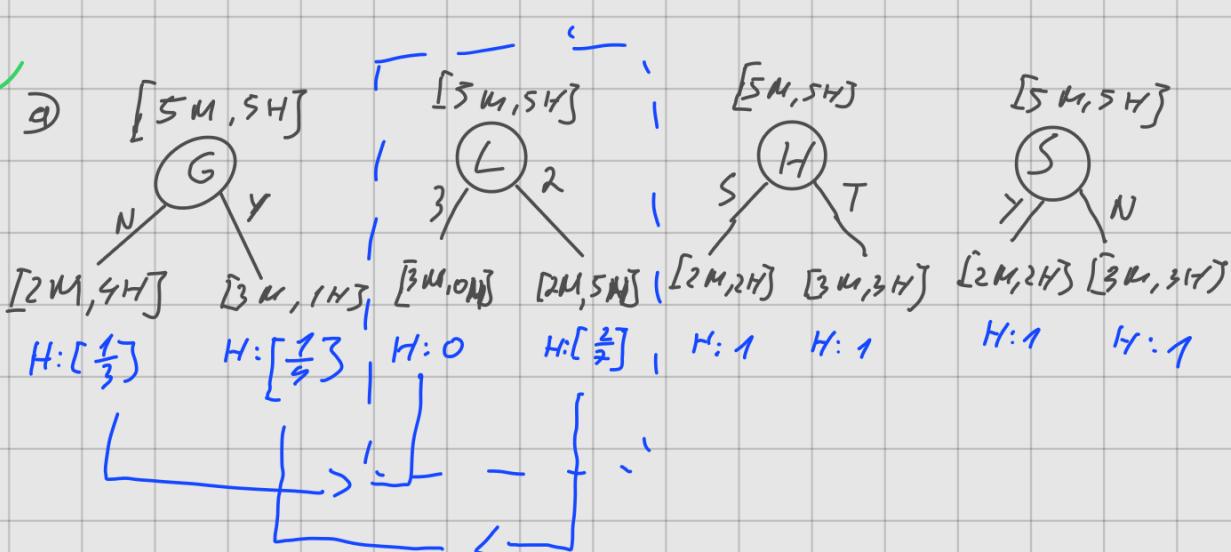
Datele de antrenament de care dispunem sunt prezentate în tabelul alăturat.

a. Învătați un arbore de decizie folosind algoritmul ID3 și trasați arborele respectiv.

	Species	Green	Legs	Height	Smelly
1	M	N	3	S	Y
2	M	Y	2	T	N
3	M	Y	3	T	N
4	M	N	2	S	Y
5	M	Y	3	T	N
6	H	N	2	T	Y
7	H	N	2	S	N
8	H	N	2	T	N
9	H	Y	2	S	N
10	H	N	2	T	Y

b. Descrieți conceptul M (marțian) ca un set de reguli conjunctive din logica propozițiilor. Spre exemplu:

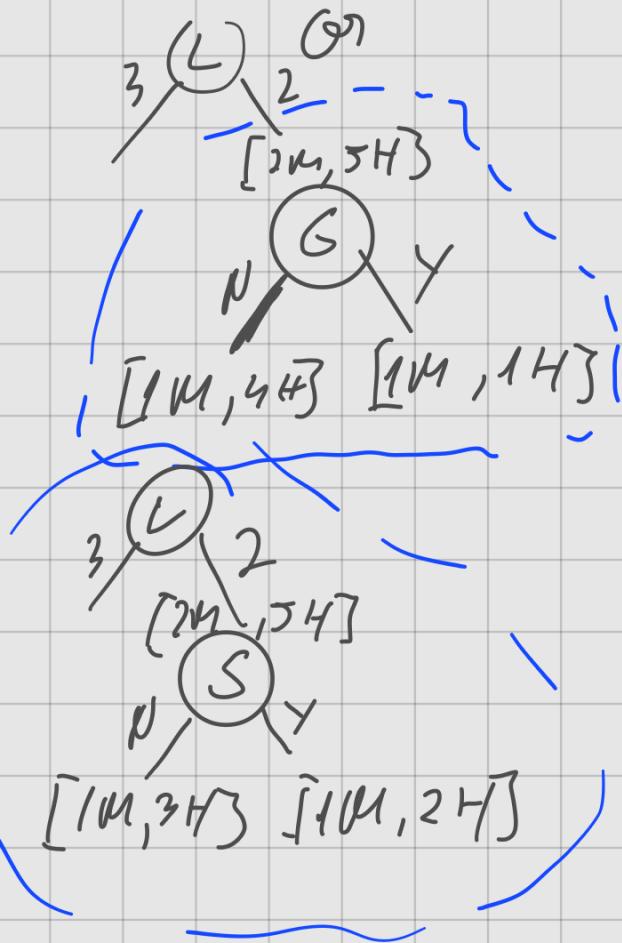
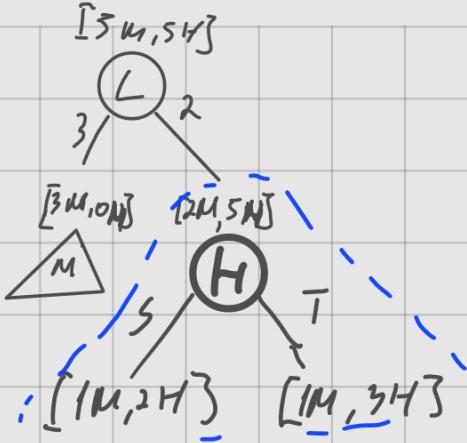
if  $\text{Green} = Y$  and  $\text{Legs} = 2$  and  $\text{Height} = T$  and  $\text{Smelly} = N$  then M;  
else  
if ... then M; else H.



