

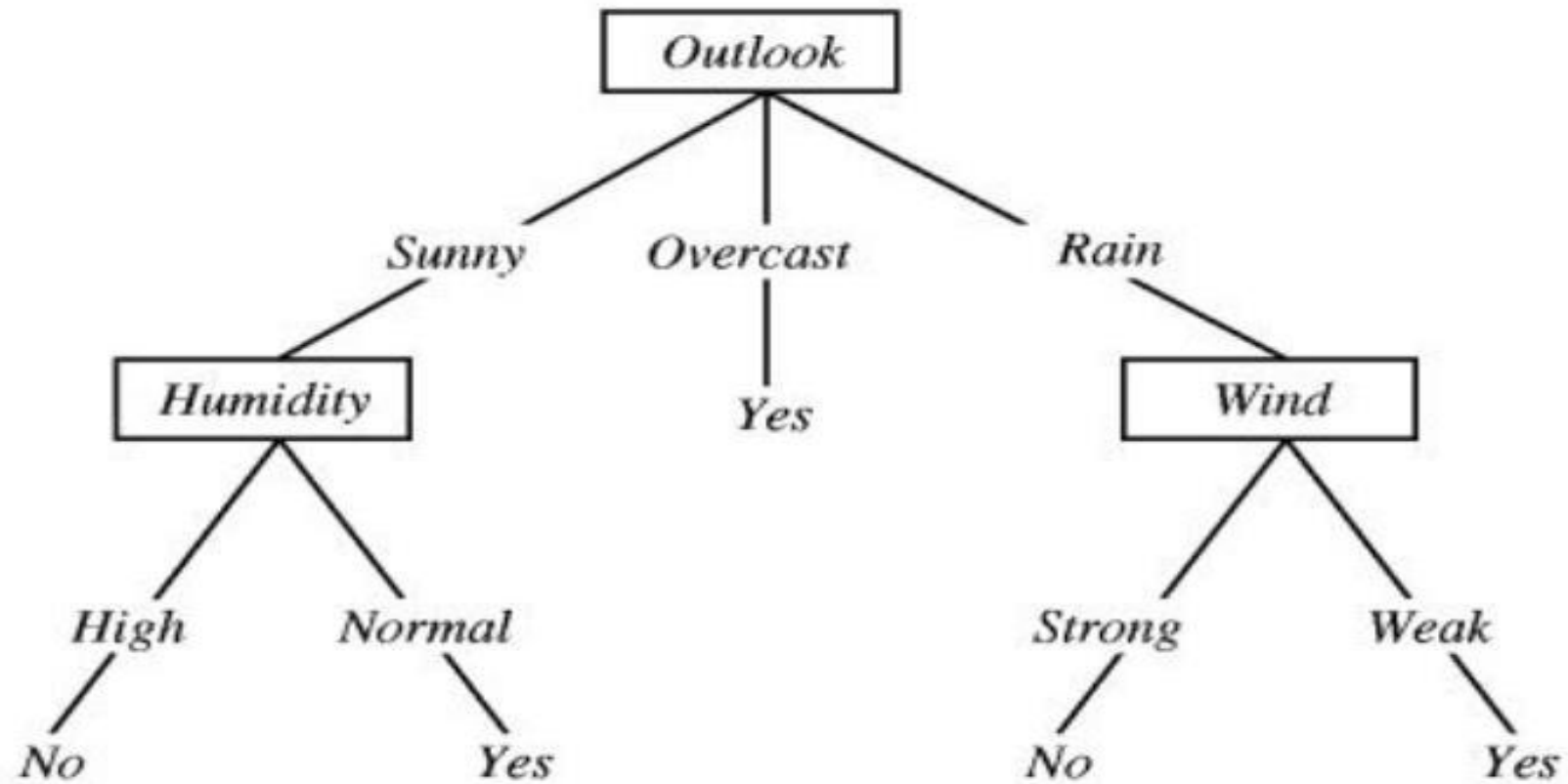
Chapitre II. . Elements d'analyse de Décision et décision sous risque

PARTIE II ARBRES DE DECISION

DR. DEKHICI L.

