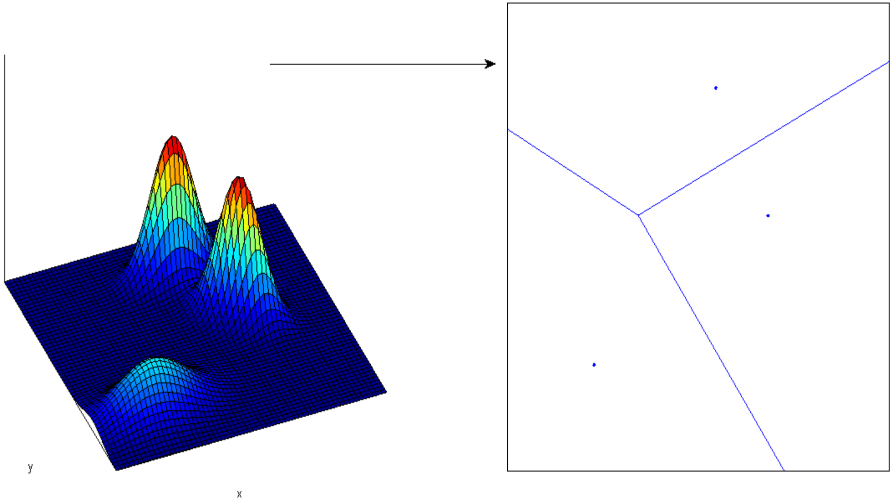
Quantification vectorielle sous condition de marge

C.Levrard

Université Paris 11 and UPMC
INRIA Projet Select

11ème Colloque Jeunes Probabilistes et Statisticiens
11 avril 2014

Principe



Principe

P une distribution sur $\mathbb{H}$, $dim(\mathbb{H}) = d \leq +\infty$

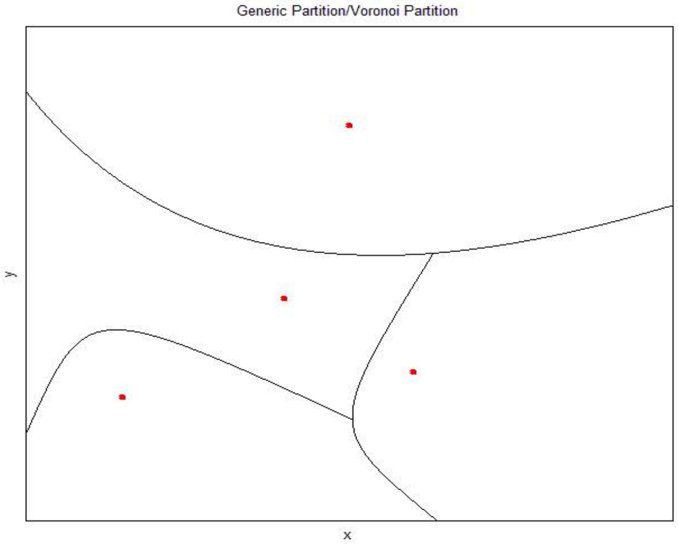
- séparer $\mathbb{H}$ selon k zones
- de manière pas totalement stupide
- en n'ayant accès qu'à un n-échantillon

Quantificateur vs dictionnaire

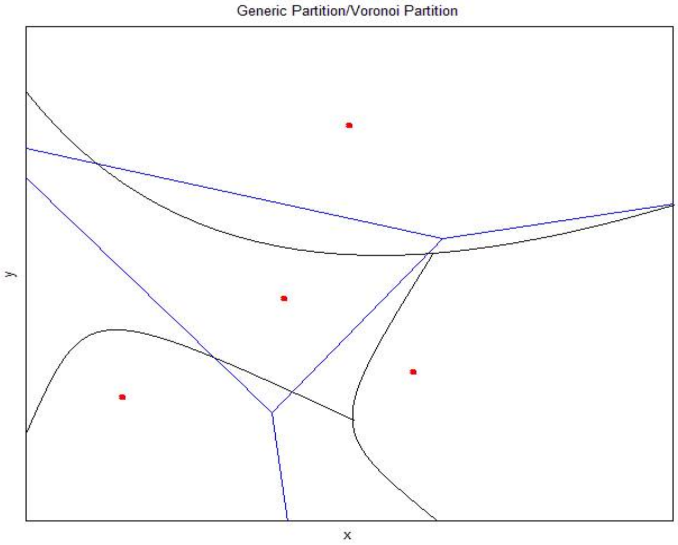
P une distribution sur $\mathbb{H}$, et Q un k-quantificateur

- $R(Q) = P\|x - Q(x)\|^2$

Quantificateur vs dictionnaire



Quantificateur vs dictionnaire



Quantificateur vs dictionnaire

P une distribution sur $\mathbb{H}$, et Q un k-quantificateur

- $R(Q) = P\|x - Q(x)\|^2$
- $\mathbf{c} = (c_1, \ldots, c_k)$ dictionnaire
- c_i mot

