

Computer Lab: Quantile Regression

Convex analysis, monotone operators and optimization

Olivier Fercoq olivier.fercoq@telecom-paristech.fr

12 December 2019

You can choose any programming language and work either alone or in pairs. Please send your code and answers to the questions to olivier.fercoq@telecom-paristech.fr before Wednesday, December 18th.

1 Data

We will be using the census dataset for this computer lab. Please download the dataset and helper file on

https://perso.telecom-paristech.fr/ofercoq/tp_qr/.

2 Quantile regression with linear kernels

For $\tau \in (0, 1)$, let us consider the pinball loss defined as $L_\tau(v) = \max\{-(1 - \tau)v, \tau v\}$.

Question 2.1

Calculate L_τ^* , $\text{prox}_{\gamma L_\tau^*}(v)$ and $\text{prox}_{\gamma L_\tau}(v)$ for $\gamma > 0$.

The quantile regression problem consists in estimating conditional quantiles. Given a pair of random variables $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ and a number $\tau \in [0, 1]$, our goal is to estimate the conditional quantile function

$$\mu_\tau(x) = \inf\{\mu \in \mathbb{R} : \mathbb{P}(Y \leq \mu \mid X = x) \geq \tau\}.$$

Given a training set $\{(x_i, y_i)\}_{0 \leq i \leq n-1}$, we estimate the conditional quantile using the solution of the following optimization problem:

$$\min_{w \in \mathbb{R}^d, w_0 \in \mathbb{R}} \frac{\alpha}{2} \sum_{j=1}^d w_j^2 + \sum_{i=1}^n L_\tau(y_i - \sum_{j=1}^d x_{i,j} w_j - w_0) \quad (1)$$

where $\alpha > 0$ is a regularization constant and set $\hat{\mu}_\tau(x) = \sum_{j=1}^d x_j w_j - w_0$. In the rest of the lab, we shall take $\alpha = 1$.

Question 2.2

Define $g : (w, w_0) \mapsto \frac{\alpha}{2} \|w\|^2$. Calculate $\text{prox}_{\gamma g}((w, w_0))$.