# Examen 2nde session (2h) - 24 juin 2024

**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.

**Exercice 1** *Questions de cours (2pts)*

**Q. 1.** Soit $d$ une mesure telle que $d : X \times X \rightarrow [0, +\infty[$. Quelles sont les propriétés que doit vérifier $d$ pour être une mesure de distance ?

**Q. 2.** Montrer que la distance de Manhattan est bien une distance.

**Q. 3.** Rappeler la définition d'une partition $P = \{C_1, \ldots, C_K\}$ d'un ensemble $\mathcal{X}$ en $K$ sous-ensembles.

**Exercice 2** *Clustering (5pts)*

Dans cet exercice, on considère la base d'apprentissage $\mathcal{B}$ suivante :

| *exemple* | $e_0$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ | $e_7$ | $e_8$ |
|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | -0.5 | 0 | -1.5 | -1.4 | 1.1 | -0.9 | 1.3 | 0.6 | 0 |
| $X_2$ | -1 | -0.9 | -0.4 | -1.5 | -0.3 | -1.1 | 0 | -0.2 | 1 |

**Q. 1.** Avec la distance euclidienne[1] et l'approche "centroid linkage", appliquer en détaillant toutes les étapes et les calculs, l'algorithme de clustering hiérarchique, méthode par agglomération sur la base $\mathcal{B}$.

**Q. 2.** Construire le dendrogramme correspondant à cette base et proposer une coupure permettant d'obtenir une partition en 3 groupes.

**Q. 3.** Appliquer (en détaillant les étapes) l'algorithme des $k$-moyennes sur la base $\mathcal{B}$ en prenant $k = 3$ et les point $e_0$, $e_1$ et $e_5$ comme centres initiaux. Vous proposerez un critère de convergence adapté à ce problème.

**Q. 4.** Représenter graphiquement $\mathcal{B}$ et la frontière de séparation des clusters obtenus par l'application précédente des $k$-moyennes.

**Exercice 3** *Apprentissage supervisé (5pts)*

On considère la base d'apprentissage représentée dans la figure donnée en Annexe. Cette base contient 20 exemples, dont la description est le couple représenté par leurs coordonnées $(x, y)$, et la classe est soit *rond* (notée $R$) soit *carré* (notée $C$).

**Dans ce qui suit : utiliser l'Annexe (à rendre) pour les réponses graphiques demandées.**

**Q. 1.** En utilisant l'algorithme des $k$ plus proches voisins, avec $k = 3$ et la distance euclidienne, représenter graphiquement la frontière de séparation des classes et donner, en justifiant, la classe des 5 points suivants : le point $L$ de coordonnées $(3, 3)$, $M$ de coordonnées $(10, 3)$, $N$ de coordonnées $(7, 4)$, $P$ de coordonnées $(9, 5)$, et $Q$ de coordonnées $(4, 2)$. En cas d'égalité de distances, les points de classe $R$ seront considérés en priorité.

**Q. 2.** On considère les 4 seuils de coupure suivants :

— sur $X$ : coupure $c_1 : x \leq 3.5$ et coupure $c_2 : x \leq 8.5$

— sur $Y$, coupure $c_3 : y \leq 4.5$ et coupure $c_4 : y \leq 3.5$

En utilisant uniquement ces 4 seuils, construire un arbre de décision binaire en donnant le détail des calculs d'entropie réalisés. Donner une représentation graphique de cet arbre et représenter graphiquement la frontière de séparation entre les classes correspondante.

1. Pour l'application numérique, dans ce cas, comme cela ne joue pas de rôle, on ne fera pas le calcul de la racine carrée.

**Q. 3.** On considère que la classe $R$ correspond à la valeur $+1$ et la classe $C$ correspond à la valeur $-1$ et on décide d'utiliser l'algorithme du perceptron. Sans dérouler l'algorithme, mais en justifiant votre réponse, tracer la frontière de décision obtenue. Quelle est la particularité de cette frontière ?

**Q. 4.** En fait, les points $M$ et $P$ sont de la classe *carré* et $L$, $N$ et $Q$ sont de la classe *rond*. Donner la matrice de confusion pour chacun des modèles appris dans les questions précédentes ($k$-ppv, arbre de décision, et perceptron). Quel est le taux d'erreur de chacun de ces modèles ? Lequel est préférable ?

**Exercice 4** *Perceptron linéaire à seuil (4pts)*

Un classifieur à deux classes, $C1$ et $C2$, opère sur des objets de dimension $d = 2$ : $X = \begin{bmatrix} x_{1,1} & x_{1,2} \\ \cdots & \\ \mathbf{x}_i & \\ \cdots & \\ x_{N,1} & x_{N,2} \end{bmatrix}$,

avec $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^2$. Il utilise la fonction discriminante

$$g : \mathbf{x}_i \mapsto g(\mathbf{x}_i) = w_1 x_{i,1} + w_2 x_{i,2} - \theta,$$

avec $\theta$ qui est donné. L'exemple $\mathbf{x}_i$ est mis dans la classe $C_1$ si $g(\mathbf{x}_i) > 0$ et dans la classe $C_2$ si $g(\mathbf{x}_i) < 0$.

