Perte, risque et erreur en apprentissage automatique

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 le nombre d'erreurs de classement réalisées par chaque modèle.

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. On rappelle que le risque empirique d'un modèle g pour la perte l_1 est donné par

$$\widehat{L}_1(g) = \frac{1}{10}\sum_{i=1}^{10} l_1(y_i,g(x_i)).$$

Déterminer le meilleur modèle (entre g_1 et g_2) au sens de $\widehat{L}_1$.

Question 3 On appelle matrice de confusion d'un modèle la matrice carrée dont les termes donnent les différents taux d'erreur de classement réalisés par le modèle. Chaque ligne correspond à une vraie valeur de y et chaque colonne à une valeur prédite. À l'intersection d'une ligne et d'une colonne, on indique le pourcentage des objets de la vraie valeur en ligne pour lesquels le modèle prédit la valeur en colonne. Dans le cas présent, il faut donc remplir une matrice de la forme suivante

g	$p=-1$	$p=1$
$v=-1$	pourcentage des objets $y=-1$ pour lesquels g donne -1	...
$v=1$	...	...

Calculer la matrice de confusion de g_2.

Question 4 Quel lien simple peut on faire entre $\widehat{L}_1(g)$ d'une part et la matrice représentant l_1 et la matrice de confusion de g d'autre part ?

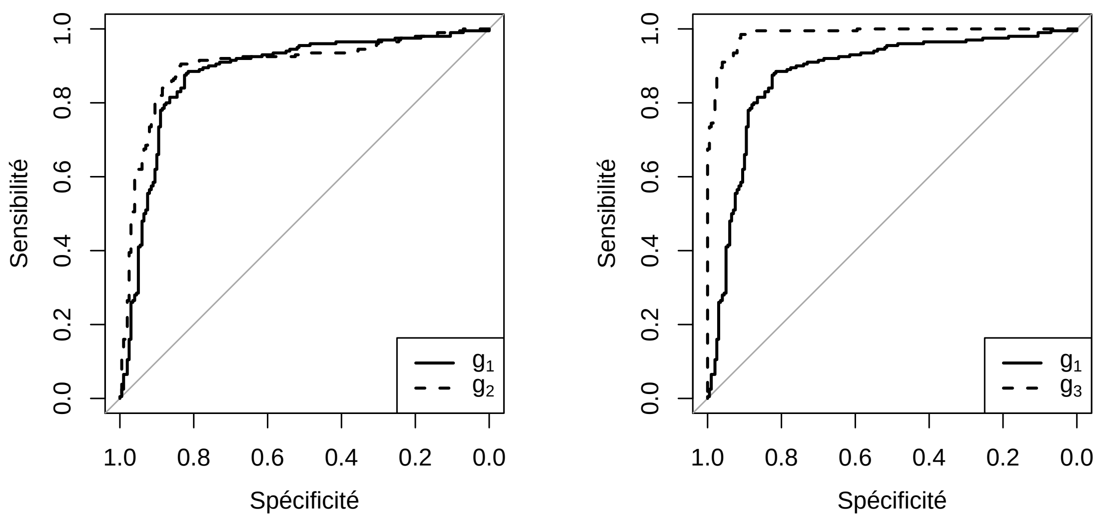
Exercice 2

On étudie un ensemble d'observations $(x_i,y_i)_{1\leq i\leq N}$, avec $x_i \in \mathcal{X}$ et $y_i \in \{-1,1\}$. Par une méthode d'apprentissage automatique, on construit deux modèles g_1 et g_2, des fonctions de $\mathcal{X}$ dans $\mathbb{R}$. On cherche à déterminer λ_1 et λ_2 tels que les modèles f_1 et f_2 définis ci-dessous soient « bons » au sens d'une fonction de perte l :

$$f_k(x) = \begin{cases} -1 & \text{si } g_k(x) \leq \lambda_k, \\ 1 & \text{si } g_k(x) > \lambda_k. \end{cases}$$

Question 1 On suppose fixée une fonction de perte l. Donner λ_k sous forme de la solution d'un problème d'optimisation (on ne peut pas expliciter λ_k, on demande simplement d'écrire le problème à résoudre).

Pour analyser les modèles g_k sans fixer de valeur à λ_k, on étudie leur courbe ROC (*Receiver Operating Characteristic*) : pour chaque $\lambda_k \in \mathcal{R}$, on calcule la *spécificité* et la *sensibilité* de f_k, et on trace la courbe obtenue ainsi. La spécificité est le taux de vrais négatifs (ici les x_i tels que $y_i = -1$ pour lesquels $g_k(x_i) \leq \lambda_k$) alors que la sensibilité est le taux de vrais positifs (ici les x_i tels que $y_i = 1$ pour lesquels $g_k(x_i) > \lambda_k$). La figure 1a représente les courbes ROC de g_1 et g_2.



(a) Courbes ROC de g_1 et g_2.

(b) Courbes ROC de g_1 et g_3.

Question 2 On considère la fonction de perte $l_{\mathbb{1}}$ donnée par $l_{\mathbb{1}}(v,p) = \mathbb{1}_{v=p}$. On suppose que

$$|\{i|y_i = -1\}| = |\{i|y_i = 1\}|.$$

Quel point de la courbe ROC de g_1 correspond au meilleur λ_1 pour la fonction de perte $l_{\mathbb{1}}$? Que peut-on dire si l'hypothèse ci-dessus n'est pas vérifiée ?

Question 3 Peut-on dire que g_1 est meilleur que g_2 (ou moins bon) *de façon générale* en utilisant la figure 1a ? Même question entre g_1 et le nouveau modèle g_3, en s'appuyant sur la figure 1b.