

2. Les Arbres de décision (Decision tree)

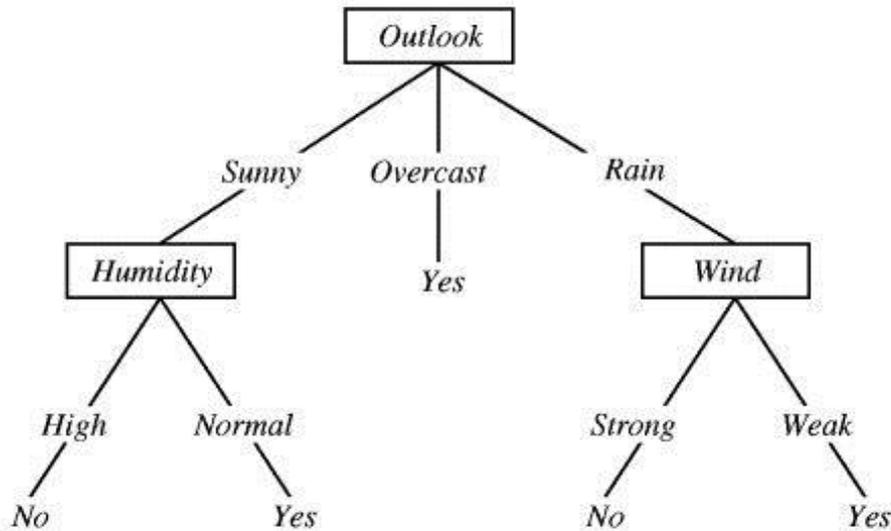
2.1. Exemple d'illustration

Un vice entraîneur a observé les décisions de l'entraîneur de jouer ou pas selon l'influence des prévisions météorologiques. Il a noté ces observations dans le tableau suivant (inspiré de l'exemple Quinlan (1993)).

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Table Données "weather" (Quinlan, 1993).

Devant un problème de prédiction, le vice entraîneur a représenté graphiquement ces observations :



Ce que le vice entraîneur a schématisé est appelé arbre de décision. Cet arbre de décision est dit aveugle si le choix de la racine temps est aléatoire. Il peut être enrichi par des indicateurs spécialement si les décisions sont incertaines.

2.2. Définition d'un arbre de décision

Les arbres de décision sont des méthodes ou des techniques très connues et utilisées dans le domaine de l'intelligence. Ce sont des méthodes graphiques pour analyser des décisions avec risque, des modèles dans lesquels les probabilités associées aux différents états de la nature sont spécifiées. Plus précisément, les arbres de décisions ont été conçus pour les problèmes faisant intervenir une séquence de décisions et événements successifs.

2.3. Applications

Les applications des arbres de décisions souvent liées sont :

- Classification
- Apprentissage
- Prédiction

Les domaines sont ceux de datamining socio-économique, industriel... enrichi des domaines de la vie quotidienne nécessitant une stratégie tels que :

- les jeux de société et jeux stratégiques
- les jeux de prédiction

2.4. Représentation par arbre de décision

Un arbre de décision permet de classer un objet à l'aide de questions : chaque nœud de l'arbre représente une question sur l'un des attributs de l'objet, chaque lien est une réponse à la question

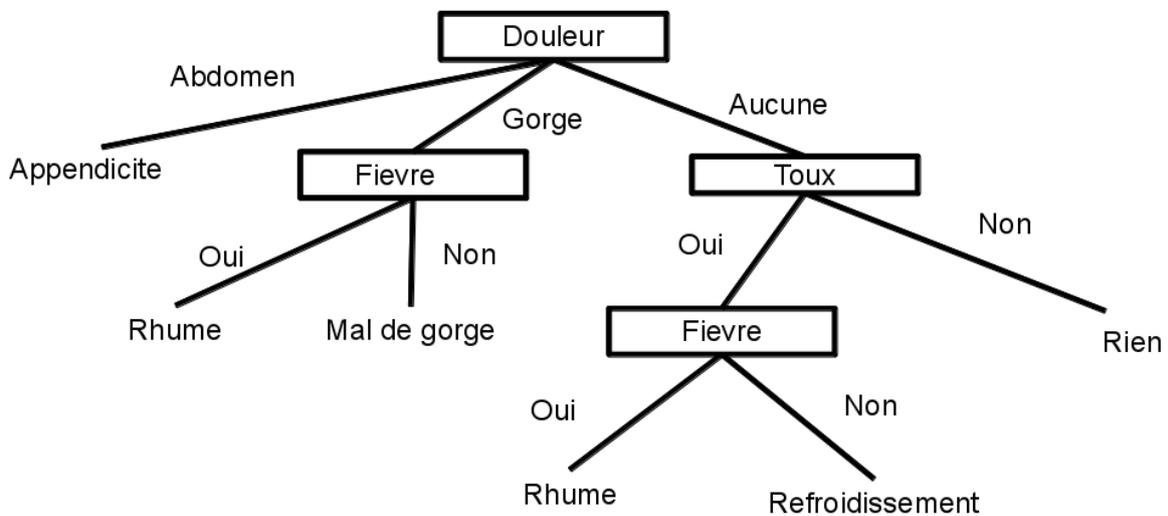
Un arbre est constitué :

