Nonparametric Regression in Orlicz spaces?

Saverio Salzo

DIMA, Università di Genova
Slipguru Research group

Convegno Italo-Francese
Ottimizzazione e processi dinamici
in apprendimento statistico e problemi inversi

Sestri Levante, 8-12 September 2014

Classical nonparametric regression

Let $(\Omega, \mathfrak{A}, \mathrm{P})$ be a probability space, let $(\mathcal{X}, \mathfrak{A}_{\mathcal{X}})$ and $(\mathbb{R}, \mathfrak{A}_{\mathbb{R}})$ be two measurable spaces, and let $(X, Y)\colon (\Omega, \mathfrak{A}, \mathrm{P}) \to (\mathcal{X} \times \mathbb{R}, \mathfrak{A}_{\mathcal{X}} \otimes \mathfrak{A}_{\mathbb{R}})$ be a random variable with distribution P on $\mathcal{X} \times \mathbb{R}$ and P has marginal $P_{\mathcal{X}}$. Suppose $\mathrm{E}|Y|^2 < +\infty$.

The L^2-*risk associated with* P is

$$R\colon \mathcal{M}(\mathcal{X}, \mathbb{R}) \to [0, +\infty]$$
$$f \mapsto \int_{\mathcal{X}\times\mathbb{R}} |y - f(x)|^2 \,\mathrm{d}P(x, y) = \mathrm{E}|Y - f(X)|^2.$$

Classical nonparametric regression

Let $(\Omega, \mathfrak{A}, \mathrm{P})$ be a probability space, let $(\mathcal{X}, \mathfrak{A}_{\mathcal{X}})$ and $(\mathbb{R}, \mathfrak{A}_{\mathbb{R}})$ be two measurable spaces, and let $(X, Y)\colon (\Omega, \mathfrak{A}, \mathrm{P}) \to (\mathcal{X} \times \mathbb{R}, \mathfrak{A}_{\mathcal{X}} \otimes \mathfrak{A}_{\mathbb{R}})$ be a random variable with distribution P on $\mathcal{X} \times \mathbb{R}$ and P has marginal $P_{\mathcal{X}}$. Suppose $\mathrm{E}|Y|^2 < +\infty$.

The L^2*-risk associated with* P is

$$\begin{aligned} R\colon \mathcal{M}(\mathcal{X}, \mathbb{R}) &\to [0, +\infty] \\ f &\mapsto \int_{\mathcal{X}\times\mathbb{R}} |y - f(x)|^2 \, \mathrm{d}P(x, y) = \mathrm{E}|Y - f(X)|^2. \end{aligned}$$

The goal is to find a regression function $f_*\colon \mathcal{X} \to \mathbb{R}$ such that $f_*(X) \approx Y$, starting from a sequence $(X_i, Y_i)_{i\in\mathbb{N}}$ of independent copies of (X, Y). This is equivalent to minimizing the L^2-risk.

L^2-risk

Consider the embedding

$$\iota\colon L^2(\mathcal{X}, P_{\mathcal{X}}) \to L^2(\mathcal{X} \times \mathbb{R}, P), \quad (\iota f)(x,y) = f(x)$$

and define $g\colon \mathcal{X} \times \mathbb{R} \to \mathbb{R}\colon (x,y) \mapsto y$. Then

$$R(f) = \int_{\mathcal{X}\times\mathbb{R}} |y - f(x)|^2 \, \mathrm{d}P(x,y) = \|g - \iota f\|_2^2.$$

Therefore, minimizing $R(f)$ corresponds to find the projection, say $\pi(g)$,

of g onto $\iota(L^2(\mathcal{X}, P_{\mathcal{X}}))$ in the space $L^2(\mathcal{X} \times \mathbb{R}, P)$. Indeed dom $R = L^p(\mathcal{X})$ and, setting $\iota f_* = \pi(g)$,

$$(\forall\, f \in L^p(\mathcal{X})) \quad R(f) - \inf_{L^2(\mathcal{X})} R = \|g - \iota f\|_2^2 - \|g - \pi(g)\|_2^2 = \|f - f_*\|_2^2$$

Some Geometry of Banach spaces

- metric projection in uniformly convex Banach spaces and variational characterization.
- the concept of orthogonality of Birkhoff-James.
- metric projection onto affine closed sets and characterization by orthogonality.

Two fundamental inequalities

We start by recalling a theorem about inequalities in Banach spaces.

