
Examen

Questions de cours (4.5 pts : 1.5 + 1.5 + 1.5)

1. Expliquer comment peut-on utiliser les règles d'association pour la classification supervisée.
2. Expliquer comment peut-on utiliser l'algorithme Apriori pour l'obtention des motifs rares.
3. Expliquer comment peut-on étendre la méthode SVM binaire pour résoudre les cas multi-classes.

Exercice 1 (15.5 pts : 6 + 4 + 3.5 + 2)

Une banque dispose des informations suivantes sur un ensemble de clients :

client	M	A	R	E	I
01	moyen	moyen	village	oui	oui
02	élevé	moyen	bourg	non	non
03	faible	âgé	bourg	non	non
04	faible	moyen	bourg	oui	oui
05	moyen	jeune	ville	oui	oui
06	élevé	âgé	ville	oui	non
07	moyen	âgé	ville	oui	non
08	faible	moyen	village	non	non

L'attribut client indique le numéro du client ; l'attribut M indique la moyenne des crédits sur le compte du client ; l'attribut A donne la tranche d'âge ; l'attribut R décrit la localité du client ; l'attribut E possède la valeur oui si le client possède un niveau d'études supérieur au bac ; l'attribut I (la classe) indique si le client effectue ses opérations de gestion de compte via Internet.

A. Règles d'association (6 pts : 1 + 2 + 1.5 + 0.5)

1. Donner la base de données formelle correspondant à cette base.
2. Calculer les motifs fréquents correspondant à un minimum de support $\sigma_s = 0.4$
3. Calculer les règles solides correspondant à un seuil de confiance = 0.9
4. En déduire un modèle de décision.

B. Arbre de décision (4 pts : 3 + 1)

1. Construire l'arbre de décision correspondant à cette base en utilisant l'algorithme ID3.
2. Donner la précision de l'arbre construit sur la base suivante :

client	M	A	R	E	I
01	moyen	âgé	village	oui	oui
02	élevé	jeune	ville	non	oui
03	faible	âgé	village	non	non
04	moyen	moyen	bourg	oui	non

B. Classification bayésienne (3.5 pts :2 + 1.5)

1. Donner le modèle de décision déduit de cette base en utilisant la classification bayésienne naïve.
2. Trouver les classes des exemples suivants :

client	M	A	R	E
01	?	âgé	?	oui
02	élevé	?	ville	?
03	faible	?	?	?
04	?	moyen	bourg	?

B. Combinaison de modèles (2 pts)

1. Utiliser le modèle combiné par vote des trois modèles précédents pour trouver les classes des exemples suivants :

client	M	A	R	E
01	moyen	jeune	bourg	oui
02	élevé	moyen	village	oui
03	faible	jeune	ville	non
04	moyen	jeune	bourg	non

Bonne Chance

Dr A.Djeffal

Corrigé type

Questions de cours (4.5 pts)

1. L'idée est de rechercher les règles solides contenant dans leur partie droite l'attribut classe, c-à-d de la forme :

$$Attribut_1 = v_{att1} \wedge Attribut_2 = v_{att2} \wedge \dots \wedge Attribut_n = v_{attn} \Rightarrow Classe = v_{classe}$$

Plusieurs études ont montré que cette technique est plus précise que certaines méthodes traditionnelles tel que les arbres de décision.

L'un des premiers algorithmes de classification associative est l'algorithme CBA (Classification-Based Association). Il utilise l'algorithme Apriori pour générer les règles d'association puis utilise une heuristique pour construire le classifieur. Les règles sont ordonnées selon leurs supports et confidences. Si plusieurs règles ont la même partie gauche, la règle de la confiance la plus élevée est utilisée dans le classifieur. Pour classer un nouveau tuple, la première règle le satisfaisant est utilisée. Le classifieur contient aussi une règle par défaut pour classer les tuple dont une règles satisfaisante n'existe pas. **1.5pt**

2. Une légère modification d'Apriori suffit pour conserver les MRM. Si le support d'un candidat est inférieur au support minimum, alors à la place de l'effacer nous l'enregistrons dans l'ensemble des motifs rares minimaux

Tous les motifs rares sont retrouvés à partir des motifs rares minimaux. Pour cela nous avons besoin de générer tous les sur-motifs possibles des MRM. **1.5pt**