- de noeuds qui testent les attributs
- de branches qui représentent chacune une valeur de l'attribut testé dans le noeud dont elles sont issues
- de feuilles (ce sont les noeuds terminaux de l'arbre) qui indiquent la classe résultante(décision finale)

Exemple illustratif d'un arbre de décision : Prédiction de maladie à partir des symptômes. En apprentissage supervisé les valeurs classes ou des décisions (ici maladie) est déjà connue pour un ensemble de cas.

Fievre	Douleur	Toux	Maladie(La classe, la décision)
oui	Abdomen	non	Appendicite
non	Abdomen	oui	Appendicite
oui	gorge	non	rhume
oui	gorge	oui	rhume
non	gorge	oui	mal de gorge
oui	non	non	aucune
oui	non	oui	rhume
non	non	oui	refroidissement
non	non	non	aucune

Un des arbres qu'on peut dessiner sans tenir compte des indicateurs, des probabilités et du type de nœud est le suivant :

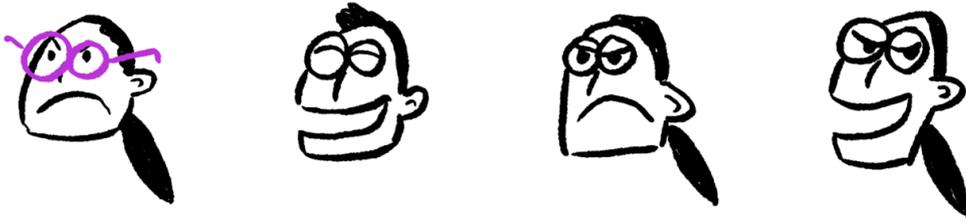


2.5. Mesures d'exploration

L'arbre est dit aveugle si le choix de la racine est aléatoire. Toutefois, des mesures sont disponibles tel que l'entropie, le gain.....etc. pour le choix de la racine et l'exploration efficace.

Exemple :

On utilise un arbre de décision pour la reconnaissance de visages. Chaque nœud de l'arbre est une question. Les feuilles sont les classes de visages. (pris de brilliant.org)¹



Comment choisir la question racine pour diviser en 2 les classes ci-dessus.

- A-t-il de cheveux longs ?
- A-t-il des lunettes ?
- a-il les yeux fermés ?
- aucune de ces questions ne peut diviser les classes en 2 groupes de 2.

La réponse avec explication en bas de page²

Le choix de la racine est alors important lors de la construction des arbres

2.6. Discrétisation

Lorsque les valeurs sont numériques, il faut les discrétiser dans des intervalles.

Par exemple le tableau d'origine de weather pour les décisions play tennis or not contenait des températures numériques et des taux d'humidité numériques. Avant de représenter, on doit d'abord diviser selon les seuils pour obtenir une humidité élevée ou normale.

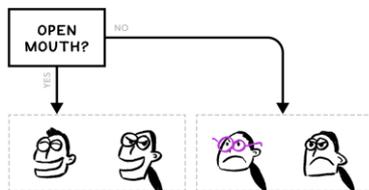
Travail à faire : Exemple Banque

Une banque souhaite réaliser un mailing pour promouvoir une offre commerciale. Afin de cibler les clients à qui elle adressera cette offre, elle procède à une enquête préliminaire pour savoir qui est susceptible d'être intéressé. Trois attributs descriptifs sont retenus : l'âge (trois tranches : [18; 34], [35; 49] et [50 et plus]), le genre (H ou F), et le fait d'être ou non propriétaire de son logement (oui,non).

L'attribut cible prend deux valeurs : oui si le client est intéressé, non sinon. Le résultat de l'enquête donne le tableau suivant :

¹ <https://brilliant.org/practice/making-decisions-2/?p=5>

² Aucune des questions ci-dessus n'est utile. Pour diviser en deux groupes équitables la question racine serait : A-t-il la bouche ouverte ? le schéma sera alors :



Age	Genre	Propriétaire	Intéressé
20	H	N	N
25	F	N	N
32	H	O	O
34	H	O	O
37	H	N	O
41	F	O	N
45	H	O	O
45	F	O	N
52	H	O	N
60	F	O	N
?	F	N	N
28	H	?	O
?	F	?	N

Discretiser les valeurs numériques. En utilisant un outil en ligne d'arbre de décision, dessiner un arbre correspondant au tableau.

//a suivre