Theorem (Xu, Roach 91')

Let $\mathcal{B}$ be Banach space and $p \in]1, +\infty[$. If $\mathcal{B}$ is uniformly convex, then

$$(\forall u \in \mathcal{B})(\forall u^* \in J_p(u))(\forall v \in \mathcal{B}) \ \ \|u+v\|^p - \|u\|^p \geq p\langle v, u^*\rangle + \sigma_p(u,v)$$

where

$$\sigma_p(u,v) = pK_p \int_0^1 \frac{(\|u+tv\| \vee \|u\|)^p}{t} \delta_{\mathcal{B}}\Big(\frac{t\|v\|}{2\|u+tv\| \vee \|u\|}\Big) \mathrm{d}t\,.$$

and $K_p > 0$ is a constant.

$$J_p(u) = \partial(1/p\|\cdot\|^p)(u) = \big\{u^* \in \mathcal{B}^* \ \big| \ \langle u, u^*\rangle = \|u\|\|u^*\|, \ \|u^*\| = \|u\|^{p-1}\big\}$$

Two fundamental inequalities

We start by recalling a theorem about inequalities in Banach spaces.

Theorem (Xu, Roach 91')

Let $\mathcal{B}$ be Banach space and $p \in]1, +\infty[$. If $\mathcal{B}$ is uniformly smooth then

$$(\forall u \in \mathcal{B})(\forall u^* \in J_p(u))(\forall v \in \mathcal{B}) \quad \|u+v\|^p - \|u\|^p \leq p\langle v, u^* \rangle + \tilde{\sigma}_p(u,v)$$

where

$$\tilde{\sigma}_p(u,v) = p\, G_p \int_0^1 \frac{(\|u+tv\| \vee \|u\|)^p}{t} \rho_{\mathcal{B}}\left(\frac{t\|v\|}{\|u+tv\| \vee \|u\|} \right) \mathrm{d}t\,.$$

and $G_p > 0$ is a constant.

$$J_p(u) = \partial(1/p\|\cdot\|^p)(u) = \left\{ u^* \in \mathcal{B}^* \;\middle|\; \langle u, u^* \rangle = \|u\|\|u^*\|,\ \|u^*\| = \|u\|^{p-1} \right\}$$

Pythagorean inequalitites

Corollary

Let $\mathcal{B}$ be a uniformly convex Banach space, $\mathcal{C}$ a closed affine subspace, $u \in \mathcal{B} \setminus \mathcal{C}$ and $p > 1$. Then the following holds

1. *If $\mathcal{B}$ has modulus of convexity of power type $q > 1$, then for every $v \in \mathcal{C}$*

$$\|u - v\|^p - \|u - \pi_{\mathcal{C}}(u)\|^p \geq \begin{cases} \dfrac{K_p c_{\mathcal{B}}}{2^p} \|v - \pi_{\mathcal{C}}(u)\|^p & \text{if } q \leq p \\ \dfrac{p}{q} \dfrac{K_p c_{\mathcal{B}}}{2^q} \dfrac{\|v - \pi_{\mathcal{C}}(u)\|^q}{\|u - v\|^{q-p}} & \text{if } q > p. \end{cases}$$

2. *If $\mathcal{B}$ is uniformly smooth and has modulus of smoothness of power type $q > 1$, then for every $v \in \mathcal{C}$*

$$\|u - v\|^p - \|u - \pi_{\mathcal{C}}(u)\|^p \leq \begin{cases} 2^{q-p} G_p b_{\mathcal{B}} \|v - \pi_{\mathcal{C}}(u)\|^p & \text{if } q \geq p \\ \dfrac{p}{q} G_p b_{\mathcal{B}} \|v - \pi_{\mathcal{C}}(u)\|^q \|u - v\|^{p-q} & \text{if } q < p. \end{cases}$$

L^p-risk

Let $p \in]1, +\infty[$ and suppose $\mathsf{E}|Y|^p < +\infty$.

The *L^p-risk associated with P* is

$$R \colon \mathcal{M}(\mathcal{X}, \mathbb{R}) \to [0, +\infty]$$
$$f \mapsto \int_{\mathcal{X}\times\mathcal{Y}} |y - f(x)|^p \, \mathrm{d}P(x, y) = \mathsf{E}|Y - f(X)|^p.$$

L^p-risk

Let $p \in]1, +\infty[$ and suppose $\mathsf{E}|Y|^p < +\infty$.

