mathematics

MDPI

Article

Supplementary: Robust Online Support Vector Regression with Truncated ε-Insensitive Pinball Loss

Xian Shan *, Zheshuo Zhang, Xiaoying Li, Yu Xie and Jinyu You

College of Science, China University of Petroleum, Qingdao 266580, China
* Correspondence: 20120029@upc.edu.cn

Definition 1. [Strong pseudo-convexity] A function $f : \chi \to R$ is said to be strongly pseudo-convex (SPC) on $\chi_1 \subset \chi$ with respect to $\overline{x} \in \chi$, if

$$f(x) - f(\overline{x}) \le K\langle f'(x), x - \overline{x}\rangle \tag{1}$$

holds for all $x \in \chi_1$, with $f'(x)$ a Clarke subgradient of f at $x, K > 0$ is a contant. If the Inequality (1) holds with respect to any $\overline{x} \in \chi_1$, f is called SPC on $\chi_1 \subset \chi$. The collection of SPC functions on χ_1 with $K > 0$ are denoted as $\mathcal{W}_K(\chi_1)$.

Proposition 1. Let $f : R \to R$ be a univariate continuous function. Assume that on each interval of $(-\infty, a], (a, b), [b, \infty), f(x)$ is convex, and $f'_-(a) < 0, f'_+(b) > 0, f'_+(a) \neq 0, f'_-(b) \neq 0$ .Then we have that the Inequality (1) holds for any fixed $\overline{x} \in R$ and $x \in R$ with

$$K = \max\left\{1, \frac{f'_-(a)}{f'_+(a)}, \frac{f'_+(a)}{f'_-(a)}, \frac{f'_-(b)}{f'_+(b)}, \frac{f'_+(b)}{f'_-(b)}, \frac{f'_-(a)}{f'_+(b)}, \frac{f'_+(b)}{f'_-(a)}\right\} \tag{2}$$

Proposition 2. Let $f : R \to R$ be a univariate continuous function. Let $a_0 < a_1 < \cdots < a_m$ be the real numbers, $a_0 = -\infty$ and $a_m = +\infty$. On each interval of $[a_i, a_{i+1}]$,$f(x)$ is convex, and $i = 0, 1, .., m-1$. Let S be the set of the minimum points of f on R. Suppose that the optimal solution set $S \in [a_q, a_{q+1}]$. With $q \in [0, \cdots, m-1]$. Moreover, suppose that $f(x)$ is strictly decreasing when $X \le InfS$ and strictly increasing when $X \ge SupS$. Then, for any fixed $\overline{x} \in [a_0, a_m]$ and $x \in [a_0, a_m]$. Inequality (1) holds with

$$\begin{aligned} K = \max\Bigg\{1, \frac{f'_+(a_\mu)}{f'_-(a_{v+1})}, \frac{f'_-(a_i)}{f'_+(a_j)} \mid q \in [0, \cdots, m-1], \mu \in [v+1, \cdots, q] \\ v \in [0, \cdots, q-1], i \in [q+1, \cdots, j], j \in [q+1, \cdots, m-1]\Bigg\} \end{aligned} \tag{3}$$

Lemma 1. Denote $\Omega_0 = [t_0 - \delta/\tau - C, t_0 - \delta/\tau] \cup [t_0 + \delta, t_0 + \delta + C]$, suppose that $0 \notin \Omega_0, f(t) = \frac{1}{2}t^2 + C \cdot L_{TIP}((t - t_0))$ is SPC on $R \backslash \Omega_0$ with $K = \max\left\{2, 1 + \frac{C\cdot\tau^2}{\delta}, 1 + \frac{C\cdot\tau}{\delta\cdot\tau-\varepsilon}\right\}$.

