

Examen

Questions. (6 Points)

1. Un motif maximal est-il un motif ferme ? expliquer ?
2. Qu'elle est la différence entre la classification supervisée et non supervisée ?
3. Dans la classification SVM, que peut-on faire pour traiter des données non-linéairement séparable ?
4. Quel est le rôle de la fonction d'activation utilisée dans les réseaux de neurones ?
5. Qu'elle est la différence entre les deux algorithmes de clustering : k-means et DBSCAN ?

Exercice 1. (10 Points)

La Table 1 représente une base d'entraînement contenant des informations sur des employés d'une entreprise. Les données numériques (Age et Salary) ont été transformées en plages (intervalles), par exemple la plage [31-40] contient tout âge compris entre 31 et 40 ans.

Table 1. Base d'entraînement.

N	Dept.	Age	Salary	Class
1	sales	[31-40]	[46K-65K]	s
2	sales	[31-40]	[31K-45K]	j
3	systems	[21-30]	[46K-65K]	j
4	systems	[31-40]	[66K-70K]	s
5	systems	[41-45]	[66K-70K]	s
6	marketing	[31-40]	[46K-65K]	s
7	marketing	[31-40]	[31K-46K]	j
8	secretary	[41-45]	[31K-45K]	s
9	secretary	[21-30]	[31K-45K]	j

Table 2. Base de test.

N	Dept.	Age	Salary	Class
1	sales	[21-30]	[66K-70K]	s
2	systems	[41-45]	?	j
3	marketing	?	?	s
4	?	[21-30]	[31K-45K]	j

1. Quel est le nom de l'opération utilisée pour transformer les valeurs numériques en intervalles ? pourquoi est-elle utilisée ?
2. Donner le modèle d'apprentissage en utilisant l'algorithme Naïve de Bayes.
3. Donner la classe de chaque exemple dans la base de test présentée dans la Table 2 selon ce modèle.
4. Donner la précision (Accuracy) du modèle appliqué sur cette base de test.

Exercice 2. (4 Points)

La Table 3 représente les notes d'examen continu x et final y d'une matière donnée pour un ensemble d'étudiants.

1. On veut prédire la note d'examen y selon la note de contrôle continu x pour un étudiant qui ne figure pas dans la base présentée dans la Table 3. **Comment appelle-t-on cette opération ?**
2. Donner les valeurs des poids a et b utilisées dans le modèle de prédiction $y = a + bx$. Noter que $\bar{x} = 12$, $\bar{y} = 11,64$, $\sum_{i=1}^7 x_i y_i = 1010$, $\sum_{i=1}^7 x_i^2 = 1169$, et $\sum_{i=1}^7 x_i = 84$.
3. Prédire la note d'examen final y pour un étudiant qui a eu une note de 13 dans l'examen continu (x).

Table 3. Base d'entraînement.

N°	MidExam (x)	Exam (y)
1	14,5	17
2	8	8
3	16	10
4	15	4
5	18,5	18
6	6	15
7	6	9,5

Bon courage.

Examen

Questions. (7 pts)

1. Un motif maximal est-il un motif ferme ? expliquer ? **Oui. Puisque un motif ferme n'a aucun sur motif fréquent, et le motif max n'a pas de sur motif.**
2. Qu'elle est la différence entre la classification supervisée et non supervisée ? **Dans la classification supervisée les classes des exemples sont connues (étiquetées) contrairement à la classification non supervisée.**
3. Dans la classification SVM, que peut-on faire pour traiter des données non-linéairement séparable ? **On utilise l'astuce du noyau (Kernel).**
4. Quel est le rôle de la fonction d'activation utilisée dans les réseaux de neurones ? **Pour mettre une valeur dans une plage données pour qu'elle soit adaptée à l'étiquetage des exemples de la base d'exemples.**
5. Qu'elle est la différence entre les deux algorithmes de clustering : k-means et DBSCAN ? **K-means est global contrairement à DBSCAN, et il doit connaître le nombre de classes k.**

Exercice 1. (10 pts)

1. Quel est le nom de l'opération utilisée pour transformer les valeurs numériques en intervalles ? pourquoi est-elle utilisée ? **C'est la discrétisation. Elle est utilisée lorsque on doit manipuler des données formelles pour qu'on puisse avoir des intervalles comme dans les arbre de décision par exemple.**
2. Donner le modèle d'apprentissage en utilisant l'algorithme Naïve de Bayes. **Voir la table 1**
3. Donner la classe de chaque exemple dans la base de test présentée dans la Table 2 selon ce modèle. **Calculer la classe de chaque exemple :**
$$P(S|sales,A,F) = P(sales|s) * P(A|s) * P(F|s) * P(s) = 1/5 * 1/8 * 2/5 * 5/9 = 0.005$$
$$P(j|sales,A,F) = P(sales|j) * P(A|j) * P(F|j) * P(j) = 1/4 * 3/7 * 1/7 * 4/9 = 0.006$$

