

VARYING QUANTILE REGRESSION WITH ONLINE SCHEME AND UNBOUNDED SAMPLING

WANG Bao-bin¹, YIN Hong²

(1. School of Mathematics and Statistics, Central South University for Nationalities,
Wuhan 430074, China)

(2. School of Information, Renmin University of China, Beijing 100872, China)

Abstract: We consider a kernel-based online quantile regression algorithm associated with a sequence of insensitive pinball loss functions. By iteration method and comparison theorem, we obtain the error bound based on the more general output space.

Keywords: quantile regression; Pinball loss; reproducing kernel Hilbert space; online algorithm

2010 MR Subject Classification: 60G15; 62H05

Document code: A **Article ID:** 0255-7797(2014)02-0281-06

1 Introduction

We consider a kernel-based online quantile regression problem when the sampling process is unbounded. Let X be an input space, $Y = R$ be an output space and denote $Z = X \times Y$. Learning algorithm is based on samples $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^T \in Z^T$, T is the sample size, which are drawn independently from the Borel measure ρ . In previous research, a lot of work was done when the learning scheme is involved with the least square loss $\phi_l(u) = u^2, u \in R$. The corresponding target function is the regression function $f_\rho : X \rightarrow Y$ by

$$f_\rho(x) = \int_Y y d\rho_x(y), x \in X,$$

where $\rho_x(\cdot)$ is the conditional distribution of ρ at each $x \in X$. However, algorithms with the least square will lose robustness if the distribution of the noise has heavy tail or abnormal variance. In 1964, Huber proposed the least modulus method instead of the least square, which weakened the noise condition. For sparsity, Vapnik [5] combined the least modulus method with the threshold value and introduced the ε -insensitive loss $\phi^\varepsilon(u) = (|u| - \varepsilon)_+, u \in$

* Received date: 2013-09-22

Accepted date: 2013-11-07

Foundation item: Supported by the Special Fund of Basic Scientific Research of Central Colleges (CZQ13015) and the Teaching Research Fund of South-Central University for Nationalities (JYX13023).

Biography: Wang Baobin(1976–), male, born at Macheng, Hubei, Ph.D., associated professor, major in stochastic analysis and its applications. E-mail:wbb1818@126.com.

R . It is a special case of the pinball loss $\phi_\tau(u) : R \rightarrow R_+$ with $\varepsilon = 0$,

$$\phi_\tau(u) = \begin{cases} (1 - \tau)u, & \text{if } u > 0, \\ -\tau u, & \text{if } u \leq 0. \end{cases}$$

Then the approximation target is a quantile regression function $f_{\rho, \tau}(x)$, which is the τ -quantile of the conditional distribution ρ_x at each $x \in X$. Here a τ -quantile of ρ_x means that there exists a value $W \in R$ satisfying

$$\rho_x(\{y \in Y : y \leq W\}) \geq \tau, \quad \rho_x(\{y \in Y : y \geq W\}) \leq 1 - \tau.$$

Steinwart and Christmann [1] conducted the error analysis of pinball loss under some noise condition. Furthermore, Xiang et al. [2] investigated the learning ability of ε -insensitive pinball loss

$$\phi_\tau^\varepsilon(u) = \begin{cases} (1 - \tau)(u - \varepsilon), & \text{if } u > \varepsilon, \\ -\tau(u + \varepsilon), & \text{if } u \leq -\varepsilon, \\ 0, & \text{otherwise} \end{cases}$$

in regularization schemes for sparsity and robustness.

In this paper, we shall associate pinball loss with the online algorithm in the reproducing kernel Hilbert space $\mathcal{H}_K$. Define a Mercer kernel $K : X \times X \rightarrow R$, which is continuous, symmetric and semi-definite, $\mathcal{H}_K$ is the completion of linear span of the function set $\{K_x = K(x, \cdot), x \in X\}$ with the inner product $\langle \cdot, \cdot \rangle_K$ satisfying $\langle K_x, K_y \rangle_K = K(x, y)$. Here we consider the varying pinball loss $\phi^{(t)}(u) = \phi_{\tau_t}(u)$, where the quantile parameter τ_t changes with the learning time t and converges to the quantity τ as t goes to infinity. One point of the paper is to observe the role of τ in the following algorithm.