Proof:

Suppose that $t_0 \le 0$

$$C \cdot L'_{TIP}(t - t_0) = \begin{cases} -C\cdot\tau, & -\delta/\tau + t_0 \le t < -\varepsilon/\tau + t_0 \\ 0, & otherwise \\ C, & \varepsilon + t_0 \le t < \delta + t_0 \end{cases} \tag{4}$$

$$\left(\frac{1}{2}t^2\right)' = t \tag{5}$$

Let $a = t_0 - \delta/\tau, b = t_0 + \delta, \Omega_1 = (-\infty, a), \Omega_2 = [a, b], \Omega_3 = (b, \infty), \Omega_0 = [a - C, a] \cup [b, b + C]$.

Case 1: $t_0 \in [-\varepsilon/\tau, 0]$

We can see that $f(t)$ is continuous on R and convex on every internal of Ω_i, i = 1,2,3. The directional derivatives of f at point a are calculated as follows:

$$f'_{-}(a) = a; f'_{+}(a) = a - C \cdot \tau \tag{6}$$

As $a < 0$, the above two directional derivatives are negative. The directional derivatives of f at point b are calculated as follows:

$$f'_{-}(b) = b + C; f'_{+}(b) = b \tag{7}$$

For $\delta > \frac{\varepsilon}{\tau}$, $b > 0$, both of the above two directional derivatives are positive. By Proposition 1, we have $f(t)$ is strongly pseudoconvex in R for any fixed $\bar{t} \in R$, where

$$K = max\{1, \frac{a - C \cdot \tau}{a}, \frac{b + C}{b}\} \leq max\{1 + \frac{C \cdot \tau^2}{\delta}, 1 + \frac{C \cdot \tau}{\tau \cdot \delta - \varepsilon}\}. \tag{8}$$

Case 2: $t_0 \in [-\delta, -\varepsilon/\tau]$

When $f'_{-}(a) = a < 0$, $f'_{+}(b) = b > 0$, $t^* \in argminf \in \Omega_2$ and $f(t)$ is continuous on R and convex on every internal of $\Omega_i, i = 1, 2, 3$. The directional derivative of f at a and b are calculated in the same way with case 1. By Proposition 1, we have $f(t)$ is strongly pseudoconvex in R for any fixed $\bar{t} \in R$, where

$$K = max\{1, \frac{a - C \cdot \tau}{a}, \frac{b + C}{b}\} \leq max\{1 + \frac{C \cdot \tau^2}{\varepsilon + \delta}, 1 + \frac{C}{b}\}. \tag{9}$$

The last equation is $0 < b < \delta - \varepsilon/\tau$. Specially, let's consider the strong pseudoconvexity of $f(t)$ on $R \backslash \Omega_0$. Define function $f(t) : R \to R$ as follows:

$$\widetilde{f} = \begin{cases} f(t), & t \in R \backslash (b, b + C) \\ f(b) + \frac{f(b+C) - f(b)}{C} \cdot t, & t \in [b, b + C] \end{cases} \tag{10}$$

Then $\widetilde{f}(t)$ is a continuous function of one variable and convex on every internal of Ω_1, Ω_2, $[b, b + C]$ and $[b + C, \infty]$. Its directional derivatives at a are the same as the directional derivatives of f calculated in Equation (6). Its directional derivatives at b and $b + C$ are calculated as

$$\widetilde{f}'_{-}(b) = b + C; \widetilde{f}'_{+}(b + C) = b + C \tag{11}$$

$$\widetilde{f}'_{-}(b + C) = b + \frac{1}{2}C; \widetilde{f}'_{+}(b) = b + \frac{1}{2}C \tag{12}$$

According to Proposition 1, we have $\widetilde{f}(t) \in \chi_K(R)$, where

$$\begin{aligned} K &= max\{1, \frac{f'_{+}(v)}{f'_{-}(v)}, \frac{f'_{-}(v)}{f'_{+}(v)}, \frac{f'_{-}(b)}{f'_{+}(b + C)} \mid v \in \{a, b, b + C\}\} \\ &= max\{1, \frac{a - C \cdot \tau}{a}, \frac{b + C}{b + \frac{1}{2}C}, \frac{b + C}{b + C}\} \\ &\leq max\{2, 1 + \frac{C \cdot \tau^2}{\delta + \varepsilon}\} \end{aligned} \tag{13}$$

The last two inequalities are true because $a < -\delta - \varepsilon/\tau < 0, b > 0$. Since $\widetilde{f}(t)$ and $f(t)$ have the same definition on $R \backslash \Omega_0$, we have $f(t) \in \chi_K(R \backslash \Omega_0)$, K is defined by the above equation.

