

## Apprentissage Statistique

### Exercice 1.

On étudie un ensemble de 10 observations,  $(x_i, y_i)_{1 \leq i \leq 10}$ , avec  $x_i \in \mathcal{X}$  et  $y_i \in \{-1, 1\}$ . Grâce à un algorithme d'apprentissage automatique, on construit deux modèles,  $g_1$  et  $g_2$ . Le tableau suivant donne les valeurs de  $g_1(x_i)$ ,  $g_2(x_i)$  et  $y_i$  pour tout  $i$  :

$x_i$	$g_1(x_i)$	$g_2(x_i)$	$y_i$
$x_1$	1	1	1
$x_2$	1	-1	1
$x_3$	-1	1	1
$x_4$	1	1	1
$x_5$	1	-1	1
$x_6$	1	-1	-1
$x_7$	-1	-1	-1
$x_8$	-1	-1	-1
$x_9$	-1	1	-1
$x_{10}$	1	-1	-1

**Question 1** Calculer les matrices de confusion des deux modèles.

**Question 2** On choisit la fonction de perte  $l_1$  définie par :

$l_1(v, p)$	$p = -1$	$p = 1$
$v = -1$	0	2
$v = 1$	1	0

où  $p$  désigne la valeur prédite et  $v$  la vraie valeur. Déterminer le meilleur modèle (entre  $g_1$  et  $g_2$ ) au sens du risque empirique construit à partir de  $l_1$ .

**Question 3** Quel modèle choisir au sens de la fonction de perte  $l_0(v, p) = \mathbb{1}_{p \neq v}$  ?

**Exercice 2.** On suppose que nous avons un ensemble de données a décrit les trois variables suivantes: Cheveux = B, D, où B=blond, D=dark. Hauteur = T, S, où T=tall, S=short. Pays = G, P, où G=Greenland, P=Poland. On vous donne l'ensemble de données d'entraînement suivant (Cheveux, Hauteur, Pays):

$$\left( \begin{array}{ccc} (B, T, G) & (B, T, G) & (B, T, P) \\ (D, T, G) & (D, T, G) & (B, T, P) \\ (D, T, G) & (D, T, G) & (B, T, P) \\ (D, T, G) & (D, T, G) & (D, T, P) \\ (B, T, G) & (B, T, G) & (D, T, P) \\ (B, S, G) & (B, S, G) & (D, S, P) \\ (B, S, G) & (B, S, G) & (B, S, P) \\ (D, S, G) & (D, S, G) & (D, S, P) \end{array} \right)$$

*Nous voulons répondre à la question suivante : si vous observez un nouvel individu grand ayant cheveux blonds, quel est son pays d'origine le plus probable?*

- 1. Donner la solution maximum a posteriori (MAP) à la question, en utilisant le cadre naïf de Bayes. Détaillez votre travail.*
- 2. Donner la solution maximum a posteriori (MAP) à la question, sans utiliser le cadre naïf de Bayes. Détaillez votre travail.*
- 3. Donner la solution de maximum de vraisemblance (MLE) à la question, en utilisant le cadre naïf de Bayes. Détaillez votre travail.*
- 4. Donner la solution de maximum de vraisemblance (MLE) à la question, sans utiliser le cadre naïf de Bayes. Détaillez votre travail.*