

Centre Universitaire de Mila

Département: informatique

Apprentissage Automatique :Travaux dirigées

Année universitaire 2024-2025

# Exercice01-solution

Construire un **arbre de décision** qui classifie par des **conditions météorologiques** si le **jouer** joue au **tennis** ou non

$$H(\mathbf{x}) = - \sum_{\mathbf{x} = x_i} p(\mathbf{x} = x_i) \log_2(p(\mathbf{x} = x_i))$$

$$\text{Gain}(\mathcal{D}, a_j) = H(\mathcal{D}) - \sum_{v \in \text{valeurs}(a_j)} \frac{|\mathcal{D}_{a_j=v}|}{|\mathcal{D}|} H(\mathcal{D}_{a_j=v})$$

Jour	Ciel	Température	Humidité	Vent	Jouer au tennis ?
1	Ensoleillé	Chaude	Élevée	Faible	Non
2	Ensoleillé	Chaude	Élevée	Fort	Non
3	Couvert	Chaude	Élevée	Faible	Oui
4	Pluie	Tiède	Élevée	Faible	Oui
5	Pluie	Fraîche	Normale	Faible	Oui
6	Pluie	Fraîche	Normale	Fort	Non
7	Couvert	Fraîche	Normale	Fort	Oui
8	Ensoleillé	Tiède	Élevée	Faible	Non
9	Ensoleillé	Fraîche	Normale	Faible	Oui
10	Pluie	Tiède	Normale	Faible	Oui
11	Ensoleillé	Tiède	Normale	Fort	Oui
12	Couvert	Tiède	Élevée	Fort	Oui
13	Couvert	Chaud	Normale	Faible	Oui
14	Pluie	Tiède	Élevée	Fort	Non

## Exercice01-solution

$$\text{Gain}(\mathcal{D}, a) = H(D) - \frac{8}{14} H(\mathcal{D}_{a=\text{alpha}}) - \frac{6}{14} H(\mathcal{D}_{a=\text{beta}})$$

$$H(D) = -\frac{9}{14} \log_2 \left( \frac{9}{14} \right) - \frac{5}{14} \log_2 \left( \frac{5}{14} \right) = 0.940$$

$$H(\mathcal{D}_{a=\text{alpha}}) = -\frac{6}{8} \log_2 \left( \frac{6}{8} \right) - \frac{2}{8} \log_2 \left( \frac{2}{8} \right) = 0.811$$

$$H(\mathcal{D}_{a=\text{beta}}) = -\frac{3}{6} \log_2 \left( \frac{3}{6} \right) - \frac{3}{6} \log_2 \left( \frac{3}{6} \right) = 1$$

$$\text{Gain}(\mathcal{D}, a) = 0.940 - \frac{8}{14} 0.811 - \frac{6}{14} 1 = 0.940 - 0.8920 = 0.048$$

## Exercice01-solution

le gain d'information de l'attribut Ciel:

- $X_{\text{Ciel}}=\text{Ensoleillé}$ : 5 exemples.
- $X_{\text{Ciel}}=\text{Couvert}$ : 4 exemples.
- $X_{\text{Ciel}}=\text{Pluie}$ : 5 exemples.
- Alors :

$$\text{Ciel } [9(+), 5(-)] \Rightarrow \left\{ \begin{array}{l} 5 \text{ Ensoleille} : [2 (+) 3(-)]. \\ 4 \text{ Couvert} : [4 (+) 0(-)]. \\ 5 \text{ Pluie} : [3 (+) 2(-)]. \end{array} \right.$$

$$\text{Gain}(X, \text{Ciel}) = \mathbf{E(X)} - \frac{5}{14} \mathbf{E(X=Ensoleillé)} - \frac{4}{14} \mathbf{E(X=Couvert)} - \frac{5}{14} \mathbf{E(X=Pluie)}.$$

$$\Rightarrow \mathbf{E(X)} = - \frac{9}{14} \log_2 \left( \frac{9}{14} \right) - \frac{5}{14} \log_2 \left( \frac{5}{14} \right) = \mathbf{0.940}.$$

## Exercice01-solution

- $E(X=\text{Ensoleillé}) = -\frac{2}{5} \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \log_2 \left(\frac{3}{5}\right) = 0.971.$
- $E(X=\text{Couvert}) = 0.$
- $E(X=\text{Pluie}) = -\frac{3}{5} \log_2 \left(\frac{3}{5}\right) - \frac{2}{5} \log_2 \left(\frac{2}{5}\right) = 0.971.$
- Donc :  $\text{Gain}(X, \text{Ciel}) = 0.940 - \frac{5}{14} * (0.971) - 0 - \frac{5}{14} * (0.971) = 0.246.$
- **$\text{Gain}(X, \text{Ciel}) = 0.246.$**

## Exercice01-solution

Attribut	Gain
Ciel	0,246
Humidité	0,151
Vent	0,048
Température	0,029

# Exercice01-solution

- ✓ L'attribut Ciel peut prendre trois valeurs : Ensoleillé, Couvert et Pluie.
- On refait les calculs avec Ciel=Ensoleillé.

	Température	Humidité	Vent	Jouer
1	Chaude	Elevée	Faible	Non
2	Chaude	Elevée	Fort	Non
8	Tiède	Elevée	Faible	Non
9	Fraiche	Normale	Faible	Oui
11	Tiède	Normale	Fort	Oui

Humidité = 2 Normale : [2 + 0-].

