
Examen

Questions de cours (7 pts : 1.5 + 2.5 + 1.5 + 1.5)

1. Quel est le rôle de la fonction d'activation dans un réseau de neurone? Et quelle différence existe entre la méthode d'entraînement par descente de gradient et celle de Widrow-Hoff?
2. Soit les méthodes d'apprentissage supervisé suivantes : KPPV, GA, ID3, OneR, CBA, RNA, BN, SVM. Mettre dans chacune des catégories suivantes parmi les méthodes précédentes celles qui conviennent :
 - (a) Méthodes influencées par l'ordre de présentation des exemples.
 - (b) Méthodes nécessitant une discrétisation préalable des valeurs des attributs.
 - (c) Méthodes nécessitant une numérisation préalable des valeurs des attributs.
 - (d) Méthodes pouvant prendre en charge de nouveaux exemples sans avoir besoin de refaire l'entraînement.
 - (e) Méthodes pouvant prendre en charge des recommandations des experts.
3. Expliquer brièvement ce que c'est l'apprentissage semi-supervisé (par renforcement).
4. Expliquer brièvement le principe d'accessibilité de point de vue densité utilisé dans le clustering basé densité.

Exercice 1 Motifs fréquents (6 pts : 2 + 2 + 1 + 0.5 + 0.5)

Soit la table suivante :

Patient	Age	Résidus	Tension	Infection
01	Jeune	Positif	Élevée	oui
02	Jeune	Négatif	Elevée	non
03	Vieux	Positif	Élevée	non
04	Vieux	Positif	Normale	non
05	Jeune	Positif	Élevée	oui
06	Jeune	Négatif	Normale	non

1. Transformer la table en une base formelle
2. En considérant un support minimum de 50 %, trouver :
 - les motifs fréquents,
 - Les motifs fréquents fermés,
 - Les motifs fréquents maximaux,
 - Les motifs rares minimaux.

Exercice 2 Classification supervisée (5.5 pts : 0.5 + 1 + 2.5 + 1.5)

Considérons la table de l'exercice précédent où l'attribut Infection représente la classe :

1. Donner le modèle de décision construit en utilisant la méthode ZeroR,
2. Donner le modèle de décision construit en utilisant la méthode OneR,
3. Donner le modèle de décision construit en utilisant la méthode ID3,
4. Donner le modèle construit en utilisant une combinaison des trois modèles précédents par la méthode Stacking à base d'une classification bayésienne naïve.

Exercice 3 Clustering (1.5 pts : 0.75 + 0.75)

Soit la table suivante où l'attribut Infection représente la classe :

Patient	Age	Résidus	Tension	Infection
01	25	+1	15	oui
02	30	-1	17	non
03	65	+1	16	non
04	70	+1	12	non
05	38	+1	15	oui
06	42	-1	11	non

1. Effectuer un clustering des exemples de la base en utilisant le clustering incrémental avec les paramètres suivant :
 - Seuil de distance : 25
 - Ordre de présentation des exemples = ordre dans la table
 - Distance : City blocs
2. Donner la table de correspondance classe-cluster puis l'affectation des clusters. En déduire le taux de clustering

Bonne Chance

Dr A.Djeffal

Corrigé type

Questions de cours

1. Le rôle de la fonction d'activation est de contrôler la sortie du neurone. **(0.75 pt)**
 La méthode de descente de gradient mis à jour les poids globalement à la fin, par contre la méthode Widrow-Hoff mis à jour les poids à chaque passage d'un exemple. **(0.75 pt)**
2. (a) a : RNA **(0.5 pt)**
 (b) b : ID3, OneR, CBA, BN **(0.5 pt)**
 (c) c : RNA, SVM **(0.5 pt)**
 (d) d : OneR, BN, RNA **(0.5 pt)**
 (e) e : BN **(0.5 pt)**
3. L'apprentissage semi-supervisé utilise les informations fournies par les exemples non classés pour renforcer l'apprentissage à partir des exemples classés. **(1.5 pt)**
4. Dans le clustering basé densité,
 - deux exemples sont voisins si la distance entre eux est inférieure à un certain seuil.
 - Un exemple est dense, si le nombre de ses voisins dépasse un certain nombre
 - Un exemple est directement densité-accessible à partir d'un autre si les deux exemples sont denses et voisins,
 - Un exemple p est densité-accessible à partir d'un autre q s'il existe une suite d'exemples q_1, \dots, q_n les reliant tel que q_{i+1} est directement densité-accessible à partir de q_i **(1.5 pt)**