NB : dans cet exercice, on considère $\mathbf{x}_i$ comme un vecteur ligne et $w$ comme un vecteur colonne.

**Q. 1.** Quelle est l'équation de la frontière de décision ?

**Q. 2.** On peut mettre en bijection les points $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \in \mathbb{R}^2$ et $\mathbf{x}' = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}^T \in \mathbb{R}^3$. On construit un classifieur $g'$ de paramètres $w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$ pour traiter les points de $\mathbb{R}^3$ : $g'(\mathbf{x}') = \mathbf{x}'w$. Quelle valeur faut-il donner à $w$ pour que les deux classifeurs soient équivalents ?

**Q. 3.** On veut coder $+1$ les objets attribués à la classe $C_1$ et $-1$ les objets attribués à la classe $C_2$ ; quelle fonction $F$ faut-il utiliser pour que la composée $F \circ g$ réalise ce codage [2] ?

**Exercice 5** *Apprentissage du perceptron (4pts)*

N.B. Les notations sont les mêmes que dans l'exercice précédent.

On dispose d'une base de $N$ exemples (observations), $\{\mathbf{x}_i\}_{i=1,\ldots,N}$, dont les classes sont connues ; la classe de $\mathbf{x}_i$ est notée $d(\mathbf{x}_i)$. On utilise l'algorithme donné plus loin (une des variantes de l'algorithme du perceptron) pour apprendre automatiquement la valeur des paramètres, c'est à dire du vecteur $w$ : ($\epsilon$ est un nombre positif ; $\mathbf{x}^T$ est le transposé de $\mathbf{x}$ )

Le critère d'arrêt peut être, par exemple, qu'il n'y a pas eu d'erreur de classification pendant un certain nombre d'itérations successives.

**Q. 1.** Que signifie la condition $d(\mathbf{x}_i)\mathbf{x}_i w(t) \geq 0$ ? Expliquer le principe de l'algorithme.

**Q. 2.** Faites tourner l'algorithme du perceptron sur le problème du OU logique en itérant sur la base d'apprentissage constituée des 4 exemples distincts possibles, avec successivement pour valeur initiale w(0) :

1. $w(0) = (0; 0; 0)$ ;
2. $w(0) = (1; 1; 1)$ ;
3. $w(0) = (1; -1; 1)$.

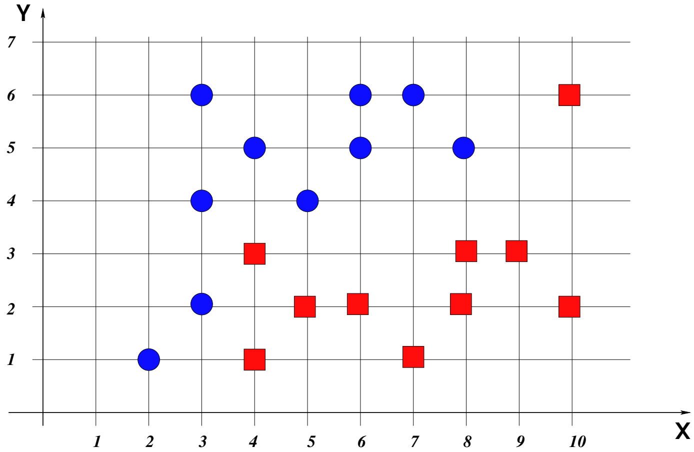
(Prendre $\epsilon = 1$) Représenter graphiquement l'évolution de la frontière de décision d'itération en itération.

2. Remarque : dans la terminologie des réseaux de neurones, $\mathbf{x}$ est une entrée, $g$ le potentiel de $\mathbf{x}$, $\theta$ le seuil, $F$ la fonction d'activation et $\{-1, +1\}$ les sorties ; le classifieur précédent est un perceptron linéaire à seuil, sans couche cachée, à une cellule de sortie.

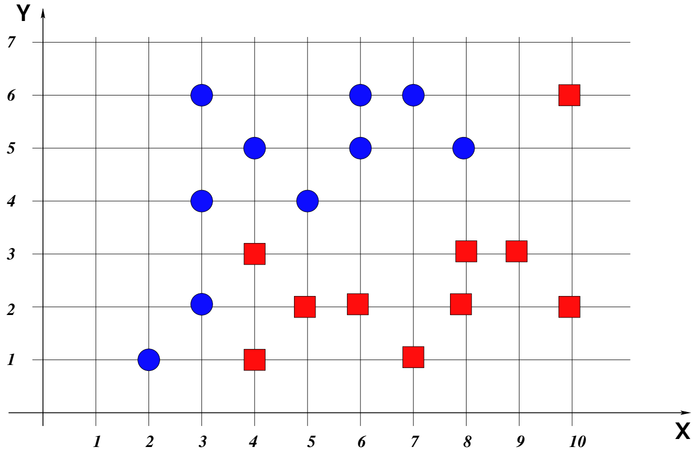
# NUMERO D'ANONYMAT :

## À rendre avec votre copie

### Exercice 3 - Question 1 : (k-ppv)




### Exercice 3 - Question 2 : (arbre)




### Exercice 3 - Question 3 : (perceptron)