Apprentissage statistique: TD6
Maximum Likelihood
& Data-Dependant Partitioning

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

20 février 2015

Exercice 1.

Soit (X, Y) 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 :

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

1. Pour les familles suivantes, donner le classifieur $\widehat{g}_n$:
   (a) $\mathcal{F}_1 = \big\{ \eta = \mathbb{1}_{[a,b]} \,, -\infty \leq a \leq b \leq \infty \big\}$
   (b) $\mathcal{F}'_1 = \big\{ \eta = c\mathbb{1}_{[a,b]} \,, -\infty \leq a \leq b \leq \infty \,, c \in [0,1] \big\}$
   (c) $\mathcal{F}_2 = \big\{ \eta = \mathbb{1}_{[a_1,b_1] \times \cdots \times [a_d,b_d]} \,, -\infty \leq a_i \leq b_i \leq \infty \big\}$
   (d) $\mathcal{F}_3 = \big\{ \eta(x) = \frac{cx}{1+cx} \,, x \geq 0 \,, c \geq 0 \big\}$
   (e) $\mathcal{F}_4 = \big\{ \eta \text{ est décroissante sur } [0,1] \big\}$
   (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, soit :

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

   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 :*

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

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

Exercice 2. On considère ici les classifieurs par partitionnement de $\mathbb{R}$ où les partitions dépendent des données D_n de façon déterministes suivant la règle $\mathcal{P}_n = \pi_n(D_n)$. Soit la séquence de règles (π_n), le classifieur associé fonctionne par vote à l'intérieur des cellules. On notera $\mathcal{F}_n$ la famille de toutes les partitions possibles pour la règle π_n : $\mathcal{F}_n = \big\{\pi_n(D) : D \in (\mathbb{R} \times \{0,1\})^n\big\}$. En utilisant le théorème et les lemmes suivants, montrer qu'un classifieur par partitionnement de $\mathbb{R}$ tel que les intervalles contiennent au moins a_n et au plus b_n points, est fortement consistant lorsque $a_n \to \infty$ et $\frac{b_n}{n} \to 0$.

Théorème 2 (Consistance forte)**.** *On note* $\operatorname{diam}(A) = \sup_{x,y\in A} \|x - y\|$, *$S_M$ la boule fermée de rayon M centrée à l'origine, et $\mathcal{P}^{(M)}$ la restriction d'une partition $\mathcal{P}$ à la boule S_M. De même si $\mathcal{F}$ est une collection de partitions, on note $\mathcal{F}^{(M)} = \big\{\mathcal{P}^{(M)} : \mathcal{P} \in \mathcal{F}\big\}$. On introduit $\Delta_n(\mathcal{F}^{(M)})$ une mesure de complexité [1] d'une famille de partition. Soient $\mathcal{F}_1, \mathcal{F}_2, \ldots$ comme précédemment, si on a :*

$$\begin{aligned}&\forall M < \infty\,, \frac{\log \Delta_n(\mathcal{F}_n^{(M)})}{n} \to 0\,,\\ &\forall S_M \text{ et } \gamma > 0\,, \mu\Big(\Big\{x : \operatorname{diam}\big(A_n(x) \cap S_M\big) > \gamma\Big\}\Big) \to 0 \ p.s,\end{aligned}$$

alors le classifieur associé est fortement consistant.

Lemme 1. *Soit $\mathcal{F}^{(M)}$ une famille de partitions telle qu'il existe une constante N vérifiant pour tout $P^{(M)} \in \mathcal{F}^{(M)}$ que $\big|P^{(M)}\big| \leq N$, alors :*

$$\Delta_n(\mathcal{F}^{(M)}) \leq 2^N \Delta_n^*(\mathcal{F}^{(M)})\,,$$

où $\Delta_n^*(\mathcal{F}^{(M)})$ est le nombre maximal de façons de partitionner n points avec des éléments de $\mathcal{F}^{(M)}$.

Lemme 2. *Soit $\mathcal{I}$ l'ensemble des intervalles de $\mathbb{R}$, $\sup\limits_{I\in\mathcal{I}} |\mu(I) - \mu_n(I)| \to 0$ p.s.*

Lemme 3. *Soit H l'entropie binaire, alors $\log\binom{s}{t} \leq sH(\frac{t}{s})$.*

---

[1] On ne donne pas sa définition ici mais on pourra la borner en utilisant le Lemme 1. Les plus curieux iront voir le "shatter coefficient" et la "VC dimension".