Quantile Regression: Multi-Task Approaches

Delanoue Pierre, Meunier Dimitri

ENSAE Paris , 3rd year engineering cycle

May 4, 2020

Motivations

- Breaking out of the dictatorship of the average: heterogeneity, robustness
- ”Non-Crossing” Problem
- Multidimensional Targets

Table of Contents

Table of Contents

Classical Quantile Regression

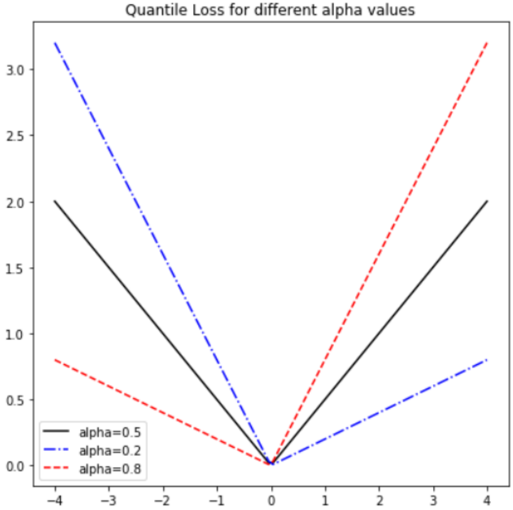
For $\alpha \in (0, 1)$, the α-th quantile of a random variable $\mathbf{y}$ on $\mathbb{R}$ is defined by:

$$q_{\mathbf{y}}(\alpha) = \inf\{x \in \mathbb{R}, F_{\mathbf{y}}(x) \geq \alpha\}$$

where $F_{\mathbf{y}}$ is the distribution function of $\mathbf{y}$.

- (i) $\alpha \longmapsto q_{\mathbf{y}}(\alpha)$ is non-decreasing
- (ii) If $U \sim \mathcal{U}([0, 1])$, then $q_{\mathbf{y}}(U) = \mathbf{y}$ with probability one.

Pinball Loss



Linear Model Estimation

$$\forall \alpha \in (0,1),\ \exists \beta_\alpha \in \mathbb{R}^q \quad s.t. \quad q_\alpha(Y|X) = \beta_\alpha^T X \tag{1}$$

$$\beta_\alpha \in arg \min_{\beta \in \mathbb{R}^q} \mathbb{E}[\rho_\alpha(Y - X'\beta)] \tag{2}$$

Given a dataset $\mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\}$, the quantile regression estimator is naturally built as:

$$\widehat{\beta}_\alpha \in arg \min_{\beta \in \mathbb{R}^q} \frac{1}{n} \sum_{i=1}^{n} \rho_\alpha(\mathrm{y}_i - x_i'\beta) \tag{3}$$

Table of Contents

Kernelised Quantile Regression

Given a positive definite kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and the associated (unique) *Reproducing Kernel Hilbert Space* $\mathcal{H} \subset (\mathbb{R})^{\mathcal{X}}$. We can *kernelized* the linear model as follows,

$$(f_\alpha, b_\alpha) \in \underset{f \in \mathcal{H}, b \in \mathbb{R}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \rho_\alpha(\mathrm{y}_i - f(x_i) - b) + \frac{\lambda}{2} ||f||_{\mathcal{H}}^2 \qquad \lambda > 0 \tag{4}$$

Applying the representer theorem, we get a Quandratic Program (QP).

The primal and the dual can be solved with interior point methods.

Table of Contents

Simultaneous Kernelised Quantile Regression

Goal: Learn p quantile functions with $0 < \alpha_1 < \ldots < \alpha_p < 1$

First approach. Learn p kernels and impose hard constraints to avoid crossing.

$$\begin{aligned} \underset{(w_{\alpha_1},\cdots,w_{\alpha_p})\in\mathbb{R}^{n\times p},b\in\mathbb{R}^p}{\text{minimize}} \quad & \sum_{k=1}^{p}\sum_{i=1}^{n}\rho_{\alpha_k}\left(y_i - \sum_{j=1}^{n} w_{\alpha_k,j}K(x_j,x_i) - b_k\right) + \frac{\lambda}{2}\sum_{k=1}^{p} w_{\alpha_k}^T K^n w_{\alpha_k} \\ \text{subject to} \quad & w_{\alpha_k,i} \leq w_{\alpha_{k+1},i}, \quad \forall i \in [\![1,n]\!], \quad \forall k \in [\![1,p-1]\!] \\ & b_k \leq b_{k+1}, \quad \forall k \in [\![1,p-1]\!] \end{aligned} \tag{5}$$

If the kernel satisfies $K(x,x') \geq 0, \quad \forall (x,x') \in \mathcal{X}^2$, the hard constraints ensure non-crossing.