Case 3: $t_0 \in (-\infty, -\delta)$

Consider that $t \in [t_0 + \varepsilon, t_0 + \delta]$. In this interval, $f(t) = C \cdot [(t - t_0) - \varepsilon] + \frac{1}{2}t^2$. Then $f'(t) = C + t$ is increasing with t. The assumption of the lemma implies that $t_0 + \delta < -C$, then $\forall t \in [t_0 + \varepsilon, t_0 + \delta]$, $f'(t) < 0$ for $t \in \Omega_1 \cup \Omega_2$. In this case, $f(t)$ is convex on each interval of $\Omega_i, i = 1, 2, 3$, and there exists a unique minimizer $t^* \in \Omega$ of $f(t)$. Moreover, $f'(t) < 0$ for $t < t^*$ and $f'(t) > 0$ for $t > t^*$. By Proposition 1, we have that $f(t)$ is strongly pseudo-convex on R with

$$\begin{aligned}
K &= max\left\{1, \frac{f'_+(v)}{f'_-(v)}, \frac{f'_-(v)}{f'_+(v)}, \frac{f'_-(a)}{f'_+(b)}, \frac{f'_+(b)}{f'_-(a)} \mid v \in \{a, b\}\right\} \\
&= max\left\{1, \frac{a - C \cdot \tau}{a}, \frac{b}{b + C}, \frac{a}{b}, \frac{b}{a}\right\} \\
&\leq max\left\{1 + \frac{C \cdot \tau^2}{\delta \cdot \tau + \delta}, 1 - \frac{C}{b + C}, 1 + \frac{\delta \cdot \tau + \delta}{C \cdot \tau}\right\}
\end{aligned} \tag{14}$$

Particularly, let's consider the strong pseudo-convexity of $f(t)$ on R Ω_0.Consider the function $\widetilde{f}(t) : R \to R$. It is observed that $\widetilde{f}$ is a continuous and piecewise-convex function. Its directional derivations at b, and $b + C$ are calculated as described above. By Proposition 2, we have that $f(t) \in \chi_K(R)$ with

$$\begin{aligned}
K &= \max\left\{1, \frac{f'_+(v)}{f'_-(v)}, \frac{f'_-(v)}{f'_+(v)}, \frac{f'_-(b)}{f'_+(b + C)} \mid v \in \{a, b, b + C\}\right\} \\
&= \max\left\{1, \frac{a - C \cdot \tau}{a}, \frac{b + C}{b + \frac{1}{2}C}, \frac{b + C}{b + C}\right\} \\
&\leq \max\left\{2, 1 + \frac{C \cdot \tau^2}{\delta \cdot \tau + \delta}, 1\right\} \\
&\leq \max\left\{2, 1 + \frac{C \cdot \tau^2}{\delta \cdot \tau + \delta}\right\}
\end{aligned} \tag{15}$$

Likewise, we can acquire the proof regarding $t_0 \geq 0$.

In summary, $K = \max\left\{2, 1 + \frac{C \cdot \tau^2}{\delta}, 1 + \frac{C \cdot \tau}{\delta \cdot \tau - \varepsilon}\right\}$.

Assumption 1: We assume that $X > 0$ exists such that $K(\mathbf{X}_t, \mathbf{X}_t) \leq X^2$ for all t. The following proportion could be obtained according to the assumption:

(1) For each $f \in \mathcal{H}$, $|f(\mathbf{X}_t)| = |\langle f, K(\mathbf{X}_t, -)\rangle| \leq X \cdot \|f\|_{\mathcal{H}}$.