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

Exemple Weather de data



Définition d'un arbre de décision

- ▶ Decision tree

Applications

- 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

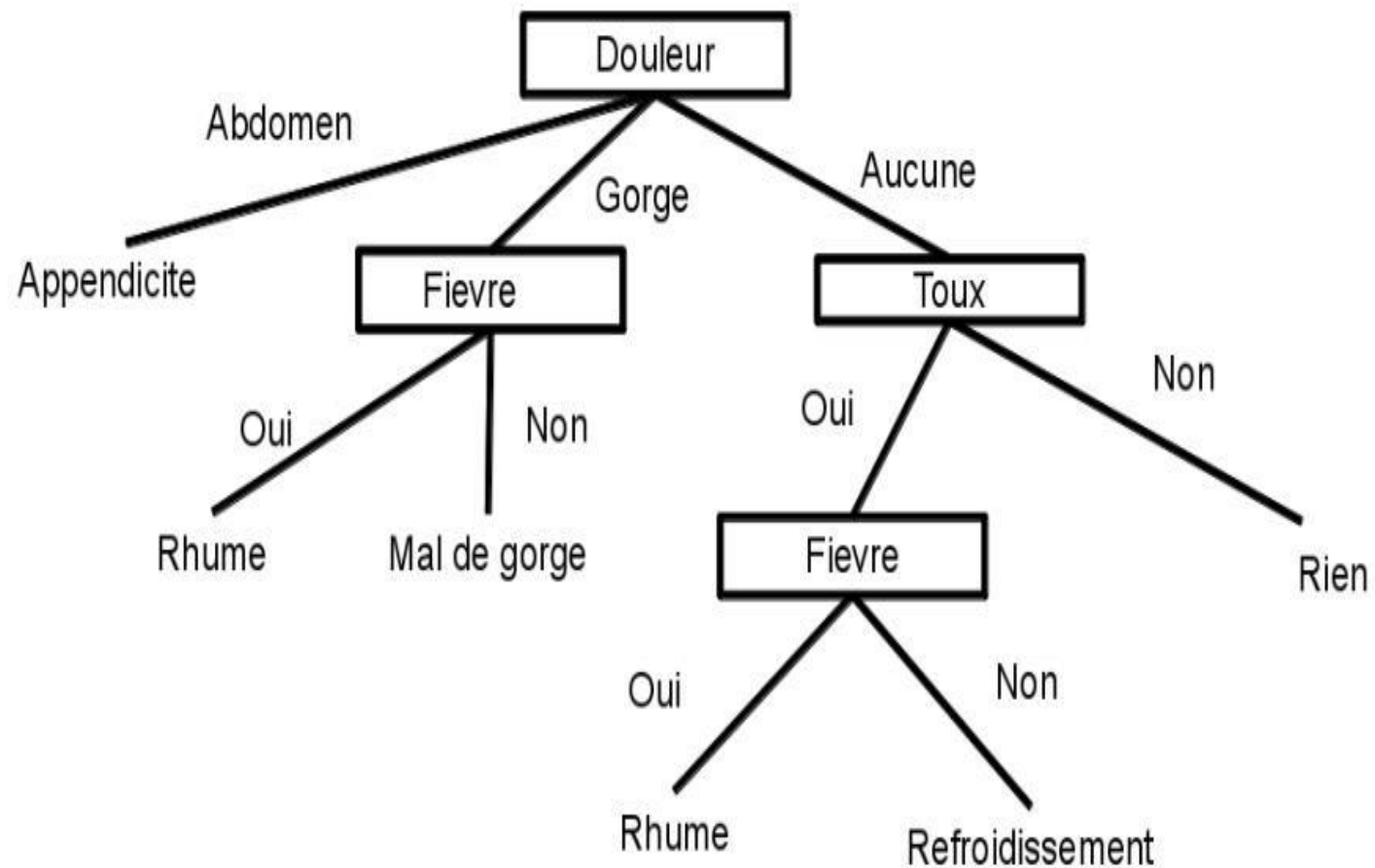
Exemple, Maladie

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 :

Arbre de décision: Maladie



Choix de la bonne racine

Comment choisir la question racine pour diviser en 2 groupes de 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.