Simultaneous Kernelised Quantile Regression

Second approach. Learn a vector valued kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^{p \times p}$

$$(f_\alpha, b_\alpha) \in \underset{f \in \mathcal{H}, b \in \mathbb{R}^p}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{p} \rho_{\alpha_k}(\mathrm{y}_i - f_k(x_i) - b_k) + \frac{\lambda}{2} ||f||_{\mathcal{H}}^2 \qquad \lambda > 0 \tag{6}$$

Computation

$$
\begin{aligned}
\underset{f\in\mathcal{H}, b\in\mathbb{R}^p, \xi^+\in(\mathbb{R}^n)^p, \xi^-\in(\mathbb{R}^n)^p}{\text{minimize}} \quad & \sum_{k=1}^{p} \alpha_k \mathbf{1}_n^T \xi_k^+ + (1-\alpha_k)\mathbf{1}_n^T \xi_k^- + \frac{\lambda}{2}||f||_{\mathcal{H}}^2 \\
\text{subject to} \quad & \xi_k^- \succeq 0, \xi_k^+ \succeq 0 \quad \forall k \in [p] \\
& \xi_{i,k}^+ - \xi_{i,k}^- = y_i - f_k(x_i) - b_k \quad \forall k \in [p] \quad \forall i \in [n]
\end{aligned}
\tag{7}
$$

Table of Contents

Optimal Transport Approach

Built a deterministic function $(u,z) \longmapsto Q_{Y|Z}(u,z)$ from $[0,1]^d \times \mathbb{R}^q$ to $\mathbb{R}^d$ where :

- (I) $(u,z) \longmapsto Q_{Y|Z}(u,z)$ being monotone with respect to u, in the sense of being a gradient of a convex function :

$$\left(Q_{Y|Z}(u,z) - Q_{Y|Z}(u',z)\right)^T (u - u') \geq 0 \qquad \forall (u,u') \in [0,1]^d \times [0,1]^d, z \in \mathbb{R}^q \tag{8}$$

- (II) Having with probability one :

$$Y = Q_{Y|Z}(U,Z), \qquad U|Z \sim \mathcal{U}([0,1]^d) \tag{9}$$

Problem to be solved

$$\max_U \{\mathbb{E}[U^T Y] : \ U \sim \mathcal{U}([0,1]^d) \textit{ and } \mathbb{E}[X|U] = \mathbb{E}[X]\} \tag{10}$$

Dual

$$\inf_{(\psi,b)} \mathbb{E}[\psi(X,Y)] + \mathbb{E}[b(U)]^T \mathbb{E}[X] : \psi(x,y) + b(u)^T x \geq u^T y \qquad \forall (y,x,u) \in \mathbb{R}^d \times \mathbb{R}^q \times \mathbb{R}^d \tag{11}$$

Solution of Dual Gives

$$\forall (u,x) \in \mathbb{R}^d \times \mathbb{R}^q, \qquad \beta_0(u)^T x = \nabla_u (b^*(u)^T x) \tag{12}$$

Discretization

$D_n = \{(Y_1, Z_1), ..., (Y_n, Z_n)\}$ and m points $(U_i)_{i \in [\![1,m]\!]}$ of $[0.1]^d$ spaced evenly.
Discrete form of our transportation problem:

$$\max_{P \succeq 0} \sum_{i,j} P_{i,j} Y_j^T U_i \qquad s.t. \qquad P^T \mathbf{1}_m = \nu[\psi], \; PX = \mu \nu^T X[b] \tag{13}$$

where the square brackets indicate the associated Lagrange multiplier.

To find :

$$\widehat{b^*} = \begin{pmatrix} b^*(U_1) \\ \vdots \\ b^*(U_m) \end{pmatrix} = \begin{pmatrix} b_1^*(U_1) \dots b_q^*(U_1) \\ \vdots \\ b_1^*(U_m) \dots b_q^*(U_m) \end{pmatrix} \tag{14}$$