(2) For given that $L'_{TPIL} < 1$, we have $\left\|\partial_f L_{TIP}(f(\mathbf{X}_t) - y_t)\right\|_{\mathcal{H}} \leq \frac{X}{C}$ and $\left\|\partial_f R_{inst}[f, \mathbf{X}_t, y_t]\right\|_{\mathcal{H}} \leq CX + \|f\|_{\mathcal{H}} \leq 2X$ for any $f : ||f||_{\mathcal{H}} \leq X$.

(3) For fixed $C > 0$ and $t = 1, 2, \cdot\cdot, f^t$ generated by TIPOSVR satisfied $f : ||f||_{\mathcal{H}} \leq X$.

Proof:

$\|f^{t+1}\|_{\mathcal{H}} \leq \|(1 - \gamma_t)f^t - C \cdot \eta_t L'_{TIP}(f(\mathbf{X}_t) - y_t)K(\mathbf{X}_t, \cdot)\|_{\mathcal{H}} \leq (1 - \gamma_t)\|f^t\|_{\mathcal{H}}$
$+ \gamma_t X$

Given $f^0 = 0$,we have $||f||_{\mathcal{H}} \leq X$.

Lemma 2: Let the sequence instance $(\mathbf{X}_t, y_t)$ satisfy $k(\mathbf{X}_t, \mathbf{X}_t) \leq X^2$. For a fixed $g \in \mathcal{H}$

$$u^t = (f^t - g)/\|f^t - g\|, t_0 = y_t - g(\mathbf{X}_t) + u^t(\mathbf{X}_t) \cdot \langle u^t, g\rangle,$$

$$\Omega_0 = \left[-\delta/\tau - \left(u^t(\mathbf{X}_t)\right)^2, -\delta/\tau\right] \cup \left[\delta, \delta + \left(u^t(\mathbf{X}_t)\right)^2\right]$$

Assuming $t_0 \notin \Omega_0, \xi_t = f^t(\mathbf{X}_t) - y_t \notin \Omega_0$, we have

$R_{inst}[f^t, \mathbf{X}_t, y_t] - R_{inst}[g, \mathbf{X}_t, y_t] \leq K \cdot \left\langle \partial_f R_{inst}[f^t, \mathbf{X}_t, y_t]\Big|_{f=f^t}, f^t - g\right\rangle_{\mathcal{H}}$ with

$$K = \max\left\{2, 1 + \frac{C \cdot \tau^2 X^2}{\delta}, 1 + \frac{C \cdot \tau X^2}{\delta \cdot \tau - \varepsilon}\right\}. \tag{16}$$

Proof:

Let $f = g + \alpha u^t, \|u^t\| = 1$.
According to the definition of R_{inst} :
$R_{inst}[f, \mathbf{X}_t, y_t] = C \cdot L_{TIP}[g(\mathbf{X}_t) + \alpha u^t(\mathbf{X}_t) - y_k] + \frac{1}{2}\|g + \alpha u^t\|^2 +$ constant,
Let $\xi = g(\mathbf{X}_t) + \alpha u^t(\mathbf{X}_t) - y_t \in R, t_0 = y_t - g(\mathbf{X}_t) + u^t(\mathbf{X}_t)(u^t, g)$,

$$\xi + t_0 = \alpha u^t(\mathbf{X}_t) + u^t(\mathbf{X}_t)(u^t, g) = u^t(\mathbf{X}_t)\left[\alpha + (u^t, g)\right]$$

