Éléments de théorie de l'information

1 Théorie de l'information

Exercice 1 (*).

1. Soit $X \sim \mathscr{U}(E)$, où E est un ensemble de cardinal $N \geqslant 1$. Montrer que $H(X) = \log_2(N)$.
2. Calculer l'entropie des variables aléatoires suivantes et interpréter les résultats obtenus.
   (a) On lance trois dés équilibrés à 4 faces et on note X_1 la v.a. correspondant à l'issue obtenue.
   (b) On lance deux dés équilibrés à 6 faces et on note X_2 la v.a. correspondant à l'issue obtenue.
   (c) On lance un équilibré à 12 faces et on note X_3 la v.a. correspondant à l'issue obtenue.

Exercice 2 (*). Geneviève se rend régulièrement à l'hippodrome, où se déroule une course entre huit chevaux. À la fin de la course, les gains de Geneviève dépendent de la performance du cheval sur lequel elle a misé : selon qu'il arrive en position $1, 2, 3, \dots, 8$, ses gains sont respectivement de 100, 50, 50, 0, -10, -50, -50, -100 euros. On supposera toujours que tous les chevaux sont aussi bons les uns que les autres.

1. On note X la variable aléatoire correspondant aux gains de Geneviève. Calculer son espérance et son entropie.
2. Lassée de perdre de l'argent sur le long terme, Geneviève décide de monétiser son activité en mettant en ligne des vidéos de réaction aux résultats des courses. Elle passe un contrat avec un network : ses gains sont désormais sécurisés et correspondent à la valeur absolue de ses gains initiaux (car le public l'apprécie tout autant quand elle s'énerve après avoir perdu). On note $Y = |X|$ la variable aléatoire correspondant aux nouveaux gains de Geneviève. Calculer la loi de Y, son espérance et son entropie.

Exercice 3 (**). Parmi les deux modèles de classification ci-dessous qui cherchent à classifier une image en chien ou en chat, lequel est le meilleur du point de vue de la fonction de coût associée à l'entropie croisée ?

	Image 1 : chien	Image 2 : chat
Modèle 1	100% chien, 0% chat	90% chien, 10% chat
Modèle 2	90% chien, 10% chat	80% chien, 20% chat

Exercice 4 (***). Calculer la divergence KL de deux lois normales unidimensionnelles. Pour quelles valeurs de leurs paramètres cette divergence KL s'annule-t-elle ? Commenter.

2 Classification binaire

Exercice 5 (**). On reprend l'exemple de classification binaire présenté en cours : le 15 avril 1912, lors de son voyage inaugural, le RMS Titanic coula après une collision avec un iceberg. On souhaite établir un lien entre la survie et l'âge des passagers voyageant en deuxième classe.

i	1	2	3	...	210
Âge x_i	30	28	18	...	18
Survie y_i	0	1	0	...	0

Pour tout $i \in \{1, \dots, N = 210\}$, on note Y_i la v.a. donnant la prédiction de la classe de x_i. Afin d'effectuer la classification, on propose un modèle logistique. On suppose ainsi que $\pi_i = P_\theta(Y_i = 1 | x_i) = \dfrac{1}{1 + e^{-(ax_i + b)}} = F(a, b, x_i)$, où $\theta = (a, b)$ est le couple de paramètres à déterminer.

1. Donner l'expression de la fonction de coût associée à l'entropie croisée, puis montrer qu'à un coefficient près, le calcul de son gradient donne

$$\nabla CE(a,b) = \left(\sum_{i=1}^{N} x_i\,(\pi_i - y_i),\, \sum_{i=1}^{N} (\pi_i - y_i) \right).$$

2. Utiliser un algorithme de descente de gradient pour déterminer les valeurs des paramètres a et b qui minimisent la fonction de coût.

Exercice 6 (***). Comparer les résultats obtenus à l'exercice précédent avec ceux donnés par la fonction de coût associée à l'erreur quadratique totale.