Computation

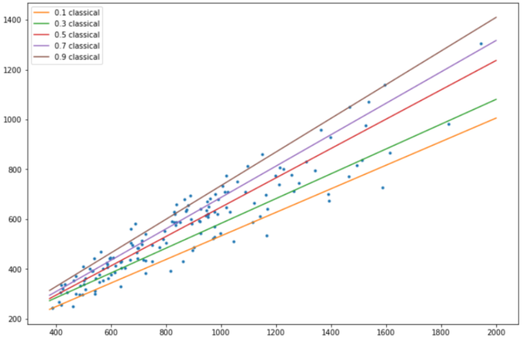
$$\beta_0(u) = \nabla b^*(u) \approx \Big(\frac{b_j^*(u^{(i)} + \epsilon, u^{-(i)}) - b_j^*(u^{(i)}, u^{-(i)})}{\epsilon}\Big)_{i \in [\![1,d]\!], j \in [\![1,q]\!]} \quad (15)$$

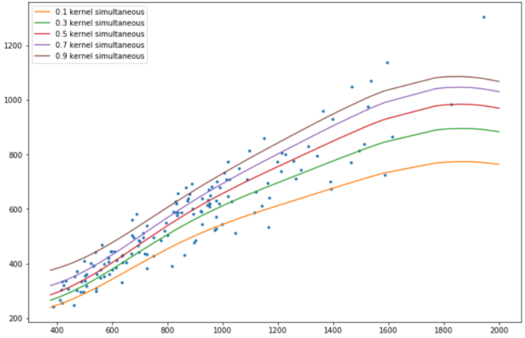
where $u = (u^{(1)}, ..., u^{(d)})$ and $\epsilon > 0$

$$\forall i \in [\![1, m]\!], \widehat{\beta}(U_i) := \Big(\frac{b_j^*(U_i^{(n:k)}) - b_j^*(U_i)}{\epsilon}\Big)_{k \in [\![1,d]\!], j \in [\![1,q]\!]} \quad (16)$$

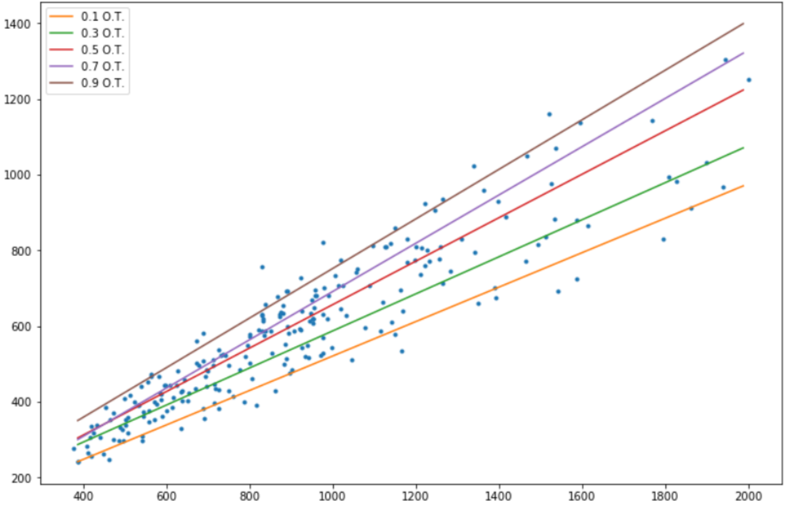
Table of Contents

Implementations

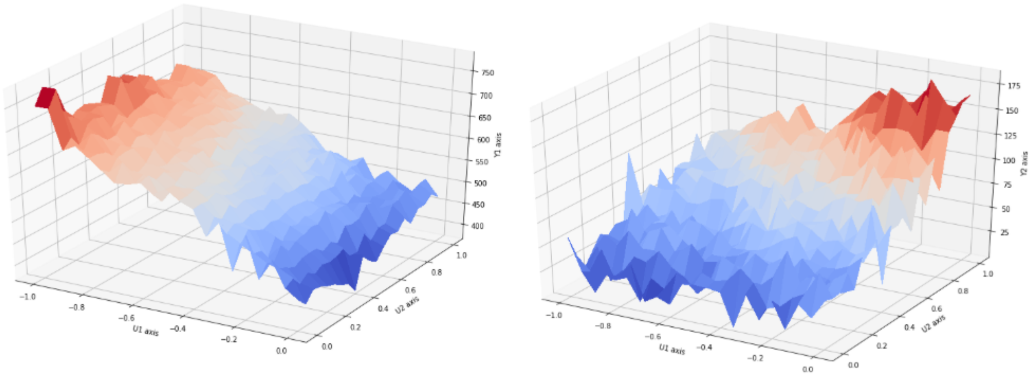




Optimal Transport Implementation



Two dimensional Case



Conclusion

- Kernel, Optimal Transport and Vector Quantile Regression research perspectives
- Interpretability, Tests and confidence intervals
- Overcome the dimension curse

Thank You

for Listening.