On remarque que $p(j|sales,A,F)$ est supérieur $p(s|sales,A,F)$ donc la classe de l'exemple 1 est j.

$$P(S|systems,C,?) = P(sales|systems) * P(A|C) * P(s) = 2/5 * 3/8 * 5/9 = 0.083$$
$$P(j|systems,C,?) = P(systems|j) * P(C|j) * P(j) = 1/4 * 1/7 * 4/9 = 0.016$$

On remarque que $p(s|sales,A,F)$ est supérieur $p(j|sales,A,F)$ donc la classe de

TABLE 1 – Table de probabilité (modèle) correspondant.

		Classe	
Attributs		$P(s) = 5/9$	$P(j) = 4/9$
Dept.	sales = 3	$P(\text{Dept.}=\text{sales} \text{classe}=s) = 1/5$	$P(\text{Dept.} \text{classe}=j) = 1/4$
	sys = 3		
	Marketing=2		
	Secretary=2		
Age	A=[21-30] = 2	$0/5 = 1+0/5+3 = 1/8$	$2/4 = 3/7$
	B=[31-40] = 5	$3/5 = 4/8$	$2/4 = 3/7$
	C=[41-45] = 2	$2/5 = 3/8$	$0/4 = 1/7$
Salary	D = [31k-45k] = 4	$1/5$	$3/4 = 4/7$
	E = [46k-65k] = 3	$2/5$	$1/4 = 2/7$
	F = [66k-70k] = 2	$2/5$	$0/4 = 1/7$

TABLE 2 – Résultats de prédiction.

N	Classe connue	classe obtenue
1	s	j
2	j	s
3	s	?
4	j	j

l'exemple 2 est s.

$$P(s|\text{marketing}, ?, ?) = P(\text{marketing}|s) * P(s) = 1/5 * 5/9 = 0.11$$

$$P(j|\text{marketing}, ?, ?) = P(\text{marketing}|j) * P(j) = 1/4 * 4/9 = 0.11$$

On remarque que $p(j|\text{sales}, A, F)$ est égale à $p(j|\text{sales}, A, F)$, donc cet exemple est inclassable.

$$P(s|?, A, D) = p(A|s) * P(D|s) * P(s) = 1/8 * 1/5 * 5/9 = 0.014$$

$$P(j|?, A, D) = p(A|j) * P(D|j) * P(j) = 3/7 * 4/7 * 4/9 = 0.10$$

On remarque que $p(j|\text{sales}, A, F)$ est supérieur à $p(s|\text{sales}, A, F)$, la classe de l'exemple 4 est j.

La table 2 résume les résultats de la classification des 4 exemples de la base de test.

- Donner la précision (Accuracy) du modèle appliqué sur cette base de test. **La précision** $= 1/3 = 0.33$

Exercice 2. (4 pts)

La Table 3 représente les notes d'examen continu x et final y d'une matière donnée pour un ensemble d'étudiants.

- On veut prédire la note d'examen y selon la note de contrôle continu x pour un étudiant qui ne figure pas dans la base présentée dans la Table 3.

— Comment appelle-t-on cette opération ?

Cette méthode est appelée la prédiction.

- Donner les valeurs des poids a et b utilisées dans le modèle de prédiction $y = a + bx$.

$$\begin{aligned} & \frac{\sum_{i=1}^7 (x_i y_i - x_i \bar{y})}{\sum_{i=1}^7 (x_i^2 - x_i \bar{x})} \\ &= \frac{\sum_{i=1}^7 x_i y_i - \bar{y} \sum_{i=1}^7 x_i}{\sum_{i=1}^7 x_i^2 - \bar{x} \sum_{i=1}^7 x_i} \end{aligned}$$

$$b = \frac{1010,5 - 11,64 \times 84}{1169,5 \times 84} = 0,2$$

$$a = \bar{x} - b\bar{y} = 12 - 0,2 \times 11,64 = 9,67.$$

3. Prédire la note d'examen final y pour un étudiant qui a eu une note de 13 dans l'examen continu (x). **Pour prédire la valeur de y correspondant à la valeur de $x=13$, on utilise la fonction de décision $y=a + bx = 9,67 + 0,2 \times 13 = 12,27$.**