3. Les méthodes des machines à vecteur support multiclasse, réduisent le problème multiclasse à une composition de plusieurs hyperplans biclasses permettant de tracer les frontières de décision entre les différentes classes. Ces méthodes décomposent l'ensemble d'exemples en plusieurs sous ensembles représentant chacun un problème de classification binaire. Pour chaque problème un hyperplan de séparation est déterminé par la méthode SVM binaire. On construit lors de la classification une hiérarchie des hyperplans binaires qui est parcourue de la racine jusqu'à une feuille pour décider de la classe d'un nouvel exemple. On trouve dans la littérature plusieurs méthodes de décomposition : 1vs1, 1vsR, DAG, ...etc. **1.5pt**

A. Règles d'association (6 pts :2 + 2 + 1.5 + 0.5)

1. Donner la base de données formelle correspondant à cette base. **2pt**

a :M=moyen, b :M=faible, c :M=élevé, d :A=âgé, e :A=moyen, f :A=jeune, g :R=village, h :R=ville, i :R=bourg, j :E=oui, k :E=non, l :I=oui, m :I=non

La base formelle

	a	b	c	d	e	f	g	h	i	j	k	l	m
1	1	0	0	0	1	0	1	0	0	1	0	1	0
2	0	0	1	0	1	0	0	0	1	0	1	0	1
3	0	1	0	1	0	0	0	0	1	0	1	0	1
4	0	1	0	0	1	0	0	0	1	1	0	1	0
5	1	0	0	0	0	1	0	1	0	1	0	1	0
6	0	0	1	1	0	0	0	1	0	1	0	0	1
7	1	0	0	1	0	0	0	1	0	1	0	0	1
8	0	1	0	0	1	0	1	0	0	0	1	0	1

2. Calculer les motifs fréquents correspondant à un minimum de support $\sigma_s = 0.4$ **2pts**

- Selon Weka : $\sigma_s = 0.4 \Rightarrow 3$ instances, les motifs fréquents :

$$F1 = \{a, b, d, e, h, i, j, k, l, m\}$$

$$F2 = \{aj, dm, hj, jl, km\}$$

$$F3 = \phi$$

$$F = F1 \cup F2$$

- Selon un calcul dur : $\sigma_s = 0.4 \Rightarrow 4$ instances

$$F1 = \{e, j, m\}$$

$$F2 = \phi$$

$$F = F1$$

3. Calculer les règles solides correspondant à un seuil de confiance = 0.9 **1.5pt**

- Selon Weka : $\sigma_s = 0.4 \Rightarrow 3$ instances, les règles solides :

$$- a \Rightarrow j, d \Rightarrow m, h \Rightarrow j, l \Rightarrow j, k \Rightarrow m$$

- Selon un calcul dur : $\sigma_s = 0.4 \Rightarrow 4$ instances, les règles solides :

Aucune

4. En déduire un modèle de décision. **0.5pt**

- Selon Weka : $\sigma_s = 0.4 \Rightarrow 3$ instances, le modèle :

$$- d \Rightarrow m$$

$$- k \Rightarrow m$$

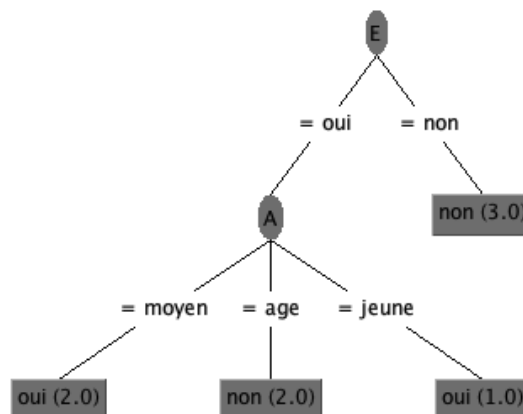
- m (par défaut)

- Selon un calcul dur : $\sigma_s = 0.4 \Rightarrow 4$ instances, le modèle :

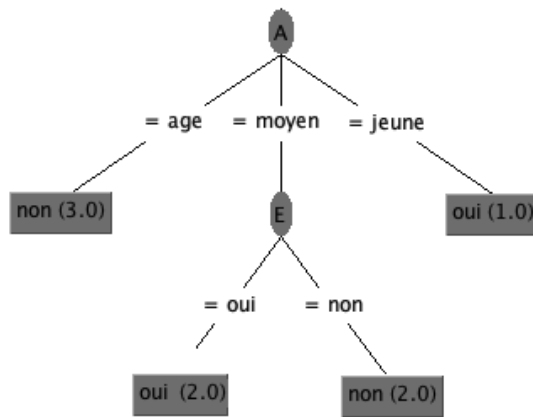
- m (par défaut : la classe la plus fréquente).

