Introduction to Statistical Learning

Exercise sheet n°1

Exercise 1 - Consider the binary classification model where the random pair (X, Y) has distribution P over $\mathbb{R} \times \{0, 1\}$ and :

$$\begin{aligned} \mathcal{L}(X \mid Y = 0) &= \mathcal{U}([0, \theta]) \\ \mathcal{L}(X \mid Y = 1) &= \mathcal{U}([0, 1]) \\ p &= \mathbb{P}(Y = 1) \end{aligned}$$

with $p, \theta \in (0, 1)$ fixed. Compute the posterior probability $\eta(x) = \mathbb{P}(Y = 1 \mid X = x)$, for any $x \in \mathbb{R}$, as a function of p, θ. What if $\theta = 1/2$?

Exercise 2 - Consider the binary classification model where the random pair (X, Y) has distribution P over $\mathbb{R}_+ \times \{0, 1\}$ and :

— the marginal distribution of X over $\mathbb{R}_+$ is denoted P_X

— the conditional distribution of Y given $X = x$ is a Bernoulli distribution with parameter $\eta(x) = \dfrac{x}{x + \theta}$, for any $x \in \mathbb{R}_+$, and for fixed $\theta > 0$.

Find the Bayes classifier for this model (i.e. the minimizer of $L(g) = \mathbb{P}(Y \neq g(X))$ over all measurable classifiers $g : \mathbb{R}_+ \to \{0, 1\}$. Give the expression of the Bayes error $L^* = L(g^*)$ in the case where $P_X = \mathcal{U}([0, \alpha\theta])$ with $\alpha > 1$. What is the value of α that maximizes L^* ?

Exercise 3 - Let $X = (T, U, V)^T$ où T, U, V IID real-valued random variables with exponential distribution $\mathcal{E}(1)$. Define $Y = \mathbb{I}\{T + U + V < \theta\}$ with fixed $\theta > 0$.

1. Find the Bayes classifier $g^*(T, U)$ when V is not observed. Give the expression of the classification error of g^* (also called Bayes error). Compute it for $\theta = 9$.
2. Now assume that only T is observed, and address the same questions as above.
3. Propose a classifier for X when none of T, U, V are observed. What is its classification error ?

Exercise 4 - Consider (X, Y) a random pair that models classification data with labels in $\{0, 1\}$. Define the following classification error

$$L_\omega(g) = \mathbb{E}\big(2\omega(Y) \cdot \mathbb{I}\{Y \neq g(X)\}\big)$$

where $\omega(0) + \omega(1) = 1$.

1. Find the optimal elements (minimizer, error) for this risk.
2. Justify the interest of such an L_ω in practice ?
3. Now consider the unit square in $\mathbb{R}^2$.
   (a) Plot the curves defined by $g \mapsto (\mathbb{P}\{g(X) = 1 \mid Y = 0\}, \mathbb{P}\{g(X) = 1 \mid Y = 1\})$ when g varies such that $L_\omega(g) = C$ with C fixed, for different values of C.
   (b) Same question but assuming now that $\mathbb{P}\{g(X) = 1\} = C$ with C fixed.

Exercise 5 - Consider (X, Y) a random pair that models classification data with labels in $\{0, 1\}$. We fix $c > 0$ and we consider classifiers with reject option $g : \mathbb{R}^d \to \{R, 0, 1\}$, that are evaluated with the following risk functional :

$$L_R(g) = \mathbb{P}\{Y \neq g(X), g(X) \neq R\} + c\mathbb{P}\{g(X) = R\} \ .$$

What is the minimizing argument of $L_R(g)$ over all possible classifiers g with reject option ? Give a practical interpretation of the result.

Exercise 6 - We consider the model for classification data where X is a random vector on $\mathbb{R}^d$ and Y is a random variable taking values in $\{-1, +1\}$. We denote $\eta(x) = \mathbb{P}\{Y = +1 \mid X = x\}$ the posterior probability. We consider the following problems for which the question is to compute the optimal decision rule g^* or f^* - please also provide the main proof arguments.

1. Criterion to minimize : $R(g) = \mathbb{E}\big((Y - g(X))^2\big)$ where $g : \mathbb{R}^d \to \{-1, +1\}$
2. Criterion to minimize : $R(f) = \mathbb{E}\big((Y - f(X))^2\big)$ where $f : \mathbb{R}^d \to \mathbb{R}$
3. Criterion to minimize : $A(f) = \mathbb{E}\big(\log(1+e^{-Yf(X)})\big)$ where $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty, +\infty\}$.

Explain why such criteria are relevant for the binary classification problem.