

Arbres de décision et classification bayésienne

Exercice 1 Soit les exemples suivants ayant trois attributs et appartenant à deux classes :

N°	Att1	Att2	Att3	Classe	N°	Att1	Att2	Att3	Classe
1	A	L	T	C1	8	B	L	T	C1
2	A	H	T	C2	9	B	H	F	C1
3	A	H	F	C2	10	C	H	T	C2
4	A	H	F	C2	11	C	L	T	C2
5	A	L	F	C1	12	C	H	F	C1
6	B	H	T	C1	13	C	H	F	C1
7	B	H	F	C1	14	C	H	F	C1

En utilisant l'algorithme ID3 :

1. Calculer l'entropie de l'ensemble d'exemples par rapport à la valeur de la classe.
2. Construire l'arbre de décision appris de cette table en utilisant l'algorithme ID3.
3. Donner la précision de l'arbre appris sur l'ensemble d'entraînement.
4. Reconstruire l'arbre en utilisant le Gini Index.
5. Construire le classifieur bayésien naïf correspondant
6. En donner les classes des exemples suivants :

N°	Att1	Att2	Att3
1	A	H	F
2	B	L	T
3	C	L	F

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. Reconstruire l'arbre en utilisant le Gini Index.
3. Construire le classifieur bayésien naïf correspondant.
4. Donner le risque du patient ayant les attributs (Jeune,V,F) selon les trois modèles. construits.

Exercice 3 Nous considérerons l'ensemble d'exemples représentant la nature de différents échantillons de champignons : toxique ou non selon les critères de couleur, taille, forme et le milieu de croissance :

Couleur	Taille	Forme	Milieu	Toxique
marron	petit	plat	terre	oui
jaune	petit	sphère	terre	oui
marron	moyen	conique	bois	non
blanc	moyen	sphère	terre	non
blanc	grand	plat	terre	non

1. Construire l'arbre de décision correspondant à cet ensemble en utilisant l'algorithme ID3.
2. Reconstruire l'arbre en utilisant le Gini Index.
3. Donner la précision des deux arbres et leurs moyennes harmoniques sur la table de test suivante :

Couleur	Taille	Forme	Milieu	Toxique
marron	grand	plat	bois	non
blanc	moyen	conique	terre	oui
jaune	moyen	sphère	terre	oui
marron	moyen	conique	bois	non
jaune	petit	plat	terre	oui

4. Construire le classifieur bayésien naïf correspondant à la base d'entraînement.
5. Calculer la précision de ce modèle sur la base de la question 3.

Responsable de la matière :
Dr A.Djeffal