B. Arbre de décision (4 pts :3 + 1)

1. Construire l'arbre de décision correspondant à cette base en utilisant l'algorithme ID3. **3pts**



ou



2. Donner la précision de l'arbre construit sur la base suivante :

1pt

client	M	A	R	E	I
01	moyen	âgé	village	oui	oui
02	élevé	jeune	ville	non	oui
03	faible	âgé	village	non	non
04	moyen	moyen	bourg	oui	non

Précision = $\frac{2}{4} = 50\%$

B. Classification bayésienne (3.5 pts :2 + 1.5)

1. Donner le modèle de décision déduit de cette base en utilisant la classification bayésienne naïve. **2pts**

I		oui (3/8)	non (5/8)
M	moyen	2/3	1/5
	élevé	0	2/5
	faible	1/3	2/5
A	moyen	2/3	2/5
	âgé	0	3/5
	jeune	1/3	0
R	village	1/3	1/5
	ville	1/3	2/5
	bourg	1/3	2/5
E	oui	3/3	2/5
	non	0	3/5

En utilisant l'estimateur de Laplace :

I		oui (3/8)	non (5/8)
M	moyen	3/6	1/5
	élevé	1/6	2/5
	faible	2/6	2/5
A	moyen	3/6	3/8
	âgé	1/6	4/8
	jeune	2/6	1/8
R	village	1/3	1/5
	ville	1/3	2/5
	bourg	1/3	2/5
E	oui	4/5	2/5
	non	1/5	3/5

2. Trouver les classes des exemples suivants :

client	M	A	R	E
01	?	âgé	?	oui
02	élevé	?	ville	?
03	faible	?	?	?
04	?	moyen	bourg	?

– Exemple 1 :

$$P(X/I=oui) = P(A=\hat{\text{a}}g\acute{e}/I=oui).P(E=oui/I=oui).P(I=oui)$$

$$= 1/6 \cdot 4/5 \cdot 3/8 = 0,05$$

$$P(X/I=non) = P(A=\hat{\text{a}}g\acute{e}/I=non).P(E=oui/I=non).P(I=non)$$

$$= 4/8 \cdot 2/5 \cdot 5/8 = 0,125$$

Classe = non

0.5pt

– Exemple 2 :

$$P(X/I=oui) = P(M=\acute{e}l\acute{e}v\acute{e}/I=oui).P(A=ville/I=oui).P(I=oui)$$

$$= 1/6 \cdot 1/3 \cdot 3/8 = 0,02$$

$$P(X/I=non) = P(M=\acute{e}l\acute{e}v\acute{e}/I=non).P(A=ville/I=non).P(I=non)$$

$$= 2/5 \cdot 2/5 \cdot 5/8 = 0,1$$

Classe = non

0.5pt

– Exemple 3 :

$$P(X/I=oui) = P(M=faible/I=oui).P(I=oui)$$

$$= 2/6 \cdot 3/8 = 0,125$$

$$P(X/I=non) = P(M=faible/I=non).P(I=non)$$

$$= 2/5 \cdot 5/8 = 0,25$$

Classe = non

0.5pt

– Exemple 4 :

$$P(X/I=oui) = P(A=moyen/I=oui).P(R=bourg/I=oui).P(I=oui)$$

$$= 3/6 \cdot 1/3 \cdot 3/8 = 0,06$$

$$P(X/I=non) = P(A=moyen/I=non).P(R=bourg/I=non).P(I=non)$$

$$= 3/8 \cdot 2/5 \cdot 5/8 = 0,09$$

Classe = non

0.5pt

B. Combinaison de modèles (2 pts)

1. Utiliser le modèle combiné par vote des trois modèles précédents pour trouver les classes des exemples suivants :

client	M	A	R	E
01	moyen	jeune	bourg	oui
02	élevé	moyen	village	oui
03	faible	jeune	ville	non
04	moyen	jeune	bourg	non