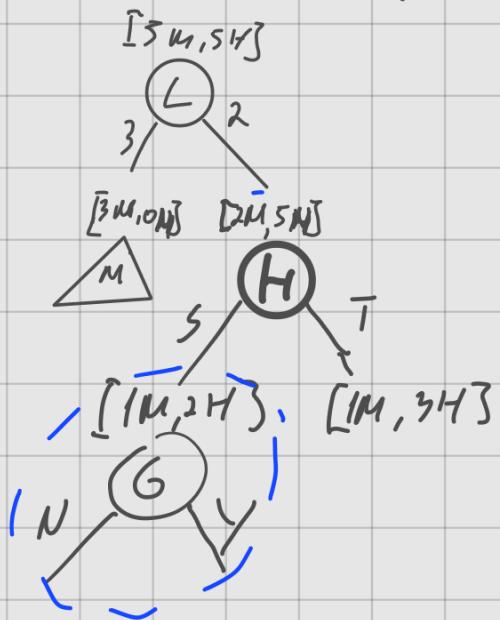
meritoriu.

$$H_{0/G} = -\left(\frac{2+4}{10} \cdot \log_2 \frac{2+4}{10} + \frac{3+1}{10} \cdot \log_2 \frac{3+1}{10}\right) = 0.9709 \quad \left\{ \begin{array}{l} \Rightarrow H_{0/L} < H_{0/G} \\ \Rightarrow H_{0/L} < H_{0/G} \end{array} \right.$$

$$H_{0/L} = -\left(\frac{3}{10} \cdot \log_2 \frac{3}{10} + \frac{7}{10} \cdot \log_2 \frac{7}{10}\right) = 0.8713 \quad \left\{ \begin{array}{l} \Rightarrow H_{0/L} - mersuabil \end{array} \right.$$



(3) aber  $H$ .



$$H_{0|H} = \frac{4+2}{2+5} \cdot \log \frac{3}{7} + \frac{4+3}{2+5} \cdot \log \frac{3}{5} = 0.2602$$

$$H_{mid|G} = \frac{2}{7} \cdot H[1M, 4H] + \frac{5}{7} \cdot H[1, 1] =$$

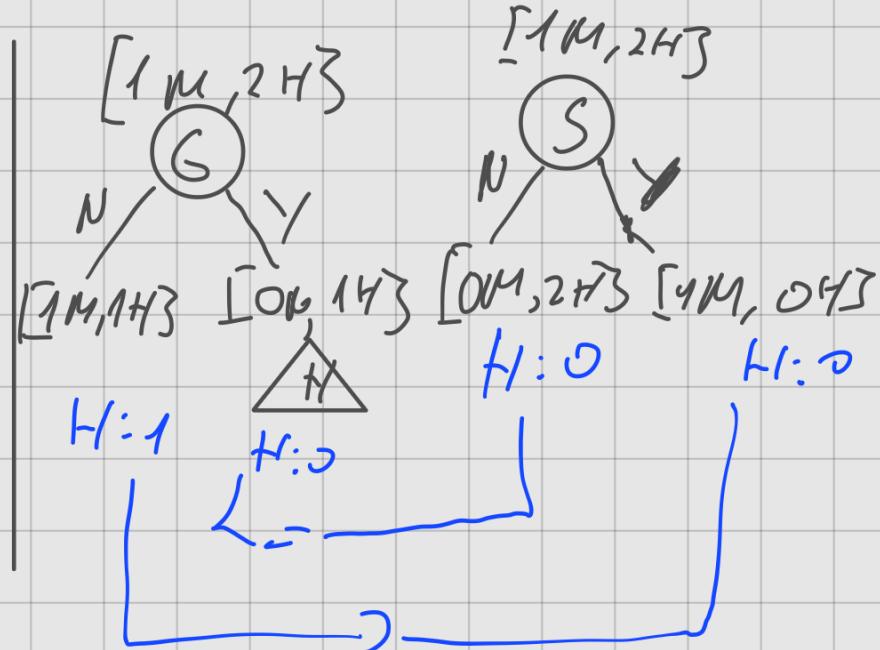
$$= \frac{2}{7} \cdot \left( \frac{1}{5} \log 5 + \frac{4}{5} \log \frac{5}{4} \right) + \frac{5}{7} = 0.9205$$