Definition 1 The online quantile regression algorithm is defined by $f_1 = 0$ and

$$f_{t+1} = f_t - \eta_t \left\{ (\phi^{(t)})'_-(f_t(x_t) - y_t) K_{x_t} + \lambda_t f_t \right\}, \quad t = 1, 2, \dots, \quad (1.1)$$

where $\lambda_t > 0$ is a regularization parameter, $\eta_t > 0$ is a step size and $(\phi^{(t)})'_-$ is the left (one-side) derivative of $\phi^{(t)}$. From the formula of $\phi^{(t)}$, we see that the learning sequence f_t with the varying quantile can be expressed as

$$f_{t+1} = \begin{cases} (1 - \lambda_t \eta_t) f_t - (1 - \tau_t) \eta_t K_{x_t}, & \text{if } f_t(x_t) - y_t > 0, \\ (1 - \lambda_t \eta_t) f_t + \tau_t \eta_t K_{x_t}, & \text{if } f_t(x_t) - y_t \leq 0. \end{cases}$$

In error analysis, the parameters λ_t, η_t are adapted to accelerate the learning rate. The second point of our paper is to abandon the boundness of the output value y . Following the framework of [3], the moment hypothesis is exploited.

Moment Hypothesis There exists constants $M \geq 1$ and $C > 0$ such that

$$\int_Y |y|^l d\rho_x \leq C l! M^l, \quad \forall l \in N, x \in X.$$

From the above, one can find that it's a generalization of boundness assumption. The main purpose of this paper is to state the error bound $\|f_{T+1} - f_{\rho,\tau}\|$, where the learning time T is large enough.

2 Main Result and Error Analysis

From the quantile regression problem, we can estimate the learning performance $\{f_t\}$ defined in (1.1) by the excess generalization error $\varepsilon(f_t) - \varepsilon(f_{\rho,\tau})$. Here the generalization error $\varepsilon(f)$ with a function $f : X \rightarrow Y$ and the pinball loss $\phi_\tau(u)$ is defined by

$$\varepsilon(f) = \int_Z \phi_\tau(u)(f(x) - y) d\rho.$$

In the following, we denote the generalization error $\varepsilon^{(t)}(f)$ with the varying τ_t as

$$\varepsilon^{(t)}(f) = \int_Z \phi_{\tau_t}(u)(f(x) - y) d\rho.$$

The relation between the excess generalization error $\varepsilon(f) - \varepsilon(f_{\rho,\tau})$ and the error bound $\|f - f_{\rho,\tau}\|_{L_{\rho_X}^r}$ is explained by the following comparison theorem, which was given in [1].

Definition 2 Let $0 < \varphi \leq \infty$ and $\xi > 1$. Denote $r = \varphi\xi/(\varphi + 1) > 0$. We say that ρ has a τ -quantile of φ -average type ξ if there exist two positive functions ω_τ and b_τ on X such that $\{b_\tau\omega_\tau^{\xi-1}\}^{-1} \in L_{\rho_X}^\varphi$ and for any $x \in X$ and $\omega \in (0, \omega_\tau(x)]$, there hold

$$\rho_X(\{y : f_{\rho,\tau}(x) < y < f_{\rho,\tau}(x) + \omega\}) \geq b_\tau(x)\omega^{\xi-1}$$

and

$$\rho_X(\{y : f_{\rho,\tau}(x) - \omega < y < f_{\rho,\tau}(x)\}) \geq b_\tau(x)\omega^{\xi-1}.$$

In our analysis we shall make use of the following comparison theorem.