Quantificateur vs dictionnaire

P une distribution sur $\mathbb{H}$, et Q un k-quantificateur

- Risk/Distortion $R(\mathbf{c}) = \mathbb{E}(\min_{i=1,\dots,k} \|X - c_i\|^2) = P\gamma(\mathbf{c}, .)$
- $\mathbf{c} = (c_1, \dots, c_k)$ dictionnaire
- c_i mot
- Fonction de contraste

$$\gamma : \begin{cases} (\mathbb{R}^d)^k & \times & \mathbb{R}^d & \longrightarrow & \mathbb{R} \\ (\mathbf{c} & , & x) & \longmapsto & \min\limits_{i=1,\dots,k} \|c_i - x\|^2 \end{cases}$$

Stratégie ERM

$X_1, \ldots, X_n$ un n-échantillon.

- On veut : $\mathbf{c}^* = \arg\min P\gamma(\mathbf{c}, .)$
- On a : $\frac{1}{n}\sum_{i=1}^{n} \min_{j=1,\ldots,k} \|X_i - c_j\|^2 = P_n\gamma(\mathbf{c}, .)$

  $\rightarrow$ $\hat{\mathbf{c}}_n \in \arg\min P_n\gamma(\mathbf{c}, .)$

Stratégie ERM

$X_1, \ldots, X_n$ un n-échantillon.

- On veut : $\mathbf{c}^* = \arg\min P\gamma(\mathbf{c}, .)$
- On a : $\frac{1}{n}\sum_{i=1}^{n} \min_{j=1,\ldots,k} \|X_i - c_j\|^2 = P_n\gamma(\mathbf{c}, .)$

  $\rightarrow$ $\hat{\mathbf{c}}_n \in \arg\min P_n\gamma(\mathbf{c}, .)$

Ce qu'on veut majorer/minorer

$$\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \stackrel{(\mathrm{def})}{=} R(\hat{\mathbf{c}}_n) - R(\mathbf{c}^*) \quad \textit{or/and} \quad \mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*)$$

Vitesses faibles

Linder 94

On suppose que $\mathrm{Supp}(P) \subset \mathcal{B}(0,1)$. Alors

$$\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \lesssim \frac{\sqrt{kd}}{\sqrt{n}}.$$

Vitesses faibles

Biau, Devroye, Lugosi 2008

On suppose que $\mathrm{Supp}(P) \subset \mathcal{B}(0,1)$. Alors

$$\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \lesssim \frac{k}{\sqrt{n}}.$$

Vitesses faibles

Biau, Devroye, Lugosi 2008

On suppose que $\mathrm{Supp}(P) \subset \mathcal{B}(0,1)$. Alors

$$\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \lesssim \frac{k}{\sqrt{n}}.$$

Vitesse minimax (Bartlett, Linder and Lugosi 1998)

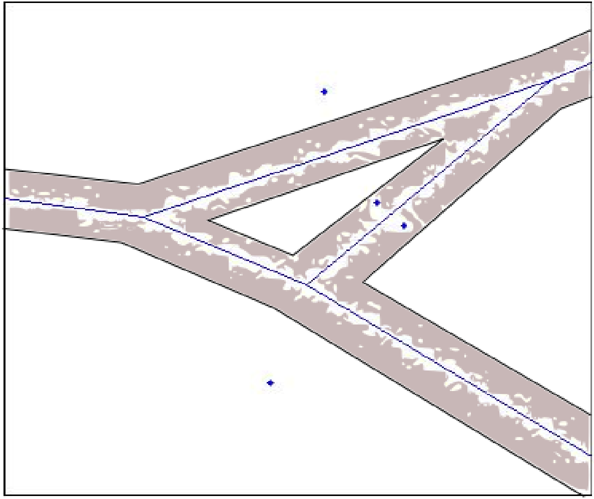
Soit $\mathcal{P}$ l'ensemble des distributions à support inclus dans $\mathcal{B}(0,1)$. Quels que soient k, d et $n \geq a_0 k$,

$$\sup_{P \in \mathcal{P}} \mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \gtrsim k\sqrt{\frac{k^{1-4/d}}{n}}.$$

Condition de marge

- N^* squelette de Voronoï
- $N^*(\varepsilon) = \{x | d(x, N^*) \leq \varepsilon\}$ ε-voisinage
- $p(\varepsilon) = \mathbb{P}(N^*(\varepsilon))$

Condition de marge



Condition de marge

- N^* squelette de Voronoï
- $N^*(\varepsilon) = \{x | d(x, N^*) \leq \varepsilon\}$ ε-voisinage
- $p(\varepsilon) = \mathbb{P}(N^*(\varepsilon))$

Condition de marge

$\mathrm{Supp}(P) \subset \mathcal{B}(0,1)$, et

$$\exists r_0 \quad x \leq r_0 \quad \Rightarrow \quad p(x) \leq ax,$$

où a est déterminé en fonction de P et k.