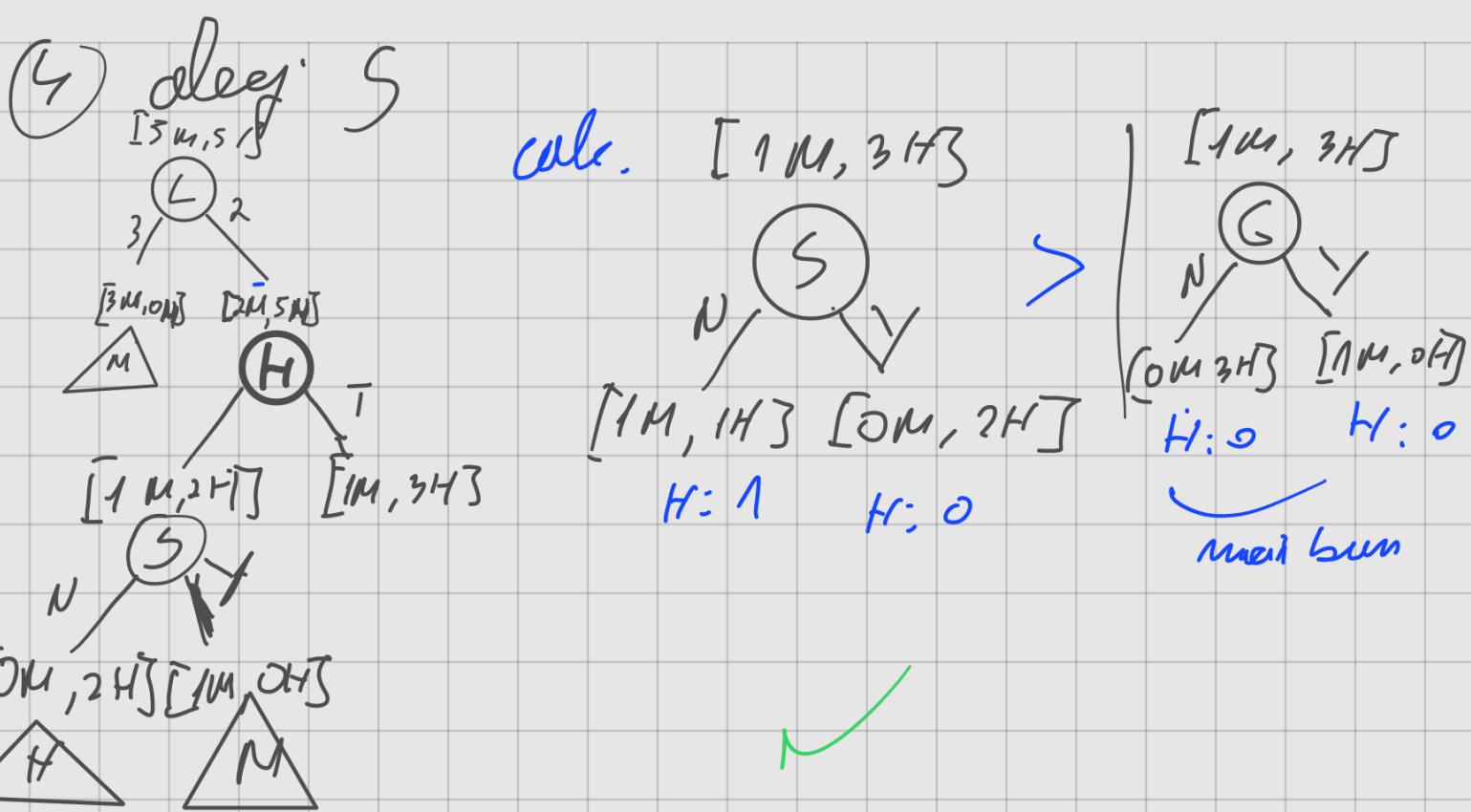
$$H_{mid|S} = \frac{2}{7} \cdot H[1, 3] + \frac{5}{7} \cdot H[1, 2] =$$

$$= \frac{2}{7} \left( \frac{1}{3} \log 5 + \frac{3}{5} \log \frac{5}{3} \right) + \frac{5}{7} \cdot \left( \frac{1}{3} \log 3 + \frac{2}{3} \log \frac{2}{1} \right) = 0.5636 S$$

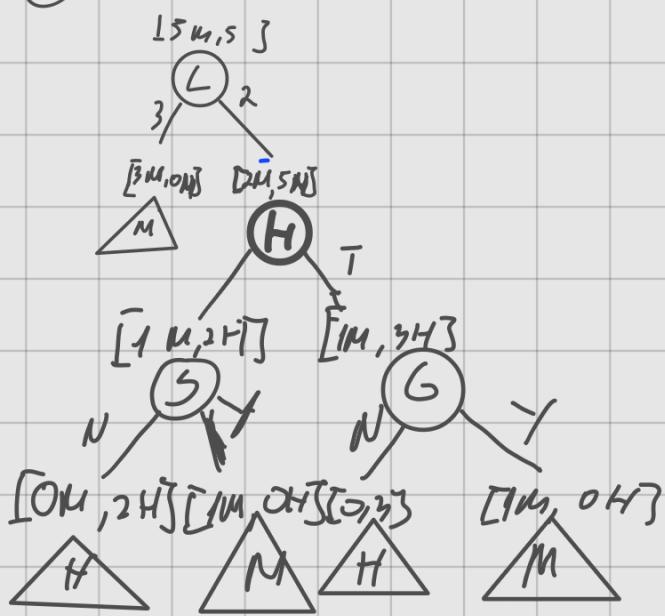
$$= 0.2312 + 0.5026 = 0.6338$$

Gilt. da  $H_{mid|S}$  nur  $H_{mid|H}$  &  $H_{mid|G}$ .