3 Elevé : [0 + 3-].

Gain ( $X_{\text{Ensoleillé}}$ , Humidité) =  $E(X_{\text{Ensoleillé}})$  -  $2/5 E(X_{\text{Ensoleillé}}=\text{Normale})$  -  $3/5 E(X_{\text{Ensoleillé}}=\text{Élevé})$ .

$E_{\text{Ensoleillé}} = -2/5 \log_2(2/5) - 3/5 \log_2(3/5) = \mathbf{0.970}$ .

$E_{\text{Ensoleillé}}(X_{\text{Ensoleillé}}=\text{Élevée}) = -3/5 \log_2(3/5) = 0$ .

$E_{\text{Ensoleillé}}(X_{\text{Ensoleillé}}=\text{Normale}) = -2/5 \log_2(2/5) = 0$ .

Donc, Gain ( $X_{\text{Ensoleillé}}$ , Humidité) = **0.970**.

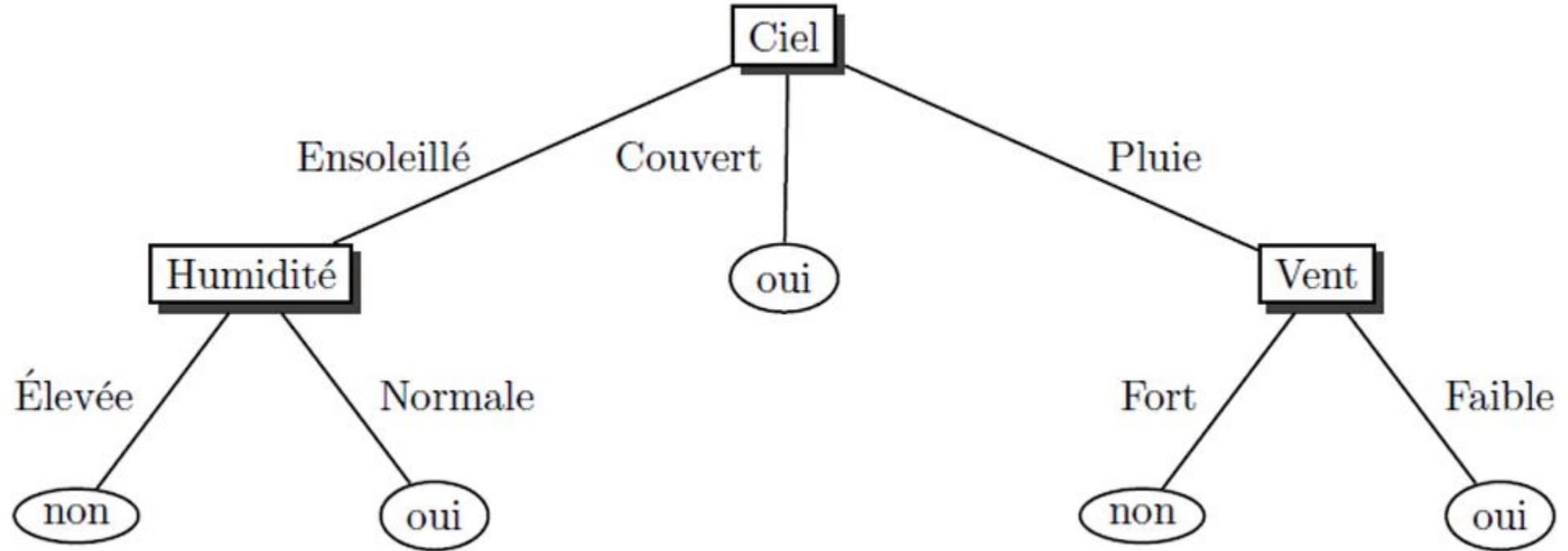
## Exercice01-solution

Attribut	Gain
Humidité	0,970
Vent	0,570
Température	0,019

# Exercice01-solution

- Pour la branche Couvert de la racine Pluie, on a 4 exemples (x3, x7, x12, x13) où tous les exemples sont positives ainsi nous effectuons la classe "Oui" à cette feuille.
- Pour la branche Pluie, on a 4 exemples (x4, x5, x6, x10, x14) donc on cherche le gain maximal entre les attributs.

# Exercice01-solution



## Exercice 2 :

Dans un hôpital, on souhaite construire un arbre de décision pour la prédiction du risque des patients d'avoir une certaine maladie en fonction de leur age et de deux symptômes booléens (vrai ou faux) appelés S1 et S2. Le risque est évalué selon trois valeurs F (faible), M (moyen) et E (élevé), l'age est discrétisé selon trois classes (jeune, adulte et senior). L'hôpital dispose de la table suivante :

N°	Age	S1	S2	Risque	N°	Age	S1	S2	Risque
1	Jeune	F	V	F	6	Jeune	F	F	F
2	Jeune	V	V	E	7	Adulte	V	F	M
3	Adulte	F	F	F	8	Adulte	V	V	M
4	Senior	V	F	E	9	Senior	F	F	F
5	Senior	F	V	M	10	Senior	V	V	E

Questions :

1. Construire l'arbre souhaité en utilisant l'algorithme ID3.
2. Donner le risque du patient ayant les attributs (Jeune, V, F) selon l'arbre construit.

# Exercice02-solution