Exercice 1

1. Table formelle : **(2 pts)**

Patient	Age		Résidus		Tension		Infection	
	Jeune	Vieux	Positif	Négatif	Élevée	Normale	oui	non
01	1	0	1	0	1	0	1	0
02	1	0	0	1	1	0	0	1
03	0	1	1	0	1	0	0	1
04	0	1	1	0	0	1	0	1
05	1	0	1	0	1	0	1	0
06	1	0	0	1	0	1	0	1

2. $F1 = \{\text{Age}=\text{Jeune}, \text{Résidus}=\text{Positif}, \text{Tension}=\text{Élevée}, \text{Infection}=\text{non}\}$,
 $F2 = \{\text{Age}=\text{Jeune}, \text{Tension}=\text{Élevée}, \text{Résidus}=\text{Positif}, \text{Tension}=\text{Élevée}\}$
 - Motifs fréquents : $F1 \cup F2$ **(2 pts)**
 - Motifs fréquents fermés : $F1 \cup F2$ **(1 pt)**
 - Motifs fréquents maximaux : $\{\text{Infection}=\text{non}, \text{Age}=\text{Jeune}, \text{Tension}=\text{Élevée}, \text{Résidus}=\text{Positif}, \text{Tension}=\text{Élevée}\}$ **(0.5 pt)**
 - Motifs rares minimaux : $\{\text{Age} = \text{Vieux}, \text{Résidus} = \text{Négatif}, \text{Tension} = \text{Normale}, \text{Infection} = \text{Oui}, \text{Age} = \text{Jeune}, \text{Résidus} = \text{Positif}, \text{Age} = \text{Jeune}, \text{Infection} = \text{non}, \text{Résidus} = \text{Positif}, \text{Infection} = \text{non}, \text{Tension}=\text{Élevée}, \text{Infection} = \text{non}, \text{Age}=\text{Jeune}, \text{Tension}=\text{Élevée}, \text{Résidus}=\text{Positif}\}$ **(0.5 pt)**

Exercice 2

1. Modèle ZeroR : Infection = non ; (0.5 pt)

2. Modèle OneR :

- Si Age = Jeune alors Infection = Oui
- Si Age = Vieux alors Infection = Non

(1 pt)

3. Modèle ID3 :

(2.5 pts)

- Si Age = Jeune alors
 - Si Résidu = Positif alors Infection = oui
 - Si Résidu = Négatif alors Infection = non
- Si Age = Vieux alors Infection = non
- Table de décision des modèles :

(0.5 pt)

Patient	ZeroR	OneR	ID3	Infection
01	non	oui	oui	oui
02	non	oui	non	non
03	non	non	non	non
04	non	non	non	non
05	non	oui	oui	oui
06	non	oui	non	non

Modèle Stacking par naïve bayes :

(0.5 pt)

Attribut	Infection	
	oui (2/6)	non (4/6)
ZeroR		
oui	0/6	0/6
non	0/6	6/6
OneR		
oui	2/6	2/6
non	0/6	2/6
ID3		
oui	2/6	0/6
non	0/6	4/6

En utilisant l'estimateur de Laplace :

(0.5 pt)

Attribut	Infection	
	oui (2/6)	non (4/6)
ZeroR		
oui	1/10	1/10
non	1/10	7/10
OneR		
oui	3/10	3/10
non	1/10	3/10
ID3		
oui	3/10	1/10
non	1/10	5/10

