Examen 2ème session (2h) - 24 juin 2019

Rappels : **Tous documents autorisés.** 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 (4 points)*

Q. 1. Pour les méthodes d'ensembles, qu'est-ce qui différencie une valisation croisée de l'approche *out of the bag*? A quoi servent ces 2 méthodes et laquelle est à privilégier (et dans quels cas) ?

Q. 2. Qu'est-ce qui différencie un problème de classification et un problème de régression. Illustrer en donnant un exemple de chaque.

Q. 3. Après avoir construit une forêt de 100 arbres de décisions, on voudrait utiliser l'algorithme des k-moyennes afin de faire ressortir des groupes d'arbres homogènes dans cette forêt. Quelle est la difficulté principale rencontrée ici ? Expliquer, en étant le plus précis possible et en donnant un maximum de détails, comment on pourrait appliquer cet algorithme dans ce cas.

Exercice 2 *Classification supervisée (9 points)*

Soit $\mathbb{X}$ une base d'apprentissage contenant n exemples définis par p dimensions et un label associé (les labels sont pris dans $\mathbb{Y} = \{-1, +1\}$). Dans cet exercice, on utilise la distance euclidienne.

Q. 1. On condidère que les p dimensions sont toutes numériques (ie. $\mathbb{R}$). Soit x_1 et x_2 deux exemples de $\mathbb{X}$, donner l'expression analytique de $d_E(x_1, x_2)$.

Q. 2. On décide d'appliquer l'algorithme des k plus proches voisins (kppv) en utilisant $\mathbb{X}$ pour classer un nouvel exemple x, combien de calculs de distance sont nécessaires au plus ?

Q. 3. On considère la base suivante : $\mathbb{X} = \{((1,2),+1), ((1,4),+1), ((2,5),+1), ((4,3),+1), ((3,5),+1), ((2,8),+1), ((3,2),-1), ((4,4),-1), ((5,5),-1), ((4,7),-1), ((6,2),-1), ((5,8),-1)\}$. En appliquant l'algorithme des kppv avec $k = 1$, et en détaillant les calculs réalisés, donner la classr de l'exemple $(3, 6)$.

Q. 4. Représenter graphiquement la base de la question précédente et tracer la frontière de séparation des classes lors de l'application des kppv pour $k = 1$.

Q. 5. Même question mais pour $k = 3$. En déduire la classe de l'exemple $(3, 6)$ dans ce cas.

Q. 6. Toujours en utilisant la base $\mathbb{X}$ de la question 3, mais en éliminant les exemples $(4, 4)$ et $(4, 3)$, construire un arbre de décision. Vous détaillerez les étapes et les calculs de réalisés.
N.B. : Une table de valeurs d'entropie est fournie en annexe.

Q. 7. Représenter graphiquement la base de la question précédente et tracer la frontière de séparation des classes fournie par l'arbre de décision construit.

Annexe

n / d		dénominateur (d)									
		1	2	3	4	5	6	7	8	9	10
numérateur (n)	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

-(n/d) * log(n/d)		dénominateur (d)									
		1	2	3	4	5	6	7	8	9	10
numérateur (n)	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 – Table de valeurs d'entropie