Margin conditions for vector quantization

C.Levrard[1]

[1]Université Paris 11 and UPMC
INRIA Projet Select
clement.levrard@math.u-psud.fr

5ème rencontre des jeunes statisticiens
Aussois 2013

Principle

P a distribution

- sum up P with k points
- without too much loss
- according to a n-sample

Codebooks/Risk

P a distribution on $\mathbb{R}^d$

- $\mathbf{c} = (c_1, \ldots, c_k)$ codebook
- c_i codepoint

Codebooks/Risk

P a distribution on $\mathbb{R}^d$

- $\mathbf{c} = (c_1, \ldots, c_k)$ codebook
- c_i codepoint
- Risk/Distortion $R(\mathbf{c}) = \mathbb{E}(\min_{i=1,\ldots,k} \|X - c_i\|^2) = P\gamma(\mathbf{c}, .)$
- Contrast function

$$\gamma : \left\{ \begin{array}{ccccc} (\mathbb{R}^d)^k & \times & \mathbb{R}^d & \longrightarrow & \mathbb{R} \\ (\mathbf{c} & , & x) & \longmapsto & \min_{i=1,\ldots,k} \|c_i - x\|^2 \end{array} \right.$$

ERM Strategy

$X_1, \ldots, X_n$ a n-sample.

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

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

ERM Strategy

$X_1, \dots, X_n$ a n-sample.

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

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

Quantity of interest

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

Basic rate

Linder, Lugosi and Zeger, 1994

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$, then $\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \leq \frac{C(P,k,d)}{\sqrt{n}}$.

Basic rate

Linder, Lugosi and Zeger, 1994

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$, then $\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \leq \frac{C(P,k,d)}{\sqrt{n}}$.

Minimax rate (Bartlett, Linder and Lugosi 1998)

Let $\mathcal{P}$ denotes the set of distributions supported on $\mathcal{B}(0,1)$. For every k, d and $n \geq a_0 k$,

$$\sup_{P \in \mathcal{P}} \mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \geq a_1 k \sqrt{\frac{k^{1-4/d}}{n}}.$$

Technical margin condition

A-TMC

There exists $A > 0$ such that

$$\forall \mathbf{c} \quad AP(\gamma(\mathbf{c},.) - \gamma(\mathbf{c}^*,.)) \geq \mathrm{Var}(\gamma(\mathbf{c},.) - \gamma(\mathbf{c}^*,.)).$$

Antos et al., 2004

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$ and P satisfies a TMC, for some $A > 0$, then

$$\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \leq C(P,k,d)\frac{\log(n)}{n}.$$

Technical margin condition

A-TMC

There exists $A > 0$ such that

$$\forall \mathbf{c} \quad AP(\gamma(\mathbf{c},.) - \gamma(\mathbf{c}^*,.)) \geq \mathrm{Var}(\gamma(\mathbf{c},.) - \gamma(\mathbf{c}^*,.)).$$

L. 2011

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$, P has a continuous density, and P satisfies a TMC, then

$$\mathbb{E}\ell(\hat{\mathbf{c}}_n, \mathbf{c}^*) \leq \frac{C(P,k,d)}{n}.$$

Pollard's regularity condition

Suppose that P has a continuous density f and $\mathbb{P}(\|X\| \leq 1) = 1$, then

- $\mathbf{c} \longmapsto P\gamma(\mathbf{c}, .)$ is twice differentiable
- its hessian matrix H is made of the following $d \times d$ blocks (σ means intergration w.r.t the $d-1$-dimensional lebesgue measure):

$\longrightarrow$ $2P(V_i)I_d - 2\sum_{\ell \neq i} r_{i\ell}^{-1}\sigma\left[f(x)(x-c_i)(x-c_i)^t \mathbf{1}_{\partial(V_i \cap V_\ell)}\right]$ when $i = j$

$\longrightarrow$ $2r_{ij}^{-1}\sigma\left[f(x)(x-c_i)(x-c_j)^t \mathbf{1}_{\partial(V_i \cap V_j)}\right]$ when $i \neq j$.

Pollard's regularity condition

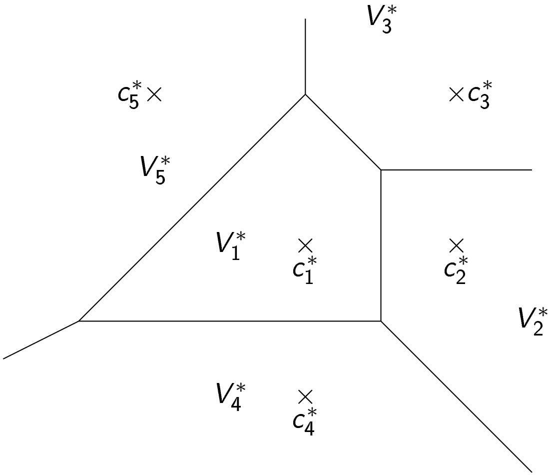
Pollard's regularity condition

Suppose that P has a continuous density f and $\mathbb{P}(\|X\| \leq 1) = 1$. Then P satisfies Pollard's regularity condition if H is positive definite at every possible optimal quantizer $\mathbf{c}^*$.