Condition de marge

Condition de marge

$\mathrm{Supp}(P) \subset \mathcal{B}(0,1)$, et

$$\exists r_0 \quad x \leq r_0 \quad \Rightarrow \quad p(x) \leq ax,$$

où a est déterminé en fonction de P et k.

Exemple :

- Gap : $p(x) = 0$ if $x \leq r$

Condition de marge

Condition de marge

Supp$(P) \subset \mathcal{B}(0,1)$, et

$$\exists r_0 \quad x \leq r_0 \quad \Rightarrow \quad p(x) \leq ax,$$

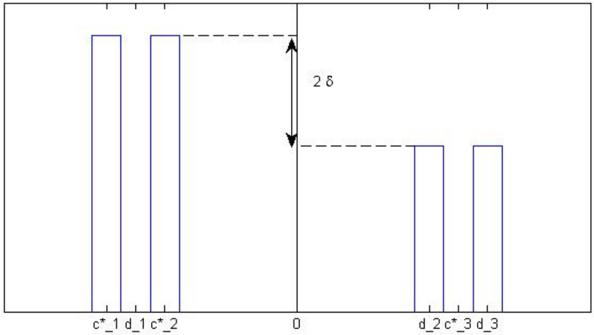
où a est déterminé en fonction de P et k.

Exemple :

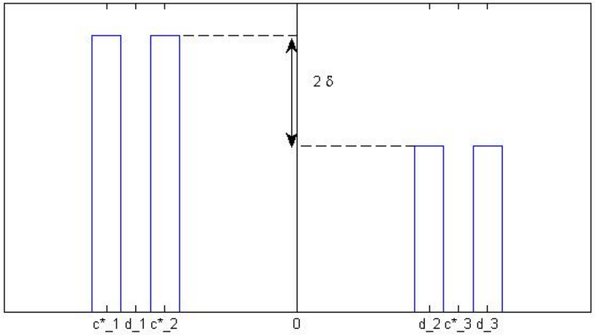
- Mélange gaussien ou presque :

$$\frac{p_{min}}{p_{max}} > \frac{512k\sigma^2}{(1-\epsilon)\tilde{B}^2(1-e^{-\tilde{B}^2/128\sigma^2})} \vee \frac{2k^2}{(1-\epsilon)\sigma^2} e^{-\tilde{B}^2/32\sigma^2},$$

Degré de séparation



Degré de séparation



$\longrightarrow$ δ écart de risque entre minimiseurs locaux seulement et minimiseurs globaux

Vitesse dimensionnée

L.2012

Si $\mathrm{Supp}(P) \subset \mathcal{B}(0,1)$, P est δ-séparée et satisfait une CM, alors

$$\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \lesssim \kappa_0(P)\frac{R(\mathbf{c}^*)}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)$$

$\longrightarrow$ $\kappa_0(P) = 4k\left(\frac{1}{\delta} \vee \frac{64}{p_{min}B^2 r_0^2}\right)$

Vitesse adimensionnée

L.2013

Si $\text{Supp}(P) \subset \mathcal{B}(0,1)$, P est δ-séparée et satisfait une CM, alors

$$\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \lesssim \kappa_0(P)\frac{C(P)}{n^{\frac{2}{3}}}$$

Procédure type Lasso

Quantification et sélection de variable selon

- $\hat{\mathbf{c}}_{n,\lambda} \in \arg\min P_n \gamma(\mathbf{c}, .) + \lambda_n \|\mathbf{c}\|_1$

Procédure type Lasso

Quantification et sélection de variable selon

- $\hat{\mathbf{c}}_{n,\lambda} \in \arg\min P_n\gamma(\mathbf{c},.) + \lambda_n \|\mathbf{c}\|_1$

Résultats verts

$\longrightarrow$ Si $\lambda \geq \frac{6k}{\sqrt{n}}$, alors

$$\mathbb{E}\ell(\hat{\mathbf{c}}_{n,\lambda}, \mathbf{c}^*) \leq \inf_{R>0} \left(\ell(\mathbf{c}_R, \mathbf{c}^*) + 2\lambda R\right) + \frac{2\lambda}{k} + \frac{c_0}{\sqrt{n}}$$

Procédure type Lasso

Quantification et sélection de variable selon

- $\hat{\mathbf{c}}_{n,\lambda} \in \arg\min P_n \gamma(\mathbf{c}, .) + \lambda_n \|\mathbf{c}\|_1$

Résultats verts

$\longrightarrow$ Si P satisfait une CM, et $\lambda \gtrsim \frac{\sqrt{k \log(kd)}}{\sqrt{n}}$, alors, pour n assez grand,

$$\|\hat{\mathbf{c}}_{n,\lambda} - \mathbf{c}_n^*\|_1 \leq \frac{C(k,d)}{\sqrt{n}},$$

avec

$$\mathbf{c}_n^* \in \arg\min \ell(\mathbf{c}, \mathbf{c}^*) + c_0 \lambda^2 \|\mathbf{c}\|_0.$$

Merci de votre attention