Apprentissage statistique: TD9
Maximum Likelihood

Emile Contal
http://econtal.perso.math.cnrs.fr/teaching

21 mars 2016

Soit (X, Y), η et D_n comme précédemment. Soit η_n une estimation de η. On définit la vraisemblance $\ell_n(\eta_n)$ et la log-vraisemblance $\mathcal{L}_n(\eta_n)$ comme suit :

$$\ell_n(\eta_n) = \prod_{i=1}^n \eta_n(X_i)^{Y_i} \big(1 - \eta_n(X_i)\big)^{1-Y_i}\,,$$
$$\mathcal{L}_n(\eta_n) = \frac{1}{n} \sum_{i=1}^n Y_i \log \eta_n(X_i) + (1 - Y_i) \log \big(1 - \eta_n(X_i)\big)\,.$$

Soit $\mathcal{F}$ une famille de fonctions de régression telle que $\eta \in \mathcal{F}$. Le classifieur par maximum de vraisemblance $\widehat{g}_n$ est donné par :

$$\widehat{g}_n(x) = \mathbb{1}_{\widehat{\eta}_n(x) \geq \frac{1}{2}} \text{ où } \widehat{\eta}_n = \operatorname*{argmax}_{\eta_n \in \mathcal{F}} \mathcal{L}_n(\eta_n)\,.$$

Exercice 1.

Montrer que lorsque $|\mathcal{F}| < \infty$ et $\eta \in \mathcal{F}$, le classifieur par maximum de vraisemblance $\widehat{g}_n$ est consistant, c'est à dire :

$$\Pr[\widehat{g}_n(X) \neq Y] \to L^\star\,.$$

On rappelle la définition de l'entropie $\mathcal{E}(\eta)$ pour tout η :

$$\mathcal{E}(\eta) = -\mathbf{E}\Big[\eta(X) \log \eta(X) + \big(1 - \eta(X)\big) \log \big(1 - \eta(X)\big)\Big]\,,$$

qui vérifie $0 \leq \mathcal{E}(\eta) \leq \log 2$, ainsi que la définition de la divergence (négative) :

$$D(\eta, \eta') = \mathbf{E}\left[\eta(X) \log \frac{\eta'(X)}{\eta(X)} + \big(1 - \eta(X)\big) \log \frac{1 - \eta'(X)}{1 - \eta(X)}\right]\,,$$

qui est négative et s'annule seulement lorsque $\eta'(X) = \eta(X)$ p.s.

Exercice 2.

1. Pour les familles suivantes, donner le classifieur $\widehat{g}_n$:

   (a) $\mathcal{F}_1 = \big\{\eta = \mathbb{1}_{[a,b]}\,, -\infty \le a \le b \le \infty\big\}$

   (b) $\mathcal{F}'_1 = \big\{\eta = c\mathbb{1}_{[a,b]}\,, -\infty \le a \le b \le \infty\,, c \in [0,1]\big\}$

   (c) $\mathcal{F}_2 = \big\{\eta = \mathbb{1}_{[a_1,b_1]\times\cdots\times[a_d,b_d]}\,, -\infty \le a_i \le b_i \le \infty\big\}$

   (d) $\mathcal{F}_3 = \big\{\eta(x) = \frac{cx}{1+cx}\,, x \ge 0\,, c \ge 0\big\}$

   (e) $\mathcal{F}_4 = \big\{\eta$ est décroissante sur $[0,1]\big\}$

   (f) $\mathcal{F}_5 = \big\{\eta(x) = \frac{1}{1+\|x-m\|^2}\,, m \in \mathbb{R}^d\big\}$

   (g) $\mathcal{F}_6 = \big\{\eta(x) = \sin^2(\theta x)\,, \theta \in \mathbb{R}\big\}$

2. Dire pour ces familles si $\widehat{g}_n$ est universellement consistant lorsque $\eta^\star \in \mathcal{F}$.

   On pourra utiliser pour certaines le théorème suivant :

   Théorème 1 (Consistance et ϵ-entropie)**.** *Soit $\mathcal{F}$ une famille de fonctions de régression, on appelle $N(\epsilon)$ son ϵ-entropie définie comme la cardinalité du plus petit ensemble de fonctions $\mathcal{F}_\epsilon$ tel que pour tout η' dans $\mathcal{F}$ il existe η'_L et η'_U dans $\mathcal{F}_\epsilon$ avec :*

   $$\begin{aligned} &\eta'_L \le \eta' \le \eta'_U\,,\\ &\mathbf{E}\Big[\eta'_U(X) - \eta'_L(X)\Big] \le \epsilon\,. \end{aligned}$$

   Lorsque $N(\epsilon) < \infty$ pour toute distribution de X et $\epsilon > 0$, le classifieur par maximum de vraisemblance est universellement consistant.