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.

Discrétisation

- ▶ Numérique– catégorie

Par exemple le tableau d'origine de weather pour les décisions play tennis or not contenait des température numériques et des taux d'humidité numériques.

Exemple banque

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).

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

A partir d'une matrice de décision, Exemple d'Assurance,

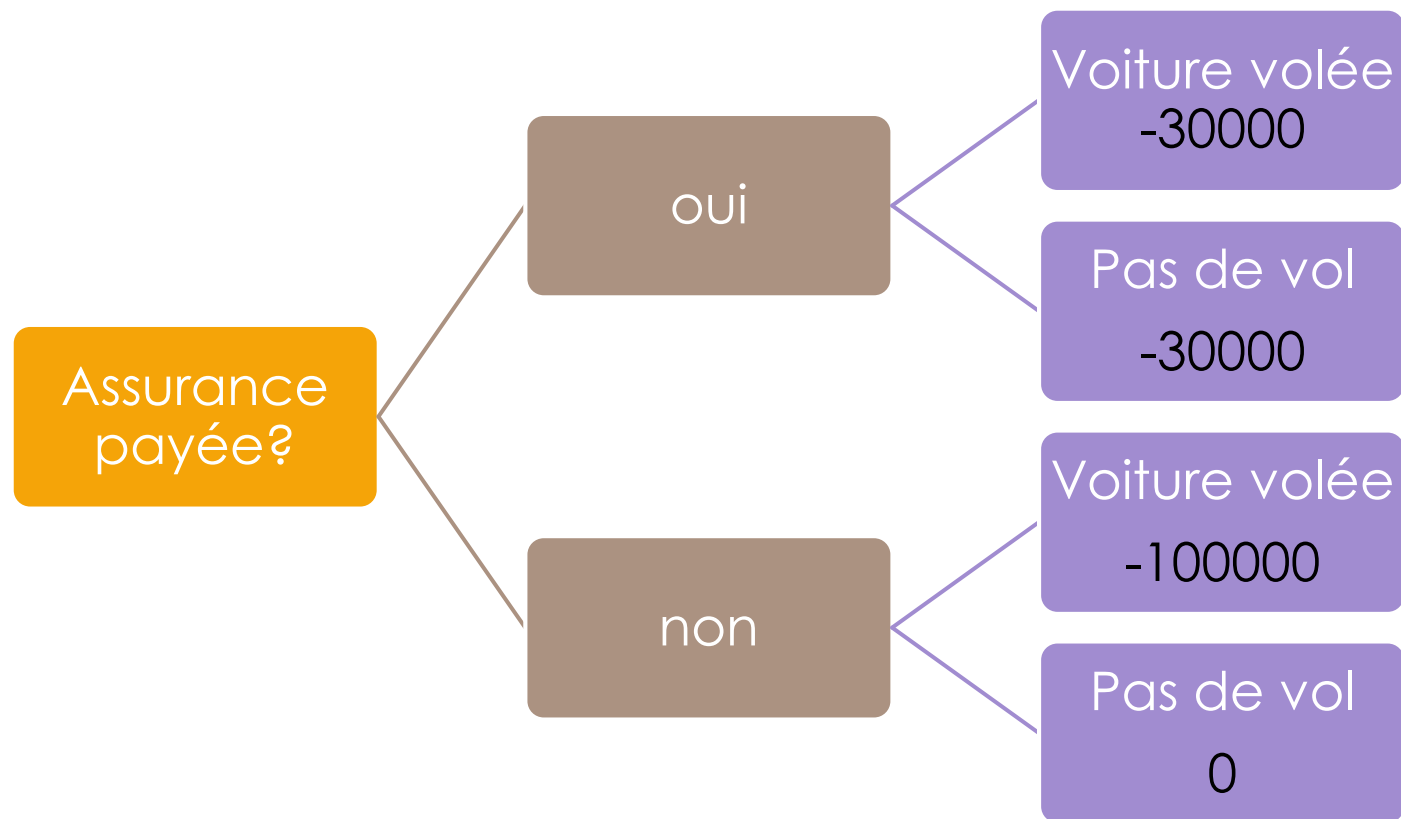
13



Exemple 1

Décisions\ Etats	E1= voiture volée	E2= pas de vol
D1= payer l'assurance de vol	-30000 DA	-30000DA
D2= ne pas payer l'assurance	-1000000 DA	0