(5) *oleg G:*



if  $L = 3$  then  $M;$

else if  $L = 2$  and  $Green = Y$  and  $Height = S$  then  $M;$

else if  $L = 2$  and  $Height = T$  and  $Smelly = Y$  then  $M;$

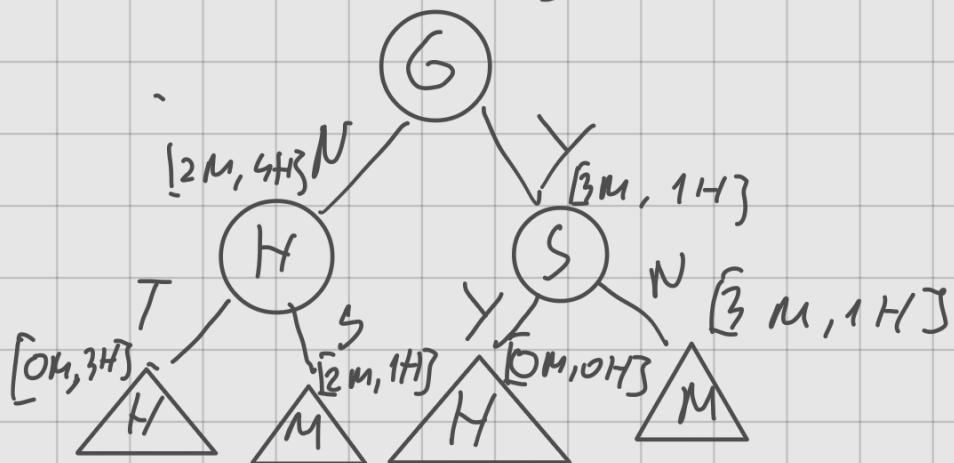
else if  $L = 2$  and  $Height = T$  and  $Smelly = N$  and  $Green = Y$

then  $M$ ; else  $H;$

c. Soluția de la punctul b de mai sus folosește cel mult 4 atribute în fiecare conjuncție. Găsiți un set de reguli conjunctive care folosesc doar 2 atribute pentru fiecare conjuncție, păstrând însă eroarea la antrenare zero. Această ipoteză mai simplă poate fi reprezentată ca un arbore de decizie de adâncime 2? Justificați răspunsul.

If  $L = 3$  then  $M;$   
 else  
 if  $L = 2 \text{ and } Green = Y$  then  $M;$   
 else  
 if  $L = 2 \text{ and } Smelly = Y$  then  $M;$

If  $H = S \text{ and } G = N$  then  $M;$   
 else  
 if  $S = N \text{ and } G = Y$  then  $M;$   
 else  $H$  [5M, 5H]



Această eroare de antrenare este 0%.

39.

(Algoritmul ID3: aplicare pe expresii booleene;  
exploatarea simetriilor operațiilor  $\vee, \wedge$  în alegerea atributelor;  
analiza „optimalității“ arborelui ID3)

\* prelucrare de Liviu Ciortuz, după  
Tom Mitchell, "Machine Learning", 1997, ex. 3.1.d

Considerăm următoarea funcție booleană:  $(A \wedge B) \vee (C \wedge D)$ . Valorile pe care le ia această funcție, calculate conform diferitelor valorii de adevăr atribuite variabilelor / atributelor  $A, B, C$  și  $D$  sunt cele cunoscute din logica propozițiilor. Dorim însă să reprezentăm această funcție ca arbore de decizie.

a. Aplicați algoritmul ID3 acestei funcții.

*Observație:* Dacă exploatați simetriile, este nevoie doar de puține calcule, altfel vă complicați în mod inutil.

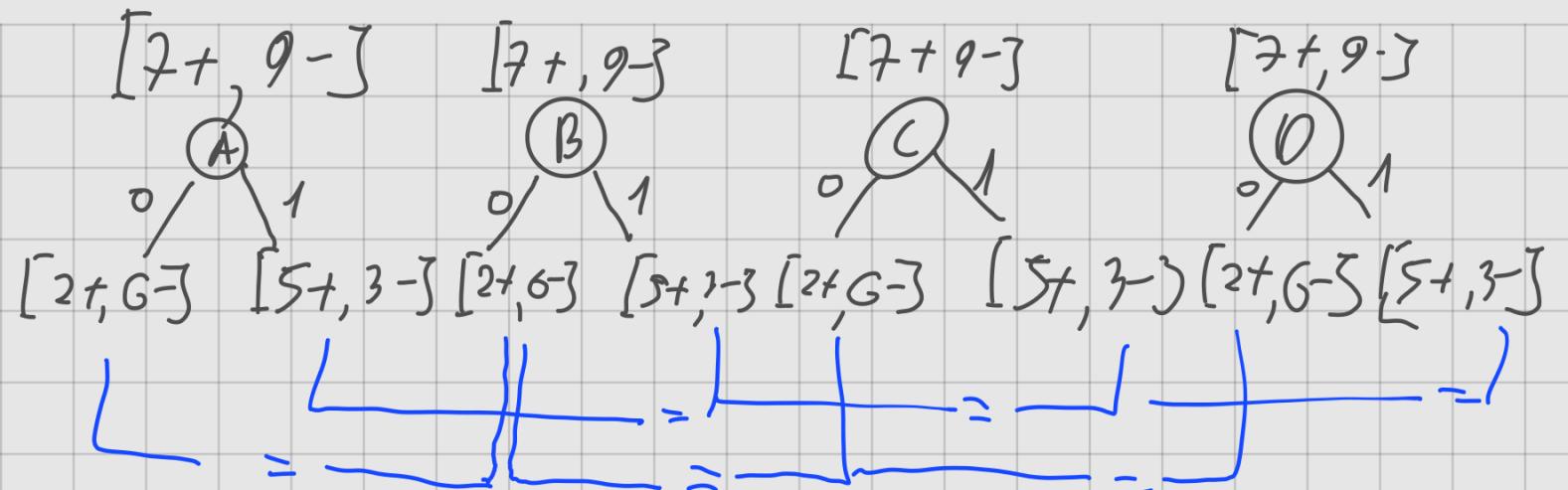
b. Arborele ID3 obținut la punctul precedent este optimal?