Exercice 3

1. – Attribut Age :
 - Si Age = Jeune alors Infection = $(65 + 40 + 20 + 80)/4 = 51,25$
 - Si Age = Vieux alors Infection = $(35 + 25)/2 = 30$
 - WSS = $(65 - 51,25)^2 + (30 - 35)^2 + (40 - 51,25)^2 + (20 - 51,25)^2 + (80 - 51,25)^2 + (30 - 25)^2 = 2175$
 - Attribut Résidus :
 - Si Résidus = Positif alors Infection = $(65 + 35 + 80 + 25)/4 = 51,25$
 - Si Résidus = Négatif alors Infection = $(40 + 20)/2 = 30$
 - WSS = $(65 - 51,25)^2 + (55 - 51,25)^2 + (80 - 51,25)^2 + (25 - 51,25)^2 + (40 - 30)^2 + (20 - 30)^2 = 2150$
 - Attribut Tension :
 - Si Tension = Élevée alors Infection = $(65 + 35 + 40 + 80)/4 = 55$
 - Si Tension = Normale alors Infection = $(20 + 25)/2 = 22,5$
 - WSS = $(65 - 55)^2 + (35 - 55)^2 + (40 - 55)^2 + (80 - 55)^2 + (20 - 22,5)^2 + (25 - 22,5)^2 = 1362,5$
 - Le min est l'attribut Tension
 - Si Tension = Normale alors
 - Si Age = Jeune alors Infection = 25
 - Si Age = Vieux alors Infection = 20
 - Si Tension = Élevée alors
 - Si Résidus = Positif alors
 - Si Age = Jeune alors Infection = 72,5
 - Si Age = Vieux alors Infection = 40
 - Si Résidus = Négatif alors Infection = 35
2. La valeur de la classe de l'exemple = 25

Exercice 4

1. Clustering Incrémental (0.75 pt)
 - Arrivé du premier exemple e_{01} : affecté au Cluster C1
 c_1 centre du Cluster1 = $\{25, +1, 15\}$
 - Arrivé de e_{02}
Distance(e_{01}, c_1) = $|25 - 30| + |1 - (-1)| + |15 - 17| = 9 < 25 \Rightarrow e_{01}$ affecté à C1
 $C1 = \{e_{01}, e_{02}\}$
 $c_1 = (\frac{25+30}{2}, \frac{+1-1}{2}, \frac{15+17}{2}) = (27,5, 0, 16)$
 - Arrivé de e_{03}
Distance(e_{03}, c_1) = $|65 - 27,5| + |1 - 0| + |16 - 16| = 38,5 > 25 \Rightarrow e_{03}$ affecté à un nouveau cluster C2 de centre $c_2 = (65, 1, 16)$
 $C1 = \{e_{01}, e_{02}\}, C2 = \{e_{03}\}$
 - Arrivé de e_{04}
Distance(e_{04}, c_1) = $|75 - 27,5| + |1 - 0| + |12 - 16| = 52,5$
Distance(e_{04}, c_2) = $|70 - 65| + |1 - 1| + |12 - 16| = 10 < 25 \Rightarrow e_{03}$ affecté à au cluster C2 de centre $c_2 = ((65 + 70)/2, (1 + 1)/2, (16 + 12)/2) = (67,5, 1, 14)$
 $C1 = \{e_{01}, e_{02}\}, C2 = \{e_{03}, e_{04}\}$
 - Arrivé de e_{05}
Distance(e_{05}, c_1) = $|38 - 27,5| + |1 - 0| + |15 - 16| = 3,5 < 25$
Distance(e_{05}, c_2) = $|38 - 67,5| + |1 - 1| + |15 - 14| = 31,5 > 25$
 $\Rightarrow e_{05}$ affecté à au cluster C1 de centre $c_1 = (\frac{27,5+38}{2}, \frac{0+1}{2}, \frac{16+15}{2}) = (32,75, 0,5, 15,5)$
 $C1 = \{e_{01}, e_{02}, e_{05}\}, C2 = \{e_{03}, e_{04}\}$

– Arrivé de e_{06}

$$\text{Distance}(e_{06}, c_1) = |42 - 32.75| + |-1 - 0.5| + |11 - 15.5| = 15.25 < 25$$

$$\text{Distance}(e_{06}, c_2) = |42 - 67.5| + |-1 - 1| + |11 - 14| = 30.5 > 25$$

$$\Rightarrow e_{06} \text{ affecté à au cluster C1 de centre } c_1 = \left(\frac{32.75+42}{2}, \frac{0.5-1}{2}, \frac{15.5+11}{2}\right) = (37.37, 0.25, 13.25)$$

$$C1 = \{e_{01}, e_{02}, e_{05}, e_{06}\}, C2 = \{e_{03}, e_{04}\}$$

2.

(0.75 pt)

Correspondance classe-cluster

Classe/Cluster	C1	C2
oui	2	0
non	2	2

Affectation des clusters : C1 classe oui , C2 classe non

Taux de clustering = $4/6 = 66.66 \%$