The *L^p-risk associated with P* is

$$R\colon \mathcal{M}(\mathcal{X}, \mathbb{R}) \to [0, +\infty]$$
$$f \mapsto \int_{\mathcal{X}\times\mathcal{Y}} |y - f(x)|^p \, \mathrm{d}P(x, y) = \mathsf{E}|Y - f(X)|^p.$$

The goal is to solve

$$\min_{f \in \mathcal{M}(\mathcal{X}, \mathbb{R})} R(f)$$

starting from a sequence $(X_i, Y_i)_{i \in \mathbb{N}}$ of independent copies of (X, Y).

The p-regression function

Theorem

There exists the regression function f_*, that is $R(f_*) = \inf_{\mathcal{M}(\mathcal{X};\mathbb{R})} R$ and it is $f_* \in L^p(\mathcal{X}, P_\mathcal{X})$. Moreover for every $f \in L^p(\mathcal{X}, P_\mathcal{X})$ the following two inequalities hold

$$R(f) - \inf_{\mathcal{M}(\mathcal{X};\mathbb{R})} R \le C_p \|f - f_*\|_p^{\min\{2,p\}} \Big(\inf_{\mathcal{M}(\mathcal{X};\mathbb{R})} R + \|f - f_*\|_p \Big)^{\max\{2,p\}-2}$$

$$\|f_* - f\|_p^{\max\{2,p\}} \le D_p \Big(R(f) - \inf_{\mathcal{M}(\mathcal{X};\mathbb{R})} R \Big) R(f)^{\frac{2-\min\{2,p\}}{p}}$$

where

$$C_p = \begin{cases} G_p/p & \text{if } p \le 2 \\ p(p-1)G_p/4 & \text{if } p > 2 \end{cases} \qquad D_p = \begin{cases} 64/(p(p-1)K_p) & \text{if } p < 2 \\ p2^{2p}/K_p & \text{if } p \ge 2\,. \end{cases}$$

Feature map in Banach spaces

Let $A\colon \mathcal{F} \to \mathbb{R}^{\mathcal{X}}$ be a linear and continuous operator for the topology of pointwise convergence on $\mathbb{R}^{\mathcal{X}}$ and let $\Lambda\colon \mathcal{X} \to \mathcal{F}^*$ be the associated feature map, in the sense that

$$Au(x) = \langle u, \Lambda(x) \rangle$$

Theorem

Then the following hold:

1. $\Lambda\colon \mathcal{X} \to \mathcal{F}^*$ *is measurable if and only if* $\operatorname{ran} A \subset \mathcal{M}(\mathcal{X}, \mathbb{R})$.
2. *Let* $p \in [1, +\infty]$ *and suppose that* $\Lambda \in L^p(\mathcal{X}, \mu; \mathcal{F}^*)$. *Then* $\operatorname{ran} A \subset L^p(\mathcal{X}, \mu)$ *and,* $A\colon \mathcal{F} \to L^p(\mathcal{X}, \mu)$ *is continuous with* $\|A\| \leq \|\Lambda\|_p$.

When $\mathcal{F}$ is strictly convex and smooth then

$$K(x, y) = \left\langle J_{\mathcal{F}}^{-1}(\Lambda(x)), J(y) \right\rangle \tag{1}$$

Example:
Let $\mathcal{X} \subset \mathbb{R}^d$ bounded with regular boundary, and $p > d$. Then $W_0^{1,p}(\mathcal{X}) \subset \mathscr{C}(\overline{\Omega})$ and the evaluation functionals $\mathrm{ev}_x \colon W_0^{1,p}(\mathcal{X}) \to \mathbb{R}$ are continuous.

The problem of consistency

Take:

- $\mathcal{F}$ a feature Banach space (such that $\mathcal{F}^*$ has type q^*), a linear map $A\colon \mathcal{F} \to L^p(\mathcal{X}, P)$ with corresponding feature map $\Lambda\colon \mathcal{X} \to \mathcal{F}^*$.
- a regularizer $G\colon \mathcal{F} \to \mathbb{R}_+$ totally convex on bounded sets with modulus of total convexity ψ.
- a vanishing sequence $(\lambda_n)_{n\in\mathbb{N}}$.

The problem is to approach the regression function f_* by means of approximate solutions $u_{n,\lambda_n}(Z_n)$ to the empirical regularized problems

$$\min_{u\in\mathcal{F}} R_n(Au, Z_n) + \lambda_n G(u), \tag{2}$$

in the sense that $\|Au_{n,\lambda_n}(Z_n) - f_*\|_p \to 0$ in probability (*weak consistency*) or almost surely (*strong consistency*), under suitable conditions on $(\lambda_n)_{n\in\mathbb{N}}$.

