

## Examen 1ère session (2h) - 11 mai 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.

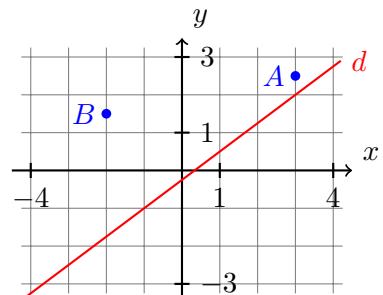
### Exercice 1 Perceptron (4 points)

On se place dans un repère orthonormé en 2 dimensions  $\mathbb{R} \times \mathbb{R}$  comme indiqué dans la figure ci-contre.

**Q. 1.** Donner un vecteur directeur de la droite  $d$  et en déduire l'équation de  $d$ .

**Q. 2.** Donner une méthode pour montrer que les points  $A$  et  $B$  sont du même côté de la droite  $d$  et l'appliquer en détaillant les calculs.

**Q. 3.** On considère que  $d$  représente un perceptron à l'étape  $t$ ,  $A$  est de label +1 et  $B$  est de label -1. En prenant un pas de mise à jour de  $\epsilon = 0.8$ , quels seront les paramètres du modèle à l'itération suivante ? Détailler les calculs obtenus et expliquer ce qui se produit si les points ne sont pas examinés dans le même ordre.



### Exercice 2 Perceptron et python (3 points)

On utilise les notations usuelles du cours. On considère l'algorithme du perceptron avec biais. On rappelle qu'une mise à jour est effectuée dans le cas où  $f(\mathbf{x}_i)y_i < 1$ . Cette mise à jour tient alors compte de l'écart entre la prédiction et le vrai label de  $\mathbf{x}_i$ .

**Q. 1.** Donner le code Python de la méthode `train` qui prend en argument  $X$ ,  $Y$  et  $\epsilon$ .

Remarque : il est demandé un code compact et efficace.

**Q. 2.** Quelles sont les erreurs qui sont le plus pénalisées dans cette variante de l'algorithme ?

### Exercice 3 Clustering (6 points)

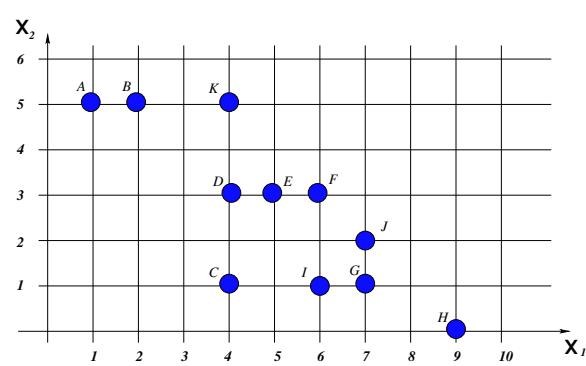
Dans cet exercice, on utilise la distance euclidienne.

**Q. 1.** Soit un ensemble d'exemples  $\mathbf{X} \in \mathbb{R}^{n \times d}$  et soit  $\mathcal{P} = \{\mathcal{C}_1, \dots, \mathcal{C}_K\}$  avec pour tout  $k = 1, \dots, K$ ,  $\mathcal{C}_k$  un sous-ensemble de  $\mathbf{X}$ . Donner les 3 propriétés que doivent vérifier les  $\mathcal{C}_k$  pour que  $\mathcal{P}$  soit une bonne partition de  $\mathbf{X}$  en  $K$  clusters.

**Q. 2.** On considère l'ensemble  $\mathbf{X} = \{A, B, \dots, H\}$  affiché dans la figure ci-contre. Pour simplifier, on considère que ces données n'ont pas besoin d'être normalisées.

a) En appliquant l'algorithme de clustering hiérarchique utilisant l'approche *simple linkage*, construire le dendrogramme correspondant à ces données en donnant les groupes construits à chaque étape de l'algorithme.

b) De même, construire le dendrogramme correspondant à l'approche *complete linkage* en donnant les groupes construits à chaque étape.

