Examen 2nde session (2h) - 28 juin 2023

Rappels : **Seul document autorisé : feuille A4 manuscrite, recto-verso.** Les calculatrices et autres appareils électroniques doivent être éteints et rangés. Le barème (sur 20) n'est donné qu'à titre indicatif.

Attention : soignez la présentation de votre copie ainsi que votre écriture !
Toute réponse illisible ou incompréhensible sera notée 0.

Exercice 1 *Questions de cours (3 points)*
On considère une base d'apprentissage $\mathcal{X}$ contenant n exemples décrits par p attributs.
Les questions suivantes sont indépendantes.

Q. 1. Qu'est-ce qui différencie l'algorithme des k-moyennes et l'algorithme des k-médoides ?

Q. 2. Quels sont les avantages et les désavantages de l'algorithme des k-moyennes vis-à-vis de l'algorithme de classification hiérarchique ?

Q. 3. Montrer que la distance de Manhattan est bien une mesure de distance.

Q. 4. Qu'est-ce que le sur-apprentissage ? Comment le mesure-t-on ? Comment l'éviter quand on utilise l'algorithme des k plus proches voisins ?

Exercice 2 *Apprentissage non-supervisé (7 points)*

On considère la base d'apprentissage de $[0, 10] \times [0, 10]$ contenant les 7 exemples suivants : $\mathcal{X} = \{(1, 2), (1, 4), (3, 4), (3, 5), (6, 2), (6, 5), (8, 3)\}$ (pour simplifier, on considère que cette base est déjà normalisée). Par la suite, le premier point est noté A, le 2e point est noté B, etc.

Q. 1. En détaillant les étapes, les calculs réalisés et les regroupements effectués, appliquer sur $\mathcal{X}$ l'algorithme de classification hiérarchique, version ascendante, en utilisant l'approche "centroid linkage" et la distance euclidienne.

Q. 2. Donner le dendrogramme obtenu par l'application de l'algorithme dans la question 1.

Q. 3. On rajoute l'exemple $(5, 5)$ dans $\mathcal{X}$, sans ré-appliquer l'algorithme donner, en les justifiant, les modifications apportées au dendrogramme précédent par l'ajout de cet exemple.

Q. 4. On décide d'appliquer l'algorithme des k-moyennes avec $k = 2$ sur la base $\mathcal{X}$ initiale (sans l'exemple $(5, 5)$). On choisit les exemples $A = (1, 4)$ et $E = (6, 2)$ pour initialiser les centres des clusters. Représenter graphiquement $\mathcal{X}$ et la frontière de séparation des 2 clusters induits par A et E.

Q. 5. Donner les coordonnées des 2 centres des deux clusters induits par A et E.

Q. 6. Donner l'inertie intra-cluster des clusters obtenus à la question précédente et en déduire l'inertie globale de la partition.

Q. 7. Au bout de combien d'étapes l'algorithme des k-moyennes converge-t-il pour cette initialisation par A et E ? Et quel est le résultat obtenu ?

Q. 8. L'algorithme des k-moyennes est sensible au choix des centres initiaux, quel choix d'exemples initiaux dans $\mathcal{X}$ entraînerait une convergence en un grand nombre d'étapes lors de l'application sur $\mathcal{X}$ avec $k = 2$?

Q. 9. A quoi servent les index de Dunn et de Xie-Beni ? Que mesurent-ils (chacun pris séparément) et qu'est ce qui les différencient ?

Exercice 3 *Perceptron (3 points)*

On considère un perceptron simple avec deux entrées notées x_1 et x_2 et une sortie y telle que :

$$y = \begin{cases} 1 \text{ si } w_1 x_1 + w_2 x_2 - w_0 > 0 \\ 0 \text{ sinon} \end{cases}$$

Q. 1. Trouver les poids pour que le perceptron calcule la fonction ET logique.

Q. 2. Même question avec la fonction OU logique.

Q. 3. Essayer de trouver des poids pour la fonction XOR et commenter.

Exercice 4 *Apprentissage supervisé (4 points)*

On considère un problème d'apprentissage supervisé pour lequel on possède une base d'apprentissage $\mathbf{X}$ de n exemples décrits par d variables (ou attributs). Chaque exemple $\mathbf{x}_i \in \mathbf{X}$ est associé à un label (ou classe) $y_i \in \mathbf{Y} = \{y_1, y_2\}$.

Q. 1. Donner l'expression de l'entropie de Shannon de la classe pour un attribut v possédant les 3 valeurs v_1, v_2, et v_3.

Q. 2. Expliquer comment obtenir le gain d'information de l'attribut v de la question précédente.

Q. 3. Pour la construction d'un arbre de décision, l'entropie de Shannon peut être remplacée par l'index de diversité de Gini $H_G(Y) = 1 - \sum_{y \in Y} p(y)^2$. Tracer les courbes de H_S (l'entropie de Shannon) et de H_G pour un cas à 2 classes. Comparer les résultats et conclure sur la différence des arbres obtenus en utilisant ces 2 mesures.

Q. 4. On considère la base suivante : $\mathbb{X} = \{((1,2),+1),((1,4),+1),((2,5),+1),((4,3),+1),((3,5),+1),$ $((2,8),+1),((3,2),-1),((4,4),-1),((5,5),-1),((4,7),-1),((6,2),-1),((5,8),-1)\}$. Représenter graphiquement la base de la question précédente et tracer la frontière de séparation des classes lors de l'application des k plus proches voisins pour $k = 3$.

Exercice 5 *Perceptron II (3 points)*

Soit le perceptron dont le vecteur de poids est $(w_0, w_1, w_2) = (2, 1, 1)$ et qui calcule le score :

$$y = signe(w_0 + w_1 x_1 + w_2 x_2).$$

Q. 1. Dessiner dans $\mathbb{R}^2$ la frontière de décision de ce perceptron.

Q. 2. Parmi les 4 modèles suivants, quel perceptron donne la même décision que le perceptron précédent ?

	$(w_0, w_1, w_2)^\top$
(I)	$(1, 0.5, 0.5)^\top$
(II)	$(200, 100, 100)^\top$
(III)	$(\sqrt{2}, \sqrt{1}, \sqrt{1})^\top$
(IV)	$(-2, -1, -1)^\top$

Q. 3. Soit les données ci-dessous où $\mathcal{P}$ correspond aux exemples positifs et $\mathcal{N}$ aux exemples négatifs :

$$(1,1)^T \in \mathcal{P}, \quad (1,0)^T \in \mathcal{N}, \quad (0,0)^T \in \mathcal{P}, \quad (0,1)^T \in \mathcal{N}$$

Soit le vecteur initial $w = (1, 0, 0)$. En prenant les exemples dans l'ordre, de manière cyclique, appliquer l'algorithme du perceptron (en considérant un pas d'apprentissage = 1). Comment peut-on voir qu'il ne sera pas possible d'apprendre un perceptron dans ce cas ?