With the definition of $\|u^t\| = 1$, $R_{inst}[f, \mathbf{X}_t, y_t]$ satisfies the strong pseudoconvex inequality, i.e.,

$$R_{inst}[f^t, \mathbf{X}_t, y_t] - R_{inst}[g, \mathbf{X}_t, y_t] \le K \cdot \left\langle \partial_f R_{inst}[f^t, \mathbf{X}_t, y_t]\Big|_{f=f^t}, f^t - g \right\rangle_{\mathcal{H}}$$

holds as the univariate function on the right-hand side of as α is strongly pseudoconvex.
We have

$$\phi(\xi) = R_{inst}[f, \mathbf{X}_t, y_t] = C \cdot L_{TIP}[\xi] + \frac{1}{2[u^t(\mathbf{X}_t)]^2}[\xi + t_0]^2 + constant$$

Therefore, according to the above lemma, $\phi(\xi + t_0)$ satisfies strong pseudoconvexity when $\xi \in R \backslash \Omega_0$ with

$$\begin{aligned} K &= \max\left\{2, 1 + \frac{C \cdot \tau^2 [u^t(\mathbf{X}_t)]^2}{\delta}, 1 + \frac{C \cdot \tau [u^t(\mathbf{X}_t)]^2}{\delta \cdot \tau - \varepsilon}\right\} \\ &\le \max\left\{2, 1 + \frac{C \cdot \tau^2 X^2}{\delta}, 1 + \frac{C \cdot \tau X^2}{\delta \cdot \tau - \varepsilon}\right\}. \end{aligned} \tag{17}$$

Theorem 1: Set example sequence $S = \{(\mathbf{X}_t, y_t)\}_{t=0}^T$ be $k(\mathbf{X}_t, \mathbf{X}_t) \le X^2$ holds for all t. $(f^0, \cdots, f^T)$ represents a hypothetical sequence produced by TIPOSVR, $R_{inst}[g, S] = \frac{1}{T}\sum_{t=1}^T R_{inst}[g, \mathbf{X}_t, y_t]$, and $\hat{g} = \arg\min_{g \in \mathcal{H}} R_{inst}[g, S]$. Fixed $C, \varepsilon > 0, 0 < \eta < C$ and set the learning rate $\eta_t = \eta \cdot t^{-1/2}$. We assume that each hypothesis f^t generated by TIPOSVR satisfies the hypothesis stated in Lemma 2, for $t = 0, 1, 2 \cdots T$. And then we have the following expression

$$\frac{1}{T}\sum_{t-1}^{T} R_{inst}[f^t, \mathbf{X}_t, y_t] \le R_{inst}[\hat{g}, S] + \alpha T^{-1/2} + o\left(T^{-1/2}\right) \tag{18}$$

Among them, $\alpha = \frac{2KX^2}{\eta} + 4KX^2\eta$, $K = \max\left\{2, 1 + \frac{C \cdot \tau^2 X^2}{\delta}, 1 + \frac{C \cdot \tau X^2}{\delta \cdot \tau - \varepsilon}\right\}$
Proof:
From the strong pseudo-convexity of R_{inst} in lemma 2 and the Lipschitz property of R_{inst} in the first argument, the following expressions are derived:

$$\begin{aligned} &\|f^t - \hat{g}\|_{\mathcal{H}}^2 - \left\|f^{t+1} - \hat{g}\right\|_{\mathcal{H}}^2 \\ &= -\left\|f^{t+1} - f^t\right\|_{\mathcal{H}}^2 - \left\langle f^{t+1} - f^t, f^{t+1} - \hat{g}\right\rangle_{\mathcal{H}} \\ &= -\eta_t^2 \left\|\partial_f R_{inst}[f^t, \mathbf{X}_t, y_t]\right\|_{\mathcal{H}}^2 + 2\eta_t \left\langle \partial_f R_{inst}[f^t, \mathbf{X}_t, y_t]\Big|_{f=f^t}, f^t - \hat{g}\right\rangle_{\mathcal{H}} \\ &\ge -4\eta_t^2 X^2 - \frac{2\eta_t}{K}\left(R_{inst}[\hat{g}, \mathbf{X}_t, y_t] - R_{inst}[f^t, \mathbf{X}_t, y_t]\right) \end{aligned}$$