Altfel spus, puteți găsi alt arbore de decizie de adâncime mai mică sau cu număr mai mic de noduri (de test) pentru această funcție? (Tineți cont că în fiecare nod al unui arbore de decizie se poate testa un sigur atribut.)

a)

$$X: (A \wedge B) \vee (C \wedge D)$$

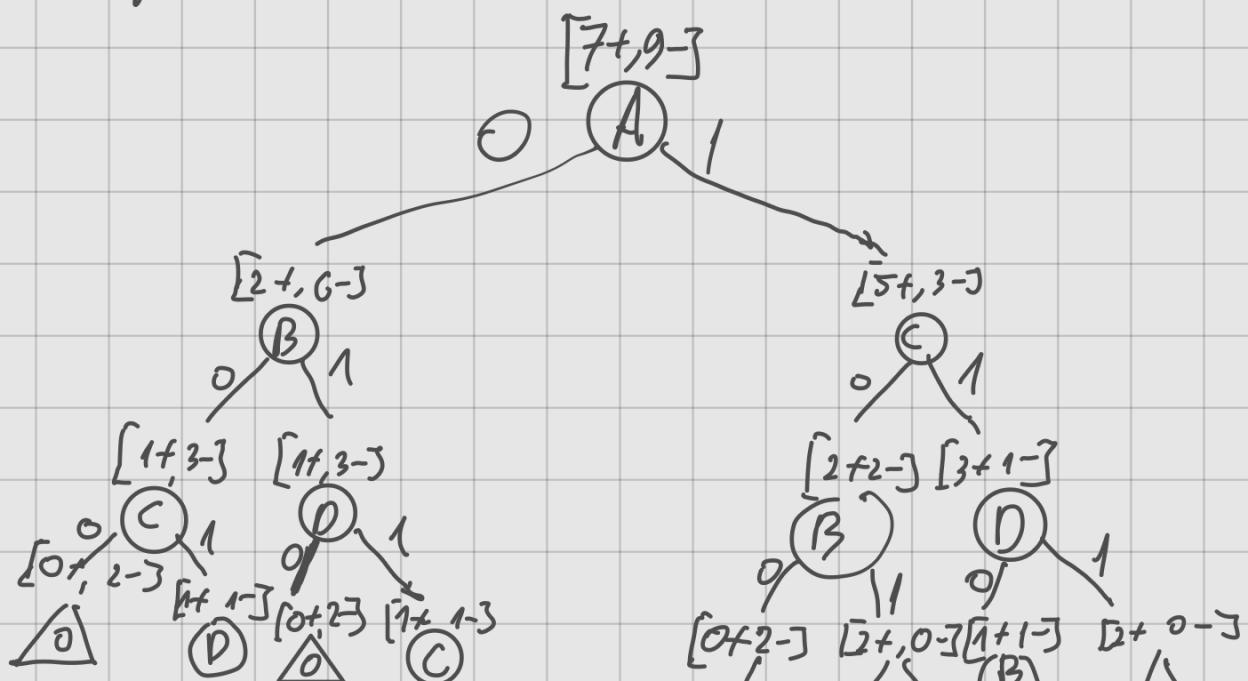
A	B	C	D	X
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	1
1	0	0	0	0
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1

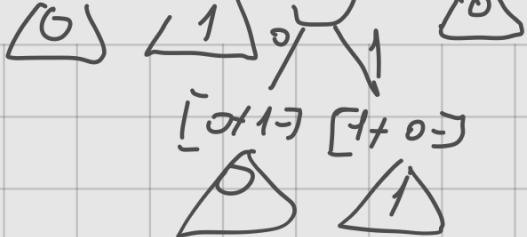
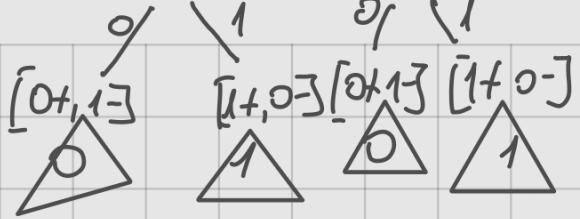


$A, B, C, D$ , sunt egli  $\Rightarrow$  au același număr de rădăcini, deci putem alege ca rădăcinile să din acestea.

