

85.

(Algoritmul ID3: aplicare;
calculul erorii la antrenare, respectiv la validare)

* CMU, 2003 fall, T. Mitchell, A. Moore, midterm exam, pr. 1

Folosiți următorul set de date pentru a învăța cu ajutorul unui arbore de decizie dacă o ciupercă este sau nu comestibilă, utilizând atributele discrete *Formă*, *Culoare* și *Miros*.

<i>Formă</i>	<i>Culoare</i>	<i>Miros</i>	<i>Comestibilă</i>
C	B	1	1
D	B	1	1
D	W	1	1
D	W	2	1
C	B	2	1
D	B	2	0
D	G	2	0
C	U	2	0
C	B	3	0
C	W	3	0
D	W	3	0

a. Calculați entropia condițională specifică $H(\text{Comestibilă} \mid \text{Miros} = 1 \text{ sau } \text{Miros} = 3)$.

Sugestie: Aplicați (deci, scrieți mai întâi) definiția noțiunii de entropie condițională specifică.

b. Ce atribut va alege algoritmul ID3 ca rădăcină a arborelui de decizie?

c. Elaborați întregul arbore de decizie care va fi învățat din datele de mai sus (fără pruning).

d. Exprimați cu ajutorul unui set de reguli din calculul propozițional clasificarea produsă de arborele de decizie obținut. (IF ... THEN *Comestibilă*; IF ... THEN \neg *Comestibilă*.)

e. Să presupunem că avem un set de date de validare:

<i>Formă</i>	<i>Culoare</i>	<i>Miros</i>	<i>Comestibilă</i>
C	B	2	0
D	B	2	0
C	W	2	1

Care va fi eroarea produsă de arborele de decizie pe mulțimea de date de antrenare respectiv pe datele de validare? (Exprimați răspunsul ca număr de exemple clasificate greșit.)