De la décision vers les résultats



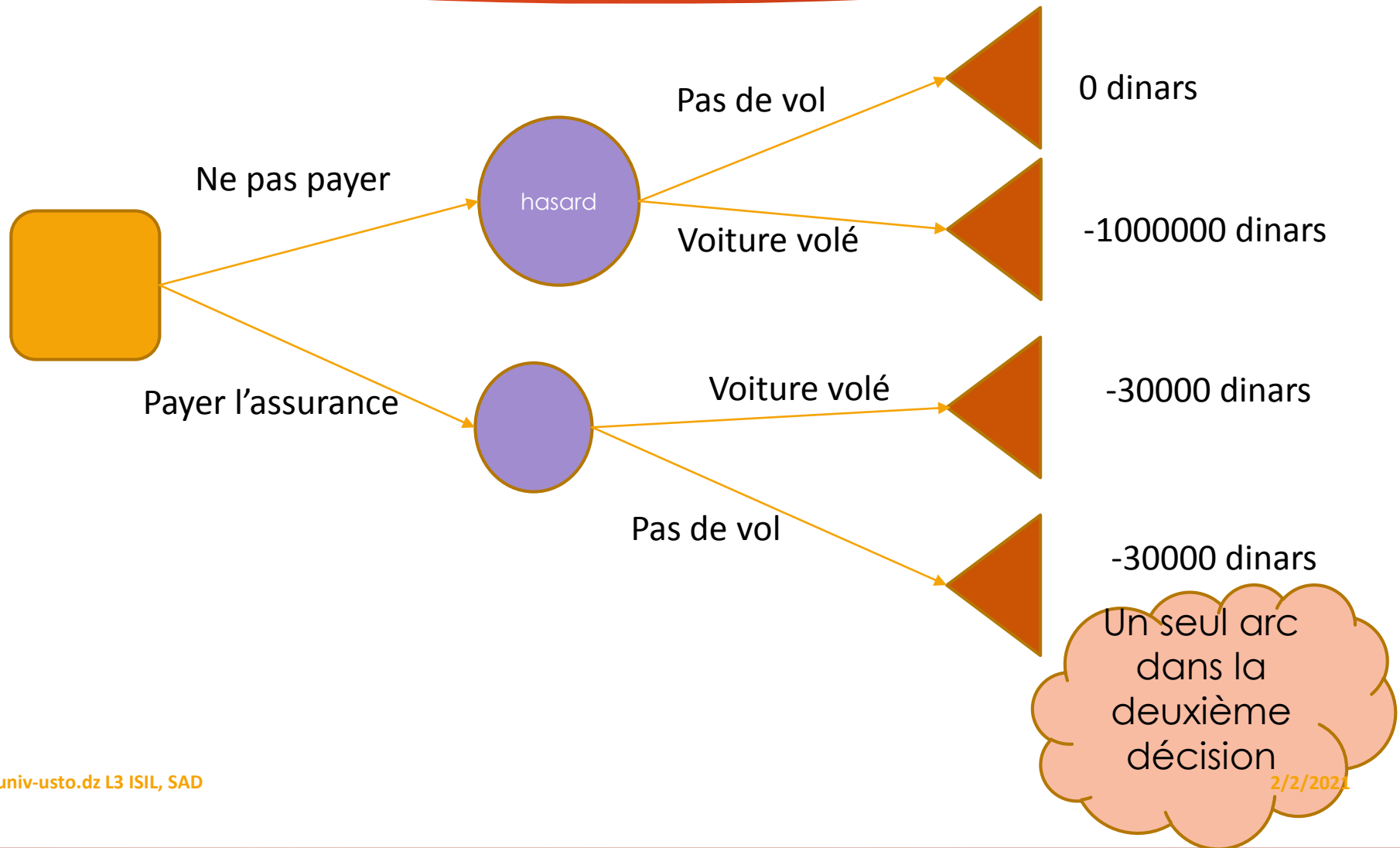
D'autres formalismes sont possibles



Avec d'autres formalismes



15



A partir d'une matrice de décision, Exemple 2



Exemple 2

Décisions \ Etats	E1 = pluie	E2 = pas de pluie
D1 = prendre mon parapluie	+Pas mouillée -Encombrée	+Pas mouillée -encombrée
D2 = ne pas prendre mon parapluie	- mouillée	+Pas mouillée

