Université Paris Dauphine – Année 2013-2014
Apprentissage et grande dimension

Durée : 2 heures

Attention ! On tiendra grand compte de la qualité de la rédaction et de la présentation. Les Sections 1 et 2 sont indépendantes.

1 Condition de marge

On considère le cadre de classification usuel : (X, Y) est un couple de variables aléatoires avec $Y \in \{0, 1\}$. On note $h^\star$ le classifieur de Bayes, $R^\star$ son risque de classification et

$$\eta(x) = \mathbb{E}[Y|X = x].$$

On suppose qu'il existe $\delta \geq 0$ tel que

$$\big|\eta(x) - 1/2\big| \geq \delta \quad \text{pour tout } x.$$

1. Soit $h(x) = \mathbf{1}_{\{\tilde{\eta}(x)>1/2\}}$ où $\tilde{\eta}$ est une fonction à valeurs dans $[0, 1]$. On note $R(h)$ son risque de classification. Montrer que

$$R(h) - R^\star \leq \mathbb{P}\left(|\tilde{\eta}(X) - \eta(X)| \geq \delta\right).$$

Soit $\mathcal{D}_n = \big((X_1, Y_1), \ldots, (X_n, Y_n)\big)$ où les (X_i, Y_i) sont de même loi que (X, Y) et indépendants. On suppose que X prends ses valeurs dans l'ensemble $\{1, \ldots, p\}$ où $p \geq 1$ est un entier.

On considère la règle de classification

$$h(x, \mathcal{D}_n) = \mathbf{1}_{\{\widehat{\eta}(x,\mathcal{D}_n)>\frac{1}{2}\}},$$

où

$$\widehat{\eta}(x,\mathcal{D}_n) = \begin{cases} \dfrac{1}{N_x}\displaystyle\sum_{i,\, X_i=x} Y_i & \text{sur} \quad \{N_x > 0\} \\ 0 & \text{sur} \quad \{N_x = 0\}, \end{cases}$$

avec $N_x = \mathrm{Card}\{i, X_i = x\}$.

2. Montrer que
$$\widehat{\eta}(x,\mathcal{D}_n) \to \mathbb{E}[Y|X=x]$$
presque-sûrement lorsque $n \to \infty$.

3. En déduire
$$\mathbb{E}\left[R_{\mathcal{D}_n}\big(h(\bullet,\mathcal{D}_n)\big)\right] \to R^\star \ \text{ lorsque } \ n \to \infty.$$

On va raffiner le résultat précédent en utilisant l'hypothèse $\delta > 0$.

4. Montrer que pour tout $x \in \{1,\dots,p\}$ on a[1]
$$\mathbb{P}\left(\left|\widehat{\eta}(x,\mathcal{D}_n) - \eta(x)\right| \geq \delta \mid N_x\right) \leq 2\exp(-2N_x\delta^2).$$

5. En déduire qu'il existe une constante $\gamma > 0$ telle que
$$\mathbb{E}\left[R_{\mathcal{D}_n}\big(h(\bullet,\mathcal{D}_n)\big)\right] - R^\star \leq \exp(-\gamma n).$$

2 Fonctionnelle quadratique

Soit $n \geq 1$. On se place dans le modèle de suite gaussienne *infinie* : on observe

$$Y_k = \theta_k + n^{-1/2}\xi_k, \ \ k = 1,2,\dots$$

où les ξ_k sont des variables aléatoires gaussiennes standard et $\theta = (\theta_k)_{k\geq 1}$ est le paramètre inconnu dans l'ellipsoide

$$\theta \in \Theta(\beta) = \Big\{\theta = (\theta_k)_{k\geq 1}, \ \ \sum_{k\geq 1} k^{2\beta}\theta_k^2 \leq 1\Big\}$$

[1]On rappelle l'inégalité de Hoeffding : si $Z_1,\dots,Z_M$ sont des variables aléatoires indépendantes telles que $a_i \leq Z_i \leq b_i$ et $\mathbb{E}[Z_i] = 0$, alors, pour $t > 0$, on a

$$\mathbb{P}\big(\sum_{i=1}^M Z_i \geq t\big) \leq \exp\big(-\frac{2t^2}{\sum_{i=1}^M (b_i-a_i)^2}\big).$$

avec $\beta > 0$. On souhaite estimer la fonctionelle $T(\theta) = \sum_{k\geq 1} \theta_k^2$.

Soit $L \geq 1$ un entier. On pose

$$\widehat{T}_L = \sum_{k=1}^{L} \big(Y_k^2 - n^{-1}\big).$$

1. Montrer que l'on a la décomposition

$$\widehat{T}_L - T(\theta) = -\sum_{k=L+1}^{\infty} \theta_k^2 + n^{-1/2} \sum_{k=1}^{L} \theta_k U_k + n^{-1} \sum_{k=1}^{L} V_k$$

où les variables aléatoires $(U_k)_{k\geq 1}$ sont indépendantes et identiquement distribuées et les $(V_k)_{k\geq 1}$ sont des variables aléatoires indépendantes et identiquement distribuées. Expliciter cette décomposition.

2. Montrer que pour tous réels a, b, c, on a : $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$.

3. En déduire

$$\mathbb{E}\left[\big(\widehat{T}_L - T(\theta)\big)^2\right] \leq C_1\big(L^{-4\beta} + n^{-1} + Ln^{-2}\big),$$

où $C_1 > 0$ est une constante que l'on déterminera.

4. En déduire que si $\beta \geq 1/4$, pour un choix de $L = L(n)$ que l'on précisera, on a

$$\sup_{\theta\in\Theta(\beta)} \mathbb{E}\left[\big(\widehat{T}_L - T(\theta)\big)^2\right] \leq C_2 n^{-1}$$

où $C_2 > 0$ est une constante que l'on déterminera.

5. Montrer que si $\beta > 1/4$, on a de plus la convergence en loi

$$\sqrt{n}\big(\widehat{T}_{L(n)} - T(\theta)\big) \xrightarrow{\text{loi}} \mathcal{N}\big(0, v(\theta)\big)$$

lorsque $n \to \infty$, où $\mathcal{N}\big(0, v(\theta)\big)$ désigne la loi normale centrée, de variance $v(\theta) > 0$. Déterminer $v(\theta)$.

6. Montrer que si $\beta < 1/4$, alors, pour un choix de $L = L(n, \beta)$ que l'on précisera, on a

$$\sup_{n\geq 1} n^{8\beta/(4\beta+1)} \sup_{\theta\in\Theta(\beta)} \mathbb{E}\left[\big(\widehat{T}_{L(\beta,n)} - T(\theta)\big)^2\right] < \infty.$$

7. Comparer avec les résultats de vitesse d'estimation du cours. Proposer une méthode adaptative pour estimer $T(\theta)$ lorsque β est inconnu mais que l'on a $\beta < 1/4$.