Examen 2e session (2h) - 23 juin 2022

Rappels : **Aucun document n'est autorisé.** Les calculatrices et autres appareils électroniques doivent être éteints et rangés. Le barème (sur 32) n'est donné qu'à titre indicatif.

Exercice 1 *Questions de cours (6 points)*

Q. 1. Soit $\mathbf{x}_1$ et $\mathbf{x}_2$, deux exemples de $\mathbf{X} \in \mathbb{R}^{n\times d}$. Donner les expressions mathématiques de la distance euclidienne d_E, de la distance de Manhattan d_M et de la distance infinie d_∞.

Q. 2. On considère un repère orthonormé en 2 dimensions X_1 et X_2, et le point $\mathbf{x}_0$ de coordonnées $(0, 0)$. Représenter graphiquement les 3 ensembles de points $\mathbf{x}$ suivants : l'ensemble des $\mathbf{x}$ tels que $d_E(\mathbf{x}_0, \mathbf{x}) = 1$, l'ensemble des $\mathbf{x}$ tels que $d_M(\mathbf{x}_0, \mathbf{x}) = 1$, et l'ensemble des $\mathbf{x}$ tels que $d_\infty(\mathbf{x}_0, \mathbf{x}) = 1$.

Q. 3. On considère la représentation de 3 groupes d'exemples : le groupe A (les ronds), le groupe B (les plus) et le groupe C (les croix), représentés dans la figure 1. En justifiant vos réponses :

1. donner les 2 groupes les plus proches parmi A, B et C en utilisant l'approche par complete linkage.
2. donner les 2 groupes les plus proches parmi A, B et C en utilisant l'approche par simple linkage.

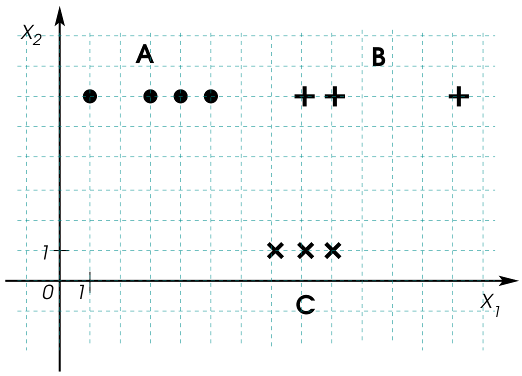

Figure 1 – Groupes de points

Exercice 2 *Expérimentations et évaluation (6 points)*

Soit $\mathbf{X} \in \mathbb{R}^{n\times d}$ un ensemble d'exemples et $\mathbf{Y} \in \{-1, +1\}^n$ l'ensemble des classes correspondant à chaque exemple de $\mathbf{X}$. Soit f_1 et f_2, deux modèles d'apprentissage supervisés (perceptron, arbre de décision, knn ou autre, on ne précise pas lesquels) défini respectivement par m_1 et m_2 hyper-paramètres et tels que $f_i : \mathbf{X} \to \{-1, +1\}$ pour $i = 1, 2$.

Q. 1. Proposer un processus complet pour la construction, le test et l'évaluation de f_1 et de f_2 à l'aide de $\mathbf{X}$ qui doit permettre de décider lequel des 2 modèles est à privilégier, en précisant les différents angles selon lesquels on peut en privilégier un plutôt qu'un autre.

Q. 2. Que mesure $S(f_1, f_2) = \frac{1}{n} \sum_{x\in\mathbf{X}} f_1(x) \cdot f_2(x)$? Expliquer et donner des cas remarquables.

Exercice 3 *Arbres de décision (5 points)*

On considère un ensemble d'exemples $\mathbf{X}$ d'un espace à 2 dimensions $X_1 = [0, 10]$ et $X_2 =]-100, 100[$. Chaque exemple est associé à un label de $\mathcal{Y} = \{-1, 1\}$. La frontière réelle $\mathcal{F}$ de séparation des classes est $x_2 = (x_1 - 5)^2 + 3$.

Q. 1. Donner graphiquement une base d'apprentissage $\mathcal{B}_1$ composée de 10 points de $X_1 \times X_2$ à partir de laquelle vous représenterez une partition de $\mathbf{X}$ caractéristique d'un arbre de décision et estimant $\mathcal{F}$.

Q. 2. Donner l'arbre de décision correspondant à la partition de la question précédente et évaluer sa performance sur un ensemble $\mathcal{B}_2$ de 10 points choisi intelligemment et de façon uniforme dans l'espace.

Q. 3. Quels sont les avantages et les inconvénients d'utiliser l'algorithme d'apprentissage par arbre de décision pour estimer $\mathcal{F}$?

Q. 4. Proposer une approche permettant d'appliquer l'algorithme d'apprentissage par arbres de décision pour estimer $\mathcal{F}$ de façon plus efficace.

Exercice 4 *Lenteurs et optimisations des k-ppv (2 points)*

Q. 1. Discuter la complexité de l'algorithme des k-ppv et identifier l'opération la plus coûteuse en temps.

Q. 2. Proposer plusieurs solutions pour réduire la complexité de l'algorithme

Exercice 5 *perceptron dual (7 points)*

On s'intéresse à des classifieurs linéaires pour discriminer deux classes. On note $D = \{x_i, y_i\}_{i=1,\ldots,N}$ un ensemble d'apprentissage avec $y_i = 1$ si $x_i \in C_1$ et $y_i = -1$ si $x_i \in C_2$. $\mathbf{w}$ est le vecteur de poids du classifieur linéaire.

1. Rappeler l'algorithme du perceptron
2. On suppose que l'algorithme est initialisé avec $\mathbf{w}_0 = 0$.
   (a) Montrer qu'à une étape de l'algorithme, il existe des coefficients α_i tels que $\mathbf{w}_t = \sum_{i=1}^{N} \alpha_i y_i x_i$. Pour cela, commencer par calculer $\mathbf{w}_1$, puis $\mathbf{w}_2$ et procéder par récurrence.
   (b) Exprimer en fonction des α_i la condition qui indique que le perceptron fait une erreur sur x_i.
   (c) Reformuler l'algorithme du perceptron avec les seuls α_i comme paramètres.
   (d) Exprimer la fonction de décision en fonction des α_i.

Exercice 6 *perceptron (6 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 ?