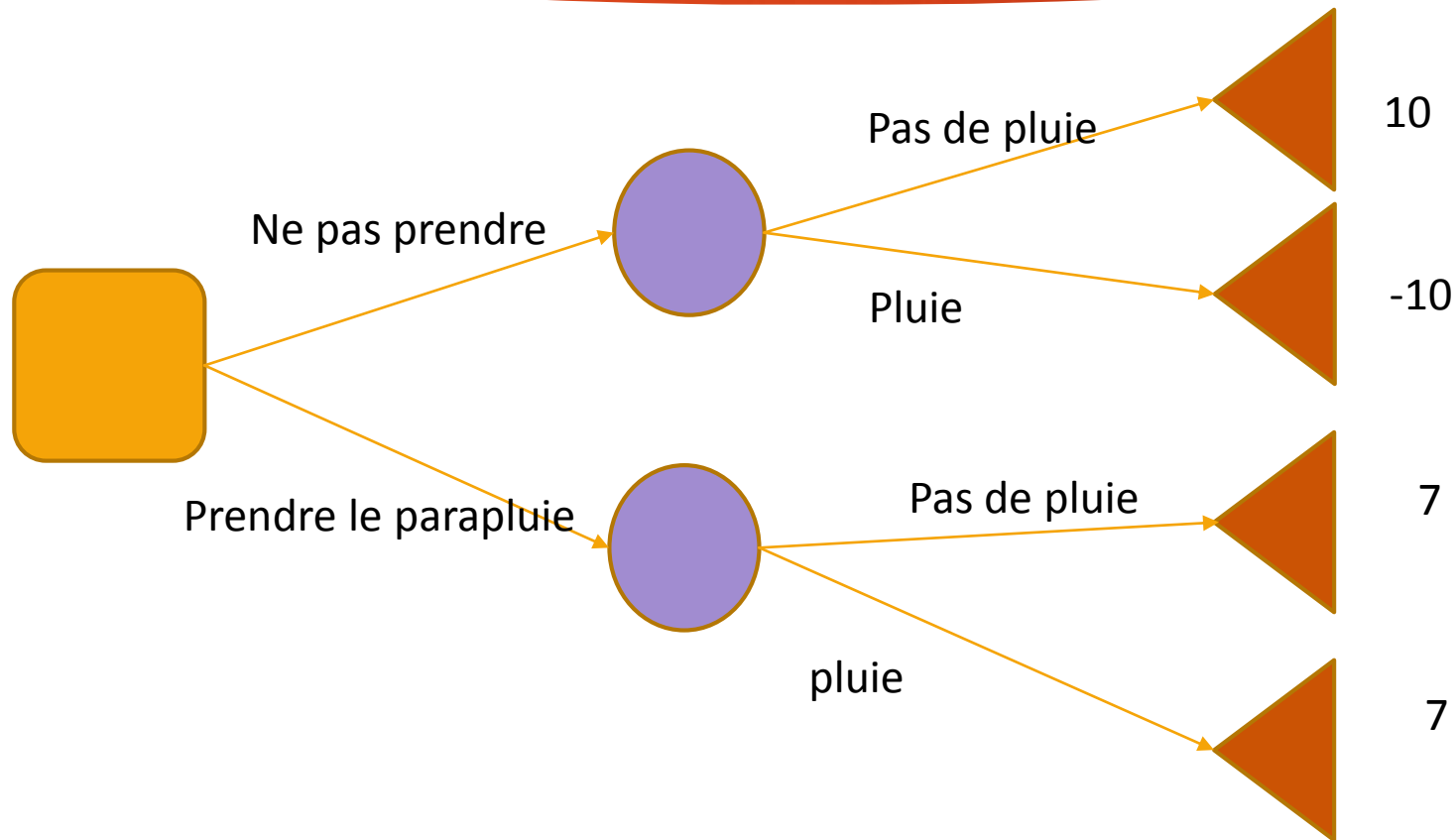
L'utilité , valeur numérique



Exemple 2

Décisions\ Etats	E1= pluie	E2= pas de pluie
D1= prendre mon parapluie	+10 favorable -3 défavorable	+10 -3
D2= ne pas prendre mon parapluie	-10	+10

Avec d'autres formalismes



Décision dans le risque

ARBRE DE DÉCISION
AVEC RISQUE

Décision dans le Risque

- ▶ Décision dans **l'incertitude** : situations de choix où les résultats des actions ne peuvent être prévus avec certitude.
- ▶ On suppose que cette incertitude est **probabilisée**, c'est à dire que le résultat obtenu ne dépend que de la réalisation d'évènements de probabilités connues. On utilise alors le terme de « **décision dans le risque** »

Les états dans les matrices de décisions peuvent avoir une probabilités

Décision avec certitude

Quel
kiosque
ouvrir?



Utilité= 4000 DA/jours



Utilité= 1000 DA/jours

Décision sous risque

Quel
kiosque
ouvrir?



$$U = 4000 \times 0.3 - 0.7 \times 1000 = 1200 - 700 = 400$$

Succes 0.3 → Utilité= 4000 DA/jours

Echec 0.7 → Utilité= -1000 DA/jours



$$U = 1000 \times 0.9 - 0.1 \times 200 = 900 - 20 = 880$$

Succes 0.9 → Utilité= 1000 DA/jours

Echec 0.1 → Utilité= -200 DA/jours

Utilité espérée (Expected Utility)

- ▶ $U(\text{noeud } j) = \text{somme } P(E_i) * U(E_i)$
- ▶ Utilités pondérées par les probabilités des états

