Plug-In et Maximization de Vraisemblance

Exercice 1 On se place dans le cadre de classification binaire $(X, Y) \in \mathbb{R}^d \times \{0, 1\}$ où la loi marginale de X admet une densité de carré intégrable par rapport à la mesure de Lebesgue. On note n observations i.i.d. $(X_i, Y_i)_{1 \leq i \leq n}$.

1. Montrer que pour $k = 0$ ou 1, la loi conditionnelle $\mathbb{P}(X \in \cdot | Y = k)$ aussi admet une densité de carré intégrable par rapport à la mesure de Lebesgue.

2. 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

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

3. Prenons une approximation

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

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

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

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

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

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

Exercice 2 On illustre un exemple de classification paramétrique par maximization de vraisemblance. Pour cela, on pose un modèle simple $(X, Y) \in \mathbb{R} \times \{0, 1\}$. Supposons 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 leurs espérances?

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