Here, $K = \max\left\{2, 1 + \frac{C \cdot \tau^2 X^2}{\delta}, 1 + \frac{C \cdot \tau X^2}{\delta \cdot \tau - \varepsilon}\right\}$,
Then there are

$$
\begin{aligned}
&\frac{1}{\eta_t}\|f^t-\hat{g}\|_{\mathcal{H}}^2-\frac{1}{\eta_{t+1}}\left\|f^{t+1}-\hat{g}\right\|_{\mathcal{H}}^2\\
&=\frac{1}{\eta_t}\left(\|f^t-\hat{g}\|_{\mathcal{H}}^2-\left\|f^{t+1}-\hat{g}\right\|_{\mathcal{H}}^2\right)+\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t+1}}\right)\left\|f^{t+1}-\hat{g}\right\|_{\mathcal{H}}^2\\
&\geq -4\eta_t X^2-\frac{2}{K}\left(R_{inst}[\hat{g},\mathbf{X}_t,y_t]-R_{inst}\left[f^{t+1},\mathbf{X}_t,y_t\right]\right)+\left(\frac{1}{\eta_t}-\frac{1}{\eta_{t+1}}\right)4X^2
\end{aligned}
$$

Considering $\left\|f^{t+1}-\hat{g}\right\|_{\mathcal{H}}\leq 2X,\eta_{T+1}=\eta(T+1)^{-1/2}$, by summing $\sum_{t=1}^{T}\eta_t\leq 2\eta T^{1/2}$, we get the following expression

$$
\begin{aligned}
&\frac{1}{\eta}\left\|f^1-\hat{g}\right\|_{\mathcal{H}}^2-\frac{1}{\eta_{T+1}}\left\|f^{T+1}-\hat{g}\right\|_{\mathcal{H}}^2\\
&\geq\sum_{t=1}^{T}-4\eta_t X^2-\frac{2}{K}\sum_{t=1}^{T}\left(R_{inst}[\hat{g},\mathbf{X}_t,y_t]-R_{inst}\left[f^t,\mathbf{X}_t,y_t\right]\right)+\left(\frac{1}{\eta}-\frac{1}{\eta_{T+1}}\right)4X^2
\end{aligned}
$$

So,

$$
\begin{aligned}
&-\frac{2}{K}\sum_{t=1}^{T}\left(R_{inst}[\hat{g},\mathbf{X}_t,y_t]-R_{inst}\left[f^t,\mathbf{X}_t,y_t\right]\right)\\
&\leq\frac{1}{\eta}\left\|f^1-\hat{g}\right\|_{\mathcal{H}}^2-\frac{1}{\eta_{T+1}}\left\|f^{T+1}-\hat{g}\right\|_{\mathcal{H}}^2-\left(\frac{1}{\eta}-\frac{(T+1)^{1/2}}{\eta}\right)4X^2+8\eta X^2T^{1/2}\\
&\leq\frac{4X^2}{\eta}-\frac{(2X)^2}{\eta_{\tau+1}}+\frac{(T+1)^2-1}{\eta}4X^2+8\eta X^2T^{1/2}\\
&\leq\frac{X^2}{\eta}+\frac{T^{1/2}}{\eta}4X^2+8\eta X^2T^{1/2}
\end{aligned}
$$

And we can get,

$$
\begin{aligned}
&-\frac{1}{T}\sum_{t=1}^{T}\left(R_{inst}[\hat{g},\mathbf{X}_t,y_t]-R_{inst}\left[f^t,\mathbf{X}_t,y_t\right]\right)\\
&\leq\frac{2KX^2}{\eta}T^{-1}+4KX^2\eta T^{-1/2}+\frac{2KX^2}{\eta}T^{-1/2}\\
&=\frac{2KX^2}{\eta}T^{-1}+\alpha T^{-1/2}
\end{aligned}
$$

$\alpha=\frac{2KX^2}{\eta}+4KX^2\eta.$