

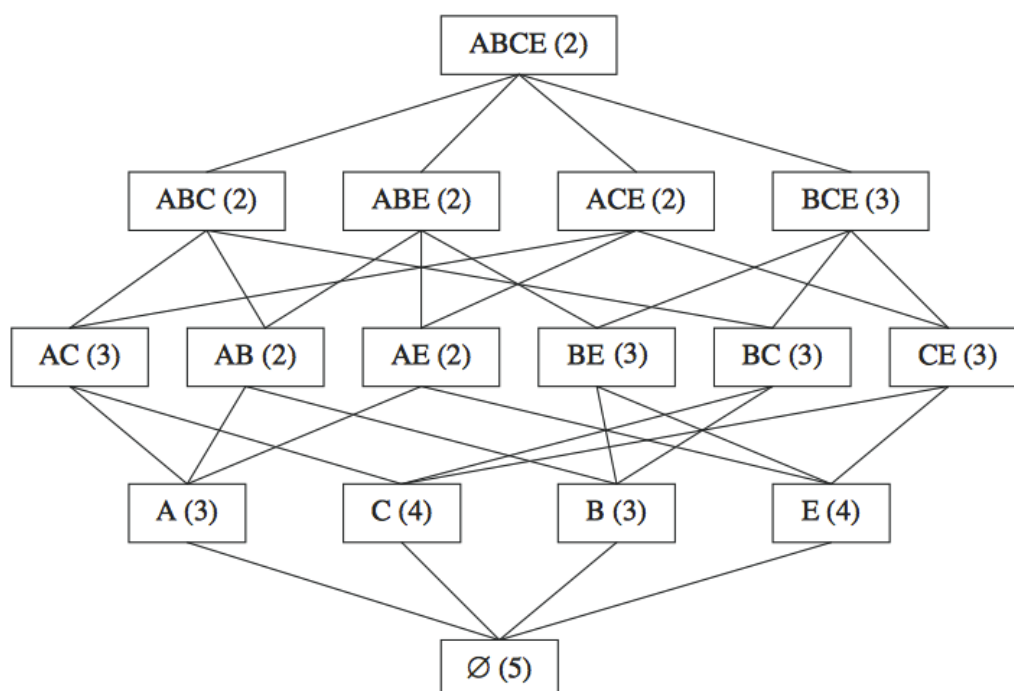
## Examen de rattrapage

### Questions de cours (4 pts : 2 + 1 + 1)

1. Deux classes d'algorithmes sont utilisées dans le clustering hiérarchique : les algorithmes divisibles et les algorithmes agglomératifs. Quelle différence existe entre les deux ?
2. Dire comment peut-on obtenir les règles de décisions à partir d'un arbre de décision.
3. Dire comment peut-on obtenir les règles de décision à partir des règles solides.

### Exercice 1 Motifs fréquents (8 pts : 2 + 2 + 2 + 2)

Soit le treillis suivant représentant les motifs obtenus à partir des items A,B,C et E avec leurs fréquences :



En supposant que la fréquence minimale est de 3 et que la confiance minimale est de 0.8, donner :

1. L'ensemble des motifs fréquents.
2. Les motifs fréquents fermés et les motifs fréquents maximaux.
3. Les motifs rares.
4. Les règles solides obtenus à partir du motif fréquent maximal le plus long.

**Tournez la page ...**

## Exercice 2 Classification (8 pts : 2 + 4 + 2)

Nous considérerons l'ensemble suivant d'exemples caractérisé par les attributs : Forme, Taille et Couleur.

Forme	Taille	Couleur	Classe
rond	petit	bleu	oui
carré	grand	rouge	non
rond	?	blanc	oui
carré	petit	bleu	oui
rond	grand	bleu	oui
carré	grand	blanc	non
carré	?	blanc	oui
carré	grand	bleu	non
carré	petit	rouge	oui
rond	grand	blanc	oui

1. Remplacer les données manquantes par le mode de l'attribut correspondant.
2. Construire l'arbre de décision correspondant par l'algorithme ID3.
3. Donner le taux de reconnaissance de l'arbre construit sur la table suivante :

Forme	Taille	Couleur	Classe
carré	petit	blanc	non
carré	grand	bleu	oui
rond	petit	rouge	oui
carré	grand	rouge	non

★★★ Bonne chance ★★★

Dr A.Djeffal

## Corrigé type

### Questions de cours (4 pts : 2 + 1 + 1)

1. Les algorithmes agglomératifs considèrent chaque exemple dans un cluster propre puis les regroupent d'étape en étape jusqu'à arriver à un seul cluster.  
Les algorithmes divisibles font l'inverse : ils considèrent tous les exemples dans un seul cluster puis les subdivisent d'étape en étape jusqu'à arriver à un cluster par exemple.
2. En prenant chaque chemin de la racine à une feuille en considérant les arcs comme des tests et la feuille comme décision.
3. En prenant les règles solides dont la partie droite égale à la classe.

### h Motifs fréquents (8 pts : 2 + 2 + 2 + 2)

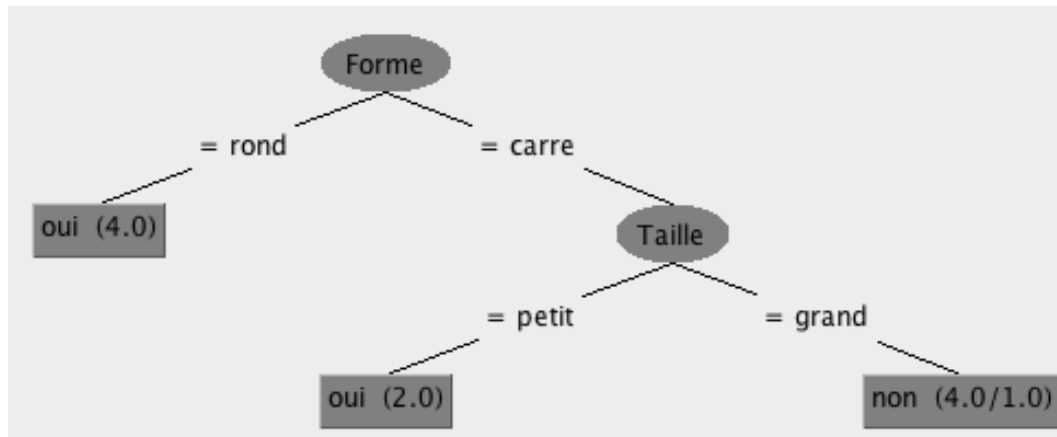
1. Motifs fréquents =  $\{A, B, C, E, AC, BE, BC, CE, BCE\}$
2. Motifs fréquents fermés =  $\{C, E, AC, BCE\}$
3. Motifs fréquents maximaux =  $\{AC, BCE\}$
4. Motifs rares =  $\{AB, AE, ABC, ABE, ACE, ABCE\}$
5. Les règles solides :
  - (a)  $BC \Rightarrow E$
  - (b)  $BE \Rightarrow C$
  - (c)  $CE \Rightarrow B$

### Classification (8 pts : 2 + 4 + 2)

1. Remplacement des données manquantes :

Forme	Taille	Coleur	Classe
rond	petit	bleu	oui
carré	grand	rouge	non
rond	<b>grand</b>	blanc	oui
carré	petit	bleu	oui
rond	grand	bleu	oui
carré	grand	blanc	non
carré	<b>grand</b>	blanc	oui
carré	grand	bleu	non
carré	petit	rouge	oui
rond	grand	blanc	oui

2. Arbre de décision



3. Précision =  $\frac{2}{4} = 50\%$