– Exemple 1 :

- Modèle associatif : Non
- Modèle d'arbre de décision : Oui
- Modèle bayésien :

$$P(X/\text{oui}) = P(\text{oui}) \cdot P(\text{moyen}/\text{oui}) \cdot P(\text{jeune}/\text{oui}) \cdot P(\text{bourg}/\text{oui}) \cdot P(\text{oui}/\text{oui})$$

$$= 3/8 \cdot 3/6 \cdot 2/6 \cdot 1/3 \cdot 4/5 = 72/4320 = 0,016$$

$$P(X/\text{non}) = P(\text{non}) \cdot P(\text{moyen}/\text{non}) \cdot P(\text{jeune}/\text{non}) \cdot P(\text{bourg}/\text{non}) \cdot P(\text{oui}/\text{non})$$

$$= 5/8 \cdot 1/5 \cdot 1/8 \cdot 2/5 \cdot 2/5 = 20/8000 = 0,0025$$

Classe : Oui

- Classe = Oui

0.5pt

– Exemple 2 :

- Modèle associatif : Non
- Modèle d'arbre de décision : Oui
- Modèle bayésien :

$$P(X/\text{oui}) = P(\text{oui}) \cdot P(\text{élevé}/\text{oui}) \cdot P(\text{moyen}/\text{oui}) \cdot P(\text{village}/\text{oui}) \cdot P(\text{oui}/\text{oui})$$

$$= 3/8 \cdot 1/6 \cdot 3/6 \cdot 1/3 \cdot 4/5 = 36/4320 = 0,0083$$

$$P(X/\text{non}) = P(\text{non}) \cdot P(\text{élevé}/\text{non}) \cdot P(\text{moyen}/\text{non}) \cdot P(\text{village}/\text{non}) \cdot P(\text{oui}/\text{non})$$

$$= 5/8 \cdot 2/5 \cdot 3/8 \cdot 1/5 \cdot 2/5 = 60/8000 = 0,0075$$

Classe : Oui

- Classe = Oui

0.5pt

– Exemple 3 :

- Modèle associatif : non
- Modèle d'arbre de décision : oui
- Modèle bayésien :

$$P(X/\text{oui}) = P(\text{oui}) \cdot P(\text{faible}/\text{oui}) \cdot P(\text{jeune}/\text{oui}) \cdot P(\text{ville}/\text{oui}) \cdot P(\text{non}/\text{oui})$$

$$= 3/8 \cdot 2/6 \cdot 2/6 \cdot 1/3 \cdot 1/5 = 12/4320 = 0,0027$$

$$P(X/\text{non}) = P(\text{non}) \cdot P(\text{faible}/\text{non}) \cdot P(\text{jeune}/\text{non}) \cdot P(\text{ville}/\text{non}) \cdot P(\text{non}/\text{non})$$

$$= 5/8 \cdot 2/5 \cdot 1/8 \cdot 2/5 \cdot 3/5 = 60/8000 = 0,0075$$

Classe : non

- Classe = non

0.5pt

– Exemple 4 :

- Modèle associatif : non
- Modèle d'arbre de décision : oui
- Modèle bayésien : Oui

$$P(X/\text{oui}) = P(\text{oui}) \cdot P(\text{moyen}/\text{oui}) \cdot P(\text{jeune}/\text{oui}) \cdot P(\text{bourg}/\text{oui}) \cdot P(\text{non}/\text{oui})$$

$$= 3/8 \cdot 3/6 \cdot 2/6 \cdot 1/3 \cdot 1/5 = 18/4320 = 0,0041$$

$$P(X/\text{non}) = P(\text{non}) \cdot P(\text{moyen}/\text{non}) \cdot P(\text{jeune}/\text{non}) \cdot P(\text{bourg}/\text{non}) \cdot P(\text{non}/\text{non})$$

$$= 5/8 \cdot 1/5 \cdot 1/8 \cdot 2/5 \cdot 3/5 = 30/8000 = 0,0038$$

Classe : Oui

- Classe = Oui

0.5pt