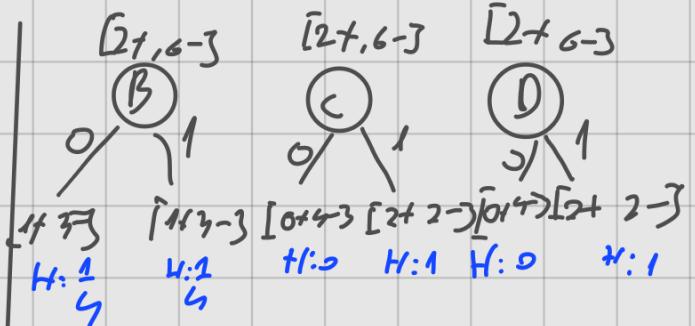
Obțin că, că  $A, B \wedge C, D$  sunt simetrice deoarece proprietatea de comutativitate a operației  $\wedge$ . Deci putem să alegem că valoarea pentru  $A, B \wedge C, D$  vor fi aceleși. Deci nu vom calcula și  $C, D$ .

① Alegem nodul  $A$  ca rădăcină.





A	B	C	D	X
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	1
1	0	0	0	0
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1



$$H_{\text{mod}}/B = \frac{1/3}{2+6} \cdot H[1+3-3] + \frac{1/3}{2+6} \cdot H[1+3-3] =$$

$$= \frac{1}{2} \cdot H[1+3-3] + \frac{1}{2} \cdot H[1+3-3]$$

$$H[1+3-3] = \frac{a}{a+b} \cdot \log \frac{a}{a+b} + \frac{b}{a+b} \cdot \log \frac{b}{a+b} =$$

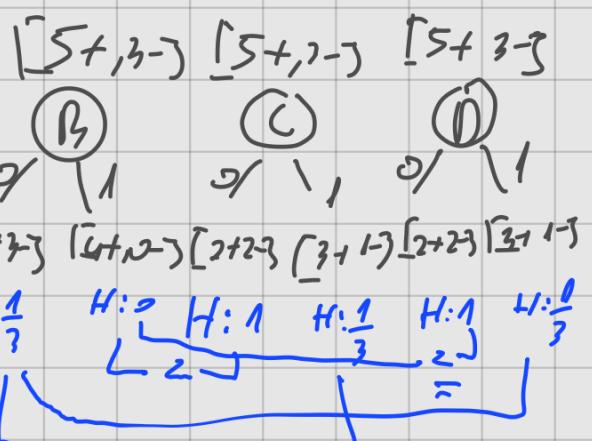
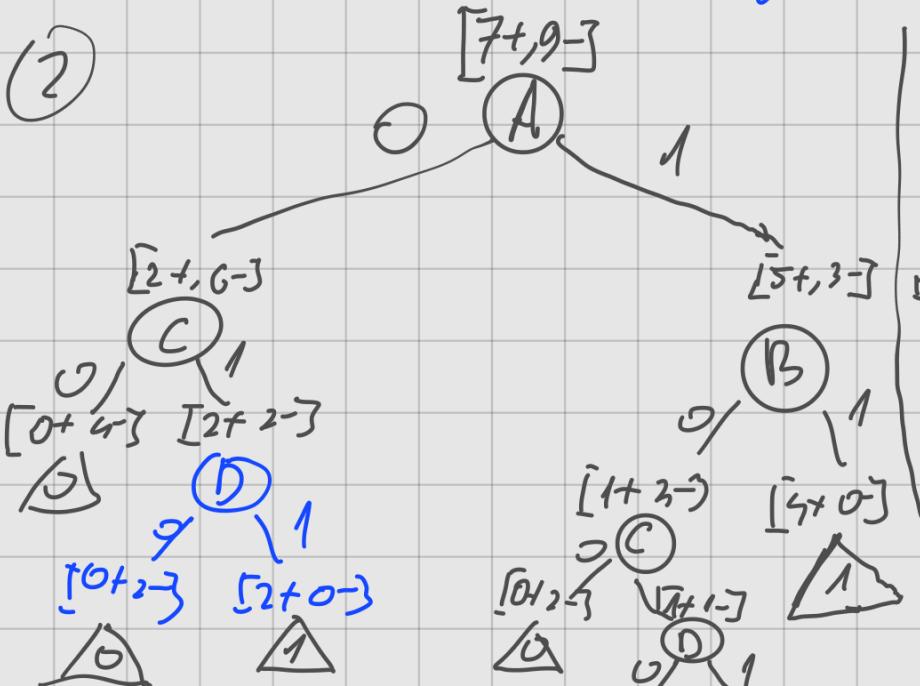
$$= \frac{1}{5} \log \frac{4}{5} + \frac{3}{5} \log \frac{3}{5} = \frac{1}{2} + \dots = 0.8112$$

$$H_{\text{mod}}/C = H_{\text{mod}}/D$$

$$H_{\text{mod}}/C = \frac{1}{8} \cdot 0 + \frac{1}{8} \cdot 1 = \frac{1}{2} = 0.5$$

alleg C zu D, nur cont. conf. reg. n'imm.