Utilité espérée Exemple d'Assurance

24



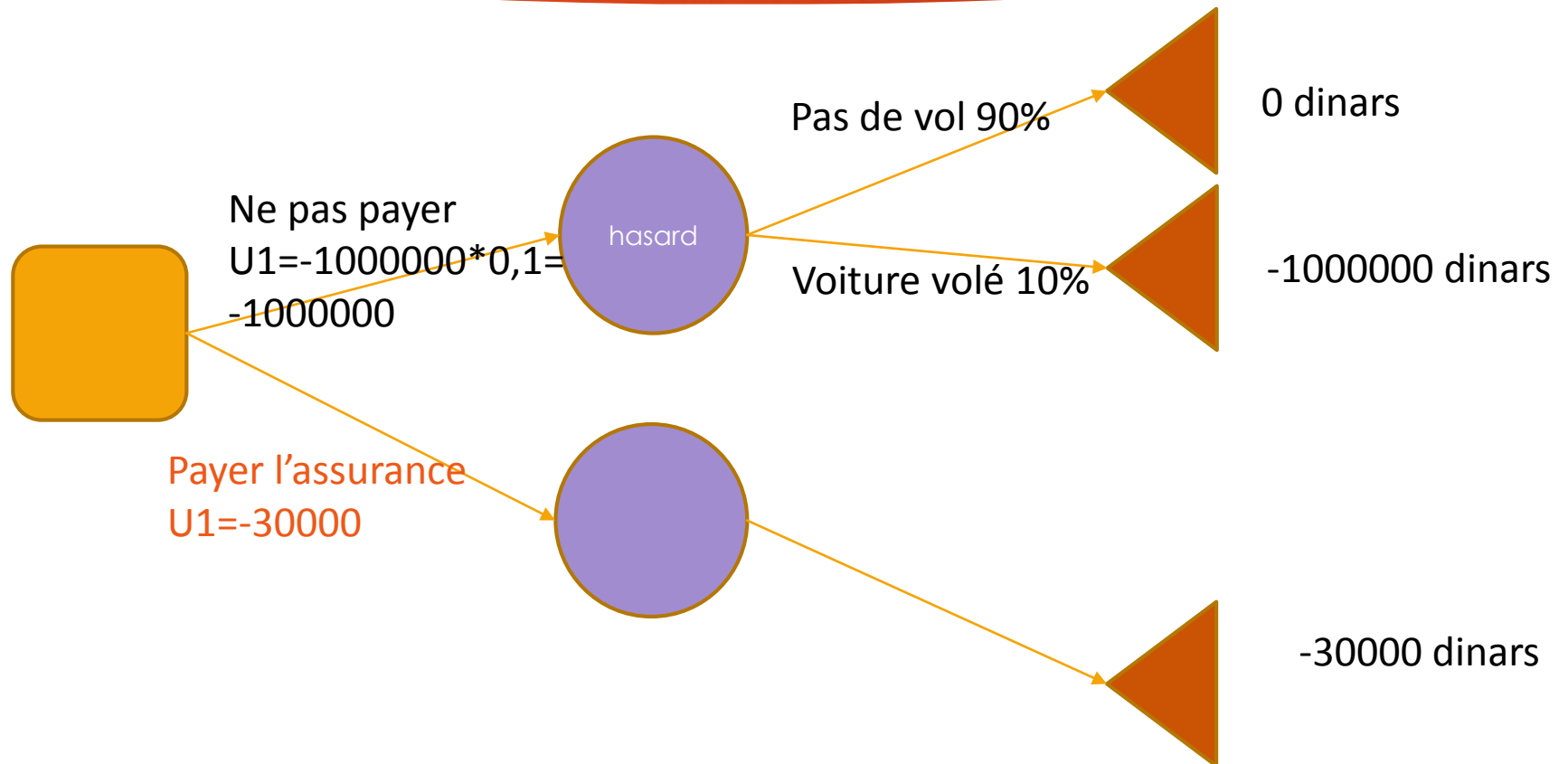
Exemple 1

Décisions \ Etats	E1 = volée (10%) voiture	E2 = pas de vol (90%)
D1 = payer l'assurance de vol	-30000 DA	-30000 DA
D2 = ne pas payer l'assurance	-1000000 DA	0

