# Modèles optimaux

## Exercice 1

On suppose connue la loi de $Z = (X,Y)$ et on cherche le modèle optimal de $Y$ comme fonction de $X$, avec $Y$ à valeurs dans $\{0,1\}$. Le modèle doit être optimal au sens de la fonction de perte $l_1$ donnée par

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

où $p$ désigne la valeur prédite et $v$ la vraie valeur.

**Question 1** Montrez que pour tout modèle $g$, on a

$$\mathbb{E}(l_1(Y,g(X))|X = x) = 2\mathbb{P}(Y = 0|X = x)\mathbb{I}_{g(x)=1} + \mathbb{P}(Y = 1|X = x)\mathbb{I}_{g(x)=0}.$$

**Question 2** En déduire que pour tout couple de modèles $g$ et $g'$, on a

$$\mathbb{E}(l_1(Y,g(X))|X = x) - \mathbb{E}(l_1(Y,g'(X))|X = x) = (\mathbb{I}_{g(x)=1} - \mathbb{I}_{g'(x)=1})(2\mathbb{P}(Y = 0|X = x) - \mathbb{P}(Y = 1|X = x)).$$

**Question 3** En déduire le modèle optimal.

## Exercice 2

On étudie des données distribuées selon le modèle $Z = (X,Y)$ suivant :

— $Y$ est une variable aléatoire à valeurs dans $\{0, 1\}$ avec $\mathbb{P}(Y = 0) = \frac{1}{3}$ ;

— $X$ est une variable aléatoire à valeurs dans $\{a, b, c\}$ dont la loi conditionnelle est donnée par :

| $x$ | $a$ | $b$ | $c$ |
|---|---|---|---|
| $\mathbb{P}(X = x\|Y = 0)$ | $0$ | $\frac{1}{3}$ | $\frac{2}{3}$ |
| $\mathbb{P}(X = x\|Y = 1)$ | $\frac{1}{3}$ | $\frac{1}{2}$ | $\frac{1}{6}$ |

**Question 1** Calculez la loi conditionnelle de $Y$ sachant $X$ dans ce modèle.

**Question 2** Déterminez le modèle optimal de $Y$ sous la forme d'une fonction de $X$ au sens de la fonction de perte $l_0(v,p) = \mathbb{I}_{p \neq v}$. Quel est le risque de ce modèle ?

**Question 3** Mêmes questions avec la fonction de perte $l_1$ de l'exercice précédent.

## Exercice 3

On étudie des données distribuées selon le modèle $Z = (X,Y)$ suivant :

— $Y$ est une variable aléatoire à valeurs dans $\{0, 1\}$ avec $\mathbb{P}(Y = 0) = \frac{1}{2}$ ;

— $X$ est une variable aléatoire à valeurs dans $\mathbb{R}^2$ dont la densité conditionnelle sachant $Y = y$ est donnée par la loi normale sur $\mathbb{R}^2$ de moyenne $\mu_y$ et de matrice de covariance $\Sigma_y$.

On rappelle que la densité de $\mathcal{N}(\mu,\Sigma)$ est donnée par

$$f(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

**Question 1** On suppose que $\Sigma_1 = \Sigma_2$. Déterminez alors le modèle optimal de $Y$ sous la forme d'une fonction de $X$ au sens de la fonction de perte $l_0(v,p) = \mathbb{I}_{p \neq v}$.

**Question 2** Même question sans l'hypothèse d'égalité des covariances.