

# Chapitre 1

## Les arbres de décisions

### 1.1 Exercices

**EXERCICE 4.1** Donner les arbres de décisions qui expriment les fonctions booléennes suivantes :

1.  $A \wedge \neg B$
2.  $A \vee (B \wedge C)$
3.  $A \oplus B$
4.  $(A \wedge B) \vee (C \wedge D)$

#### Corrigé

1. – A==VRAI==>
  - B==VRAI ==> VRAI
  - B==FAUX ==> FAUX
- A==FAUX==> FAUX
2. – A==VRAI==> VRAI
- A==FAUX==>
  - B==VRAI==>
    - C==VRAI ==> VRAI
    - C==FAUX ==> FAUX
  - B==FAUX==> FAUX
3. – A==VRAI==>
  - B==VRAI ==> FAUX
  - B==FAUX ==> VRAI
- A==FAUX==>
  - B==VRAI ==> VRAI
  - B==FAUX ==> FAUX

**EXERCICE 4.2** Soient les exemples suivants :

1. Calculer l'entropie de l'ensemble d'exemples par rapport à la valeur de la classe.

Instance	Classe	a1	a2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

2. Quel le gain de l'attribut a2.

### Corrigé

1. En appliquant :

$$I(n, p) = \left(\frac{p}{p+n}\right) \log_2\left(\frac{p+n}{p}\right) + \left(\frac{n}{p+n}\right) \log_2\left(\frac{p+n}{n}\right) \quad (1.1.1)$$

nous avons 3 exemples positifs et trois négatifs donc  $I(n, p) = 1$

2. En appliquant :

$$E(T) = \sum \left(\frac{p_i + n_i}{p+n}\right) I(p_i, n_i) \quad (1.1.2)$$

où T est un test sur l'attribut a2 on aura :

$$E(T) = \frac{4}{6} * I(2, 2) + \frac{2}{6} * I(1, 1) = 1$$

Par conséquent, le gain est égal à 0.

**EXERCICE 4.3** Soient les exemples suivants :

NUM	CIEL	TEMP.	HUMI.	VENT	Concept
1	couvert	élevé	normale	non	1
2	pluvieux	moyenne	normale	non	1
3	ensoleillé	élevé	forte	non	0
4	ensoleillé	élevé	forte	oui	0

TABLE 1.1 – Description des conditions météorologiques et du concept JouerFoot

Comparer l'arbre de décision calculé par ID3 avec les hypothèses trouvées selon l'algorithme candidate-elimination. Commenter.

**Corrigé** L'arbre obtenu est :

- HUMI==normale==>1
- HUMI==forte==>0

Candidate-elimination nous donne l'espace de version suivant :

$$S = \{ \langle ?, ?, normale, non \rangle \}$$

$$G = \{ \langle ?, ?, normale, ? \rangle \}$$

La première branche de l'arbre de décision correspond à l'hypothèse la plus générale dans l'espace.

**EXERCICE 4.4** Installer le logiciel Weka à partir du site : <http://www.cs.waikato.ac.nz/ml/Weka/>.

1. Choisir l'application *explorer*.
2. Ouvrir le fichier "data/weather.nominal".
3. A l'aide du bouton *edit* vérifier le contenu du fichier (vous trouvez les 14 exemples traités en cours).
4. Choisir *classify* et ensuite choisir via le bouton *choose* la méthode ID3 (classés dans le répertoire *trees*).
5. Appuyer sur *start* : l'arbre de décision est construit.
6. Ouvrez avec un éditeur externe le fichier weather.nominal. Étudier le codage des données pour pouvoir faire l'exercice suivant.
7. Appliquer ID3 (via weka) sur l'ensemble d'exemples suivants et comparer avec le résultat de candidate-elimination :

Pays	Constructeur	Couleur	Année	Type	Exemple
Japon	Honda	Bleue	2000	Familiale	1
Japon	Toyota	Verte	1997	Sportive	0
Japon	Toyota	Bleue	1999	Familiale	1
Étas-Unis	Chrysler	Rouge	2000	Familiale	0
Japon	Honda	Blanche	2000	Familiale	1