Lemma 1 Let $0 < \varphi \leq \infty$ and $\xi > 1$. Denote $r = \varphi\xi/(\varphi + 1) > 0$. If the measure ρ has a τ -quantile of φ -average type ξ , then for any measurable function f on X , we have

$$\|f - f_{\rho,\tau}\|_{L_{\rho_X}^r} \leq 2^{1-1/\xi} \xi^{1/\xi} \|\{b_\tau\omega_\tau^{\xi-1}\}^{-1}\|_{L_{\rho_X}^\varphi}^{1/\xi} \{\varepsilon(f) - \varepsilon(f_{\rho,\tau})\}^{1/\xi}. \quad (2.1)$$

For the error analysis, we need the approximate error $D(\lambda)$ with (p, K, τ) is defined by

$$D(\lambda) = \inf_{f \in H_K} \{\varepsilon(f) - \varepsilon(f_{\rho,\tau}) + \frac{\lambda}{2} \|f\|_K^2\}, \lambda > 0. \quad (2.2)$$

A minimization f_λ of (2.2) is called the regularization function. Throughout the paper, we assume that the error bound of $D(\lambda)$ satisfies

$$D(\lambda) \leq D_0 \lambda^\gamma, \forall \lambda > 0 \quad (2.3)$$

with the constant $D_0 > 0$ and the index power $0 < \gamma \leq 1$. Now we can present our main theorem.

Theorem 1 Assume the measure ρ has a τ -quantile of φ -average type ξ and the approximate error (2.3) holds. Take the parameters η_t, λ_t as the form

$$\eta_t = \eta_1 t^{-\alpha}, \lambda_t = \lambda_1 t^{-\beta}, \eta_1, \lambda_1 > 0$$

with

$$0 < \beta < \frac{2}{5-\gamma}, \beta < \alpha < \frac{2-3\beta+\beta\gamma}{2}. \quad (2.4)$$

Then we get

$$\mathbb{E}_{z_1, \dots, z_T} \|f_{T+1} - f_{\rho, \tau}\|_{L_{\rho_X}^r} = O(T^{-\Lambda}), \quad (2.5)$$

where Λ is given by

$$\Lambda = \min\left\{\frac{\beta\gamma}{\xi}, \frac{\alpha-\beta}{2\xi}, \frac{2+\beta\gamma-3\beta-2\alpha}{2\xi}\right\} > 0.$$

3 Proof of Main Theorem

We are in a position to present the key analysis in our study. We shall consider the sample error $\|f_{T+1} - f_{\lambda_T}\|_K$.

Lemma 2 Define the sequence f_t by (1.1). Let the parameters η_t, λ_t be the same form in Theorem 1 and the error bound (2.3) holds. Then

$$\mathbb{E}_{z_1, \dots, z_T} \|f_{T+1} - f_{\lambda_T}\|_K^2 \leq C' T^{-\theta^*}, \quad (3.1)$$

where $\theta^* = \min\{\alpha - \beta, 2 + \beta\gamma - 3\beta - 2\alpha\}$ and C' is a constant independent of T .

Proof By induction, we assume that $\|f_t\|_K \leq \frac{\kappa}{\lambda_t}$, $t \in N$ and $0 < \tau_t < 1$, $t \in N$. From the definition of f_t , we see that

$$\|f_{t+1}\|_K \leq (1 - \lambda_t \eta_t) \|f_t\|_K + \eta_t \kappa \leq (1 - \lambda_t \eta_t) \frac{\kappa}{\lambda_t} + \eta_t \kappa = \frac{\kappa}{\lambda_t} \leq \frac{\kappa}{\lambda_{t+1}}.$$

Denote $B_t = (\phi^{(t)})'_-(f_t(x_t) - y_t)K_{x_t} + \lambda_t f_t$. Then $\|B_t\|_K \leq 2\kappa$ by the fact that $\|(\phi^{(t)})'_-\|_\infty \leq 1$.

