Własności estymatorów regresji porządkowej z karą LASSO

Wojciech Rejchel

Uniwersytet Mikołaja Kopernika w Toruniu
Uniwersytet Warszawski

Badania sfinansowane ze środków Narodowego Centrum Nauki przyznanych w ramach

finansowania stażu po uzyskaniu stopnia naukowego doktora, DEC-2014/12/S/ST1/00344.

Regresja porządkowa

- $Z = (X, Y)$, $Z' = (X', Y')$ - niezależne wektory losowe o tym samym rozkładzie P
- $X, X' \in \mathcal{X}$, $Y, Y' \in \mathcal{Y} \subset \mathbb{R}$

Regresja porządkowa

- $Z = (X, Y)$, $Z' = (X', Y')$ - niezależne wektory losowe o tym samym rozkładzie P
- $X, X' \in \mathcal{X}$, $Y, Y' \in \mathcal{Y} \subset \mathbb{R}$
- X, X' - obserwowane wektory cech
- Y, Y' - nieznane zmienne losowe

Regresja porządkowa

- z jest "lepszy" niż z', jeśli $y > y'$

Regresja porządkowa

- z jest "lepszy" niż z', jeśli $y > y'$
- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

Regresja porządkowa

- z jest "lepszy" niż z', jeśli $y > y'$
- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$\text{jeśli } f(x, x') > 0, \ \text{ to przewidujemy } y > y'$$

Regresja porządkowa

- z jest "lepszy" niż z', jeśli $y > y'$
- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$\text{jeśli } f(x, x') > 0, \text{ to przewidujemy } y > y'$$

- Minimalizacja ryzyka

$$Q(f) = \mathbf{E}\, \phi \left[\operatorname{sign}(Y - Y')\, f(X, X') \right]$$

Regresja porządkowa

- z jest "lepszy" niż z', jeśli $y > y'$
- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$\text{jeśli } f(x, x') > 0, \text{ to przewidujemy } y > y'$$

- Minimalizacja ryzyka

$$Q(f) = \mathbf{E}\, \phi\left[\operatorname{sign}(Y - Y')\, f(X, X')\right]$$

- $\phi : \mathbb{R} \to \mathbb{R}$ - funkcja straty

Regresja porządkowa

- z jest "lepszy" niż z', jeśli $y > y'$
- Reguła rangująca $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$\text{jeśli } f(x, x') > 0, \text{ to przewidujemy } y > y'$$

- Minimalizacja ryzyka

$$Q(f) = \mathbf{E}\, \phi \left[\operatorname{sign}(Y - Y')\, f(X, X') \right]$$

- $\phi : \mathbb{R} \to \mathbb{R}$ - funkcja straty

$$\bar{f} = \arg \min_{f \in \mathbb{F}} Q(f)$$

Regresja porządkowa

- $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ - niezależne kopie Z

Regresja porządkowa

- $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ - niezależne kopie Z
- Minimalizacja ryzyka empirycznego

