Exercice 1 Nous nous plaçons dans le cadre de classification binaire $(X,Y) \in \mathbb{R}^d \times \{0,1\}$ et supposons que pour $\forall k \in \{0,1\}$ la loi conditionnelle $\mathbb{P}(X \in \cdot | Y = k)$ admet une densité de carré intégrable f_k par rapport à la mesure de Lebesgue. Rappelons que *Plug-In Rule* permet de majorer l'excès d'erreur par

$$L_g - L^* \leq 2\|\tilde{\eta} - \eta\|_{\mathbb{L}^1(\mathbb{R}^d,\, \mu_X)}$$

où $\tilde{\eta}(X)$ désigne un estimateur de $\eta(X) := \mathbb{P}(Y = 1|X) = \mathbb{E}[Y|X]$ et g la règle de décision

$$g(X) = 1_{\eta(X) > \frac{1}{2}}.$$

Dans cet exercice, on construit un tel estimateur et évalue sa consistance. Pour cela, on note n observations $(X_i, Y_i)_{1 \leq i \leq n}$ et

$$\begin{aligned} &p = \mathbb{P}(Y = 1) \\ &\alpha(x) = p f_1(x) - (1-p) f_0(x). \end{aligned}$$

1. Soit une base orthonormale $(\psi_j)_{j \geq 1}$ de $\mathbb{L}^2(\mathbb{R}^d, \lambda)$ vérifiant $\sup_{j,x} |\psi_j(x)| < +\infty$. Calculer l'espérance et la variance de

   $$\alpha_{n,j} = \frac{1}{n} \sum_{i=1}^n (2Y_i - 1)\psi_j(X_i).$$

   Quelle est l'interprétation de cet estimateur?

2. Prenons une approximation

   $$\tilde{\alpha}_n(\cdot) = \sum_{j=1}^{k_n} \alpha_{n,j}\psi_j(\cdot).$$

   Montrer la consistance faible du classifieur si $k_n \to \infty$ et $k_n/n \to 0$ lorsque le nombre d'observations n tends vers l'infini.

3. Montrer la consistance forte si k_n croît un peu moins vite que précédemment

   $$\lim_{n\to\infty} \frac{k_n \ln n}{n} \to 0.$$

4. Quel impact k_n a sur la performance du classifieur?

5. Suggérer une base orthonormale pour $\mathbb{L}^2(\mathbb{R}^d, \lambda)$.

6. Il est équivalent d'avoir la loi de X admettant une densité $L^2(\mathbb{R}^d, \lambda)$. Pourquoi?

Exercice 2 Dans cet exercice, on illustre l'approche de classification paramétrique par maximisation de vraisemblance. Pour cela, on pose un modèle de classification simple $(X,Y) \in \mathbb{R} \times \{0,1\}$. Suppons que la loi conditionelle sachant $Y = i$ est gaussienne $\mathcal{N}(m_i, 1)$ et qu'à nouveau, on dispose d'un jeu d'observations i.i.d.

1. Calculer les estimateurs pour (m_1, m_2) par maximisation de vraisemblance.

2. A quelle vitesse ces estimateurs convergent-t-ils en norme vers les vraie valeurs? (Indication: conditionner par la somme des Y_i.)

3. Evaluer la vitesse de décroissance de l'excès d'erreur et conclure.