From definition (1.1), we see by inner products that

$$\|f_{t+1} - f_{\lambda_t}\|_K^2 = \|f_t - f_{\lambda_t}\|_K^2 + 2\eta_t \langle f_{\lambda_t} - f_t, B_t \rangle_K + \eta_t^2 \|B_t\|_K^2. \quad (3.2)$$

For the second term, the reproducing property and the convexity of the loss function ϕ_t tell us that

$$\begin{aligned} \langle f_t - f_t, B_t \rangle_K &= (\phi^{(t)})'_-(f_t(x_t) - y_t) \{f_{\lambda_t}(x_t) - f_t(x_t)\} + \lambda_t \langle f_{\lambda_t} - f_t, f_t \rangle_K \\ &\leq \phi^{(t)}(f_{\lambda_t}(x_t) - y_t) - \phi^{(t)}(f_t(x_t) - y_t) + \lambda_t \langle f_{\lambda_t} - f_t, f_t \rangle_K \\ &\leq \phi^{(t)}(f_{\lambda_t}(x_t) - y_t) - \phi^{(t)}(f_t(x_t) - y_t) + \frac{\lambda_t}{2} \|f_{\lambda_t}\|_K^2 - \frac{\lambda_t}{2} \|f_t\|_K^2. \end{aligned}$$

Thus, taking expectation with respect to z_t , we find

$$\begin{aligned}\mathbb{E}_{z_t}\langle f_t - f_t, B_t \rangle_K &\leq [\mathcal{E}^{(t)}(f_{\lambda_t}) + \frac{\lambda_t}{2}\|f_{\lambda_t}\|_K^2] - [\mathcal{E}^{(t)}(f_t) + \frac{\lambda_t}{2}\|f_t\|_K^2] \\ &\leq -\frac{\lambda_t}{2}\|f_t - f_{\lambda_t}\|_K^2.\end{aligned}$$

Therefore, together with the bound for B_t and (3.2), we have

$$\mathbb{E}_{z_t}\|f_{t+1} - f_{\lambda_t}\|_K^2 \leq (1 - \lambda_t\eta_t)\|f_t - f_{\lambda_t}\|_K^2 + 4\kappa^2\eta_t^2. \quad (3.3)$$

We decompose $\|f_t - f_{\lambda_t}\|_K$ as $\|f_t - f_{\lambda_{t-1}}\|_K$ and the drift error $d_t = \|f_{\lambda_{t-1}} - f_{\lambda_t}\|_K$. On one hand,

$$d_t \leq \frac{1}{2}\left(\frac{\lambda_{t-1}}{\lambda_t} - 1\right)(\|f_{\lambda_t}\|_K + \|f_{\lambda_{t-1}}\|_K)$$

by the lemma in [4]. Noting the fact that $\|f_{\lambda}\|_K \leq \sqrt{\frac{D(\lambda)}{\lambda}}$ for each $\lambda > 0$. Then $d_t \leq d_0 t^{-(1-\frac{\beta}{2}+\frac{\beta\gamma}{2})}$, where $d_0 = \beta 2^{\beta+1} \sqrt{2D_0\lambda_1^\gamma/\lambda_1}$. Using the elementary inequality $2ab \leq Aa^2b^q + b^{2-q}/A$ with $0 < q < 2, A > 0$ to the case of $a = \|f_t - f_{\lambda_{t-1}}\|_K, b = d_t$, we obtain

$$\|f_t - f_{\lambda_t}\|_K^2 \leq \|f_t - f_{\lambda_{t-1}}\|_K^2 + A\|f_t - f_{\lambda_{t-1}}\|_K^2 d_t^q + d_t^{2-q}/A + d_t^2.$$

By the above estimats, we get

$$\mathbb{E}_{z_t}\|f_{t+1} - f_{\lambda_t}\|_K^2 \leq (1 + Ad_t^q - \lambda_t\eta_t)\|f_t - f_{\lambda_{t-1}}\|_K^2 + \frac{d_t^{2-q}}{A} + d_t^2 + 4\kappa^2\eta_t^2.$$