(2)

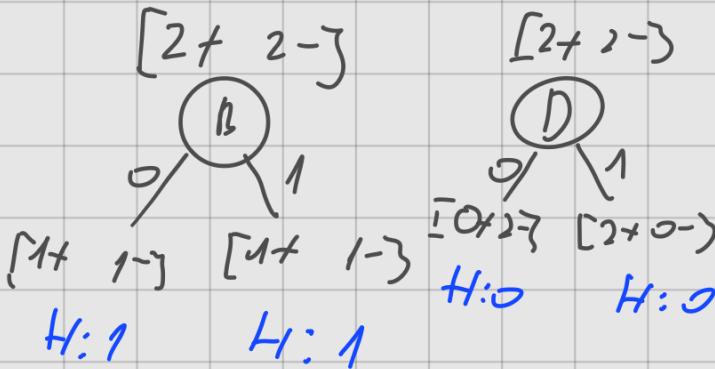


alleg B,

$[0+1-]$   $[1+0-]$

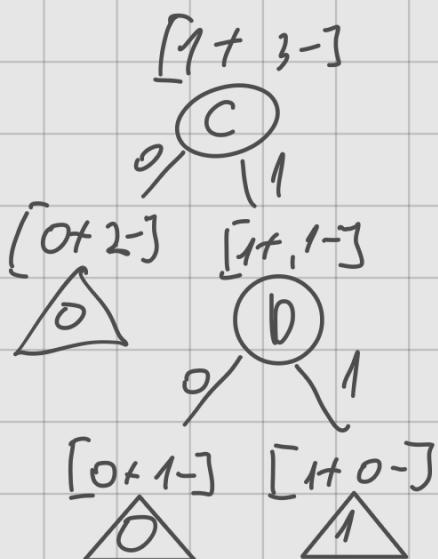


Facem nt. sub. arbrei dins al leii C.



aleg D

Facem nt. sub. arb. rămas al leii B



Conf. obv. de mai sus, alegem  
ori C ori D.

← aici alegem ultimul nod din  
adâncoră D.

5

Arborele este optimul, nu putem crea

unul cu adâncime mai mică din cauza proprietăților

de simetrie al operatorului A. Oricum am rezolvat  
nodurile, vom obține mereu arbori numai  
de nivel.

Studentul Timmy dorește să știe cum ar trebui să proceze cel mai bine ca să promoveze examenul de învățare automată. Pentru aceasta, a cules informații de la studenții care au urmat acest curs în anii precedenți și apoi a decis să-și construiască un *model* bazat pe arbori de decizie. A colectat în total nouă *instance* / exemple, descrise cu ajutorul a două trăsături (văzute în cele ce urmează ca două variabile aleatoare,  $S$  și  $A$ ): „este bine să stai și să înveți până noaptea târziu înainte de examen” ( $S$ ) și „este bine să mergi la toate cursurile și seminariile” ( $A$ ). Timmy dispune acum de următoarele „statistici” (care sunt de fapt *partitionări* ale datelor sale):

$$Set(\text{all}) = [5+, 4-]$$

$$Set(S+) = [3+, 2-], Set(S-) = [2+, 2-]$$

$$Set(A+) = [5+, 1-], Set(A-) = [0+, 3-]$$

Presupunând că se folosește drept criteriu de selecție a celei mai bune trăsături căștigul maxim de informație, ce trăsătură va alege Timmy? Care este valoarea căștigului de informație?

Puteți folosi la calcule următoarele aproximări:

$N$	3	5	7
$\log_2 N$	1.5850	2.3219	2.8073

$S+$  : „Studenți care au învățat până noaptea târziu”

$S-$  : „Studenți care nu au învățat până noaptea târziu”

$A+$  : „Studenți care au mers la facultate”

$A-$  : „Studenți care nu -”

1)  $H(\text{all}) = ?$

$$H(\text{all}) = \sum_{X \in \text{Val}(x)} p(x=x) \cdot \log_2 \frac{1}{p(x=x)} =$$

$$= p(\text{all} = S+) \cdot \log_2 \frac{1}{p(\text{all} = S+)} + p(\text{all} = S-) \cdot \log_2 \frac{1}{p(\text{all} = S-)} =$$

$$= \frac{5}{9} \cdot \log_2 \frac{9}{5} + \frac{4}{9} \cdot \log_2 \frac{9}{4} =$$

$$= - \left( \frac{5}{9} \cdot \log_2 \frac{9}{5} + \frac{4}{9} \cdot \log_2 \frac{9}{4} \right) = -0.091$$

[2+1, 2-3]

$$H(S+) = - \left( \frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \cdot \log_2 \frac{1}{2} \right) =$$

$$= -\frac{1}{2} \left( \log_2 \frac{1}{2} \cdot \frac{1}{2} \right) = -\frac{1}{2} \log_2 \frac{1}{4} =$$

$$= \frac{1}{2} \cdot \log_2^2 \frac{1}{4} = 1$$

$H(A+) = 0$

2) Calc. căștig. nt. fiecare subset.