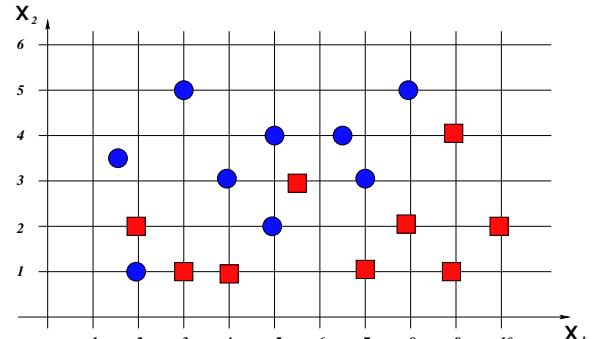


**Q. 3.** Appliquer l'algorithme des  $k$ -moyennes sur les données avec  $k = 3$  et en prenant les exemples  $A$ ,  $C$  et  $H$  comme centres initiaux. Vous donnerez les clusters obtenus à chaque itération de l'algorithme.

**Exercice 4** Apprentissage supervisé (5 points)

Dans cet exercice, on utilise la distance euclidienne.

**Q. 1.** On considère l'ensemble  $\mathbf{X}$  dont les données sont affichées dans la figure ci-contre. Dans cette figure, il y a 2 classes : les exemples de classe +1 sont les ronds, les exemples de classe -1 sont les carrés.



a) Représenter sur un schéma clair et soigné la frontière de séparation des classes pour un classifieur par 1-plus proche voisin.

b) Représenter sur un autre schéma (clair et soigné) la frontière de séparation des classes pour un classifieur par 3-plus proches voisins.

**Q. 2.** On souhaite utiliser l'algorithme de construction d'arbres de décision. Pour discréteriser les données, on a le choix entre une coupure sur  $X_1$  à la valeur seuil 6, et une coupure sur  $X_2$  à la valeur seuil 2.5. En détaillant les calculs utilisés, donner la meilleure coupure à prendre en terme de gain d'information.

**Exercice 5** Évaluations (3 points)

**Q. 1.** Citer 3 approches possibles pour évaluer un algorithme d'apprentissage supervisé sur un jeu de données fourni. Quels sont les avantages et les inconvénients de chacune de ces approches ? Quand sont-elles utilisables ou inutilisables ? Que permet chacune d'elle d'évaluer ?

**Annexe**

n / d	dénominateur (d)									
	1	2	3	4	5	6	7	8	9	10
1	1	0,50	0,33	0,25	0,20	0,17	0,14	0,13	0,11	0,10
2		1	0,67	0,50	0,40	0,33	0,29	0,25	0,22	0,20
3			1	0,75	0,60	0,50	0,43	0,38	0,33	0,30
4				1	0,80	0,67	0,57	0,50	0,44	0,40
5					1	0,83	0,71	0,63	0,56	0,50
6						1	0,86	0,75	0,67	0,60
7							1	0,88	0,78	0,70
8								1	0,89	0,80
9									1	0,90
10										1

numérateur (n)	-(n/d) * log(n/d)	dénominateur (d)									
		1	2	3	4	5	6	7	8	9	10
1	0	0,50	0,53	0,50	0,46	0,43	0,40	0,38	0,35	0,33	
2		0	0,39	0,50	0,53	0,53	0,52	0,50	0,48	0,46	
3			0	0,31	0,44	0,50	0,52	0,53	0,53	0,52	
4				0	0,26	0,39	0,46	0,50	0,52	0,53	
5					0	0,22	0,35	0,42	0,47	0,50	
6						0	0,19	0,31	0,39	0,44	
7							0	0,50	0,28	0,36	
8								0	0,15	0,26	
9									0	0,14	
10										0	

FIGURE 1 – Calculs d'entropies