- Set $F = R \circ A\colon \mathcal{F} \to F$
- Set, for every $\lambda > 0$, $u_\lambda = \mathrm{argmin}_{u \in \mathcal{F}} F(u) + \lambda G(u)$.

Then

$$R(Au) - \inf_{\mathcal{M}(\mathcal{X},\mathbb{R})} R = \Big(R(Au) - \inf_{\operatorname{ran} A} R\Big) + \Big(\inf_{\operatorname{ran} A} R - \inf_{\mathcal{M}(\mathcal{X},\mathbb{R})} R\Big)$$

Usually one further split

$$R(Au) - \inf_{\operatorname{ran} A} R = \Big(R(Au) - R(Au_\lambda)\Big) + \Big(R(Au_\lambda) - \inf_{\operatorname{ran} A} R\Big)$$

and then the (local) Lipschitz continuity of R is used.

Theorem

Set $t = \max\{2, p\}$, $m = \min\{2, p\}$. *Let* $V_p : \mathcal{F} \to \mathbb{R}_+$ *be such that for every* $v \in \mathcal{F}$

$$V_p(u_\lambda) = D_p^{1/t} R(Au_\lambda)^{\frac{t-p}{tp}} \|A\|^{-1} \big(R(Au_\lambda) - \inf R(\operatorname{ran} A)\big)^{1/t}.$$

Then for every $(u, v) \in \mathcal{F}^2$,

$$R(Au) - \inf_{\operatorname{ran} A} R \leq C_p \|A\|^p \big(\|u - u_\lambda\| + V_p(u_\lambda)\big)^m \cdot \big(\|u - u_\lambda\| + \|A\|^{-1} R(Au_\lambda)^{1/p}\big)^{p-m}.$$

We do not use any Lipschitzian property.

Theorem (Consistency)

Suppose that G is totally convex on bounded sets with modulus of convexity ψ. Set $\rho_\lambda = \psi_0^\natural((\mathsf{E}|Y|^p + 1)/\lambda)$. Then the following holds:

1. *For every $\lambda > 0$, $\tau > 1$ and $n \in \mathbb{N}^*$*

$$\mathsf{P}^*\Bigg[R(Au_{n,\lambda}(Z_n)) - \inf F(\mathcal{F}) > C_p\|A\|^p \cdot \big((\hat{\psi}_{\rho_\lambda})^\natural\big(\alpha(n,\tau,\lambda)\big) + V_p(u_\lambda)\,\big)^{\min\{p,2\}} \cdot \Big((\hat{\psi}_{\rho_\lambda})^\natural\big(\alpha(n,\tau,\lambda)\big) + \|A\|^{-1}F(u_\lambda)^{1/p}\Big)^{p-\min\{p,2\}}\Bigg] \leq 1/\tau.$$

where

$$\alpha(n,\tau,\lambda) = \kappa p T_{q^*}\tau^{1/m^*}\frac{(\kappa\rho_\lambda + \|Y\|_p)^{p-1}}{\lambda n^{1/m}}, \quad m = \max\{p,q\}.$$

Theorem (Consistency)

Suppose that

$$(\forall\, K > 1) \quad (\hat{\psi}_{\rho_{\lambda_n}})^{\natural}\left(K \frac{\rho_{\lambda_n}^{p-1}}{\lambda_n n^{1/m}}\right) \to 0\,.$$

Then $\|Au_{n,\lambda_n}(Z_n) - f_*\|_p \to 0$ *in* P*-probability, where* f_* *is the regression function.*

Theorem (Consistency)

Suppose that

$$\begin{cases} \mathcal{F} \text{ is uniformly convex with modulus of convexity of power type } q \\ G = \eta\|\cdot\|^r + H, \quad \text{where } \eta \in \mathbb{R}_{++},\ r \in \,]1, +\infty[\,,\ \text{and } H \in \Gamma_0^+(\mathcal{F}). \end{cases}$$

Set $\rho_\lambda = ((\mathsf{E}|Y|^p + 1)/\lambda)^{1/r}$ *and set* $m = \max\{p, q\}$*. Suppose that*

$$\frac{1}{\lambda_n^{p/r} n^{1/m}} \to 0\,.$$

Then $\|Au_{n,\lambda_n}(Z_n) - f_*\|_p \to 0$ *in* P*-probability,*

References

P. L. Combettes, S. Salzo and S. Villa, Consistency of Regularized Learning Schemes in Banach spaces. *In preparation*.

Z B. Xu and G.F. Roach, Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces , *Journal of Mathematical Analysis and Applications*, Vol. 157, No. 1, pp. 189–210, 1991.