$$H(S+) = - \left( \frac{3}{5} \log_2 \frac{3}{5} + \frac{2}{5} \log_2 \frac{2}{2} \right)$$

$$H(A+) = - \sum_{i=1}^6 p_{A+} \log_2 \frac{5}{6} - \frac{1}{6} \cdot \log_2 \frac{1}{6} =$$

$$= -0.263 \quad 2.587$$

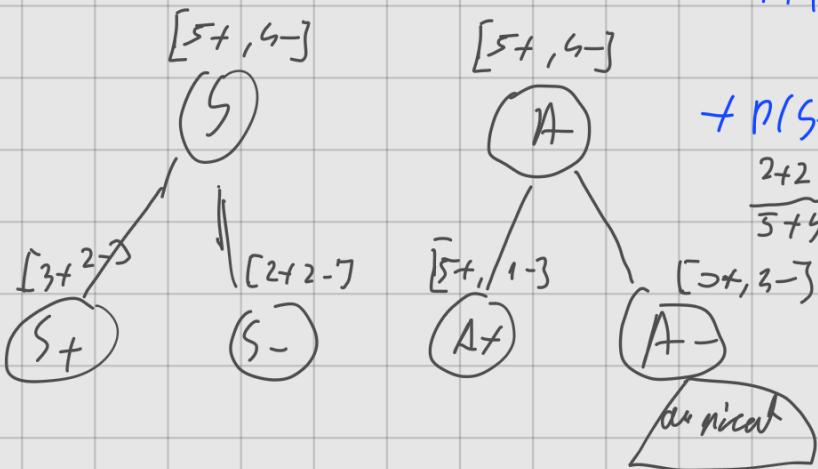
$$= -\frac{5}{6} \log_2 \frac{5}{6} + \frac{1}{6} \log_2 \frac{6}{6} = 0.670$$

Mit interessant

$$\text{IG}_{\text{all}/S} = H(\text{all}) - H(\overset{S+}{\text{all}})$$

$$= H(\text{all}) - \left( P(S+) \cdot H \left[ \begin{matrix} S+ \\ S- \end{matrix} \right] + \right.$$

$$+ P(S-) \cdot H \left[ \begin{matrix} S- \\ S+ \end{matrix} \right] \left. \right) = (\cancel{x})$$



$\text{IG}(A) = H[\text{all}] - H[\text{all}/A]$

$$H_{\text{mod}} = H[a+, b-] = \frac{a}{a+b} \log_2 \frac{a+b}{a} + \frac{b}{a+b} \log_2 \frac{a+b}{b}$$

$$H(\text{mod}/A) = \frac{c+d}{a+b} \cdot H[c+, d-] + \frac{e+f}{a+b} \cdot H[e+, f-]$$

$$\text{IG}(A) = H(\text{all}) - \left( \frac{1}{9} \cdot \frac{3}{5} \cdot \log_2 \frac{3}{5} + \frac{4}{9} \cdot \frac{2}{5} \cdot \log_2 \frac{2}{5} \right) =$$

$$= - \left( \frac{5}{9} \cdot \log_2 \frac{5}{9} + \frac{4}{9} \cdot \log_2 \frac{4}{9} \right) = 0.985 =$$

$$= \frac{5}{9} \cdot 0.736 + \frac{4}{9} \cdot 1.160 = 0.985 =$$

$$= \underbrace{0.408 + 0.519}_{0.001} - 0.985 = 0.006$$

Calc.  $\text{IG}(\text{all}/A) = H(\text{all}) - H(\text{all}/A)$

$$\begin{aligned}
 &= H(\text{all}) - \left( \frac{5+1}{5+5} \cdot H[5, 1-3] + \frac{0+3}{5+5} \cdot H[0+3, 3-3] \right) = \\
 &= 1.027 - \left( \frac{2}{3} \cdot H\left(\frac{5}{6}\right) + \frac{1}{3} \cdot 0 \right) = \\
 &= 1.027 - \frac{2}{3} \cdot 0.650 = \\
 &= 1.027 - 0.433 = 0.558
 \end{aligned}$$

Deci  $H_6(\text{all}/A) > H_6(\text{all}/S)$

$$0.558 > 0.006$$

Deci Timofeo arătura A, deoarece are un cștiig de inf. mai mare.

33.

(Expresivitatea arborilor de decizie:  
un rezultat privind funcțiile booleene)

Orice funcție booleană (care primește  $n$  argumente din mulțimea  $\{0, 1\}$  și întoarce un element din mulțimea  $\{0, 1\}$ ) poate fi reprezentată cu ajutorul unui arbore de decizie. Adevărat sau fals?

În cazul afirmativ, explicați succint cum anume poate fi construit arboarele de decizie respectiv.

În cazul negativ, dați un exemplu de funcție booleană pentru care nu se poate construi un arbore de decizie consistent cu funcția respectivă.

Aderat.

- Un oricare de decizie este construit

din : - moduri interne (Piesă mod internă reprezentă o modalitate lăcoșitoare care include  
trei valori cuprinse între 50, 1)

- trămare : (Piesă mod trămăză reprezintă un grup de valori pe care le poți exprima în mod divers)
- moduri frumoase . (Piesă mod frumoasă reprezintă ceea ce te pot aduce bucurie)