Antos 04

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$ and P satisfies Pollard's regularity condition, then P satisfies a TMC.

Pollard's regularity condition



A more practical condition for continuous densities

- f the density of P
- $B = \min_{\mathbf{c}^*, i \neq j} \|c_i^* - c_j^*\|$
- N^* the union of all boundaries of Voronoi cells for the optimal quantizer $\mathbf{c}^*$

A more practical condition for continuous densities

- f the density of P
- $B = \min_{\mathbf{c}^*, i \neq j} \|c_i^* - c_j^*\|$
- N^* the union of all boundaries of Voronoi cells for the optimal quantizer $\mathbf{c}^*$

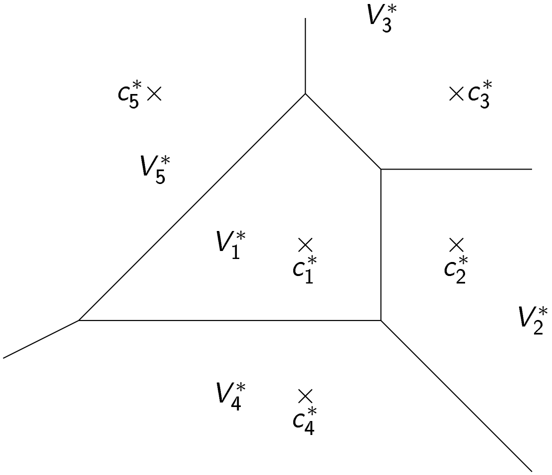
L. 2011

If $\|f_{|N^*}\|_\infty < C(d) B \inf_{i=1,\ldots k} \mathbb{P}(V_i^*)$, then P satisfies Pollard's regularity condition (thus satisfies a TMC).

A true margin-type condition

- N^* union of optimal Voronoï cells boundary
- $N^*(\varepsilon) = \{x | d(x, N^*) \leq \varepsilon\}$ ε-neighborhood
- $p(\varepsilon) = \mathbb{P}(N^*(\varepsilon))$ thickness of the boundary

A true margin-type condition



A true margin-type condition

L. 2012

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$ and suppose that

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

where a is fixed and depends on k. Then P satisfies a A-TMC with explicit constant A.

A true margin-type condition

L. 2012

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$ and suppose that

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

where a is fixed and depends on k. Then P satisfies a A-TMC with explicit constant A.

- Polynomial thickness: $p(x) \leq a_1 x + a_q x^q$, $q > 1$

  $\rightarrow$ If $a_1 \leq Bp_{min}/64$, then P satisfies a A-TMC

A true margin-type condition

L. 2012

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$ and suppose that

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

where a is fixed and depends on k. Then P satisfies a A-TMC with explicit constant A.

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

  $\rightarrow$ P satisfies a A-TMC, with $r_0 = \frac{Br}{4\sqrt{2}}$.

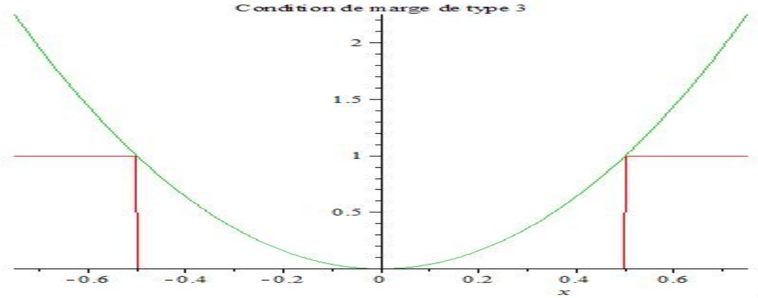
A true margin-type condition

L. 2012

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$ and suppose that

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

where a is fixed and depends on k. Then P satisfies a A-TMC with explicit constant A.



A true margin-type condition

L. 2012

Suppose that $\mathbb{P}(\|X\| \leq 1) = 1$ and suppose that

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

where a is fixed and depends on k. Then P satisfies a A-TMC with explicit constant A.

- Quasi-gaussian mixture, $d = 2$

  $\rightarrow$ Suppose that P satisfies

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

then P satisfies a A-TMC.

Thanks for your attention

- A.GERSHO and R.M.GRAY, *Vector quantization and Signal compression*, 1992
- T.LINDER, *Learning-Theoretic Methods in Vector Quantization*, CISM Courses and Lectures, Springer, Vienna, 2002
- V.KOLTCHINSKII, *Local Rademacher Complexities and Oracle Inequalities in Risk Minimization*, The Annals of Statistics, 2006
- A.ANTOS, L.GYÖRFI, and A.GYÖRGY, *Improved Convergence Rates in Empirical Vector Quantizer Design*, 2004
- C.LEVRARD, *Fast rates for empirical vector quantization*, Electronic Journal of Statistics, 2013.