Arbre sous risque



25



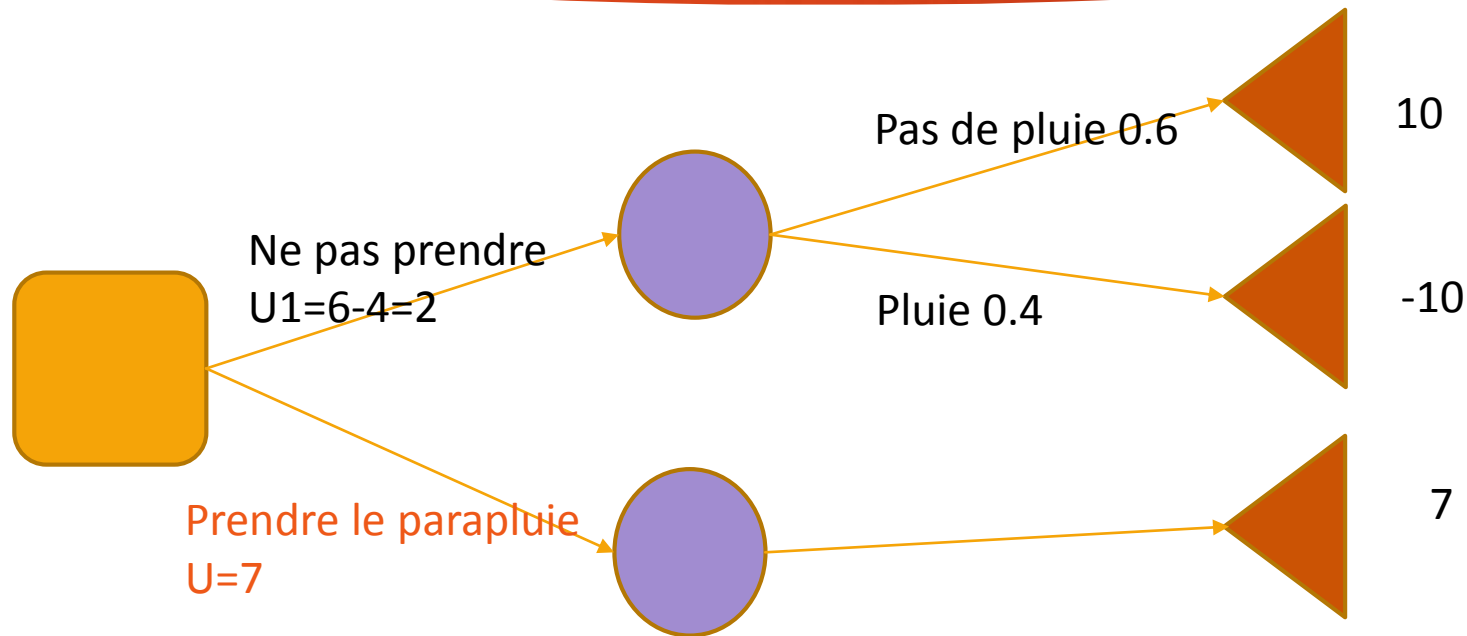
L'utilité , valeur numérique et états propobolisés



Exemple 2

Décisions\ Etats	E1= pluie (40%)	E2= pas de pluie (60%)
D1= prendre mon parapluie	+10 favorable -3 défavorable	+10 -3
D2= ne pas prendre mon parapluie	-10	+10

Utilité espérée , exemple de parapluie



Construction des arbres avec indices

- ▶ On dénombre plusieurs algorithmes pour construire des arbres de décision, parmi lesquels : ID3 (*Iterative Dichotomiser 3*), C4.5, C5 (successeurs d'ID3), CHAID (*CHi-squared Automatic Interaction Detector*) ; CART (*Classification And Regression Tree*)
- ▶ Tous ces algorithmes se distinguent par le ou les critères de segmentation utilisés, par les méthodes d'élagages implémentées, par leur manière de gérer les données manquantes dans les prédicteurs.

Implémentation des arbres avec indices

- ▶ Beaucoup de logiciels de fouille de données proposent des bibliothèques permettant d'implémenter un ou plusieurs algorithmes d'apprentissage par arbre de décision. Par exemple, le logiciel Open Source R contient plusieurs implémentations de CART, telles que rpart, party et randomForest, les logiciels libres Weka et Orange (et son module orngTree) ou encore la bibliothèque libre Python scikit-learn ; mais également Salford Systems CART, IBM SPSS Modeler, RapidMiner, SAS Enterprise Miner, KNIME, Microsoft SQL Server .