We take the constant $A = \frac{\frac{1}{2}\eta_1\lambda_1}{d_1^q}$ and $q = \frac{\alpha+\beta}{1-\frac{\beta}{2}+\frac{\beta\gamma}{2}}$, the restriction of the index power α, β implies that

$$\mathbb{E}_{z_t}\|f_{t+1} - f_{\lambda_t}\|_K^2 \leq \left(1 - \frac{\eta_1\lambda_1}{2}t^{-\alpha-\beta}\right)\|f_t - f_{\lambda_{t-1}}\|_K^2 + Ct^{-\theta}, \quad (3.4)$$

where $\theta = \min\{2 + \beta\gamma - 2\beta - \alpha, 2\alpha\}$ and $C = d_0^{2-q}/A + d_0^2 + 4\kappa^2\eta_1^2 > 0$.

Applying relation (3.4) iteratively for $t = 1, \dots, T$, we have that

$$\mathbb{E}_{z_1, \dots, z_T}\|f_{T+1} - f_{\lambda_T}\|_K^2 \leq C \sum_{t=1}^T \prod_{j=t-1}^T \left(1 - \frac{\eta_1\lambda_1}{2}j^{-\alpha-\beta}\right) t^{-\theta}. \quad (3.5)$$

Applying the following elementary inequality [4] with $0 < a_1 < 1, c, a_2 > 0$ and $t \in \mathbb{N}$,

$$\sum_{i=1}^{t-1} i^{-a_2} \exp\left\{-c \sum_{j=i+1}^t j^{-a_1}\right\} \leq \left\{\frac{2^{a_1+a_2}}{c} + \left(\frac{1+a_2}{ec(1-2^{a_1-1})}\right)^{(1+a_2)/(1+a_1)}\right\} t^{a_1-a_2}$$

to the case of $a_1 = \alpha + \beta < 1, a_2 = \theta$ and $c = \frac{\eta_1\lambda_1}{2}$, we get the desired result (3.1).

Proof of Theorem 1 By the decomposition in [6],

$$\varepsilon(f_{T+1}) - \varepsilon(f_{\rho, \tau}) \leq \varepsilon(f_{T+1}) - \varepsilon(f_{\lambda_T}) + D(\lambda_T).$$

Since ϕ is uniformly continuous, then

$$|\varepsilon(f_{T+1}) - \varepsilon(f_{\lambda_T})| \leq \|f_{T+1} - f_{\lambda_T}\|_{\infty} \leq \|f_{T+1} - f_{\lambda_T}\|_K.$$

The above bound together with the comparison theorem (2.1) and Lemma 2 give our conclusion (2.5).

References

- [1] Steinwart I, Christmann A. Estimating conditional quantiles with the help of the pinball loss [J]. Bernoulli, 2011, 17: 211–225.
- [2] Xiang D H, Hu T, Zhou D X. Learning with varying insensitive loss [J]. Appl. Math. Letters, 2011, 24: 2107–2109.
- [3] Guo Z C, Wang C. Online regression with unbound sampling [J]. Int. J. Comp. Approx., 2011, 14: 2936–2941.
- [4] Ye G B, Zhou D X. Fully online classification by regularization [J]. Appl. Comput. Harmonic Anal., 2007, 23: 198–214.
- [5] Vapnik V. Statistical learning theory[M]. New York: Wiley, 1998.
- [6] Ying Y, Zhou D X. Online regularized classification algorithms[J]. IEEE Trans. Inform. Theory, 2006, 52: 4775–4788.

无界意义下的在线变化分位数回归算法

汪宝彬¹, 殷红²

(1. 中南民族大学数学与统计学学院, 湖北 武汉 430074)

(2. 中国人民大学信息学院, 北京 100872)

摘要: 本文研究了基于核方法下的在线变化损失函数的回归算法. 利用迭代和比较原则, 得到了算法的收敛速度, 并将该结果推广到了更一般的输出空间.

关键词: 分位数回归; Pinball 损失函数; 再生核希尔伯特空间; 在线算法

MR(2010)主题分类号: 60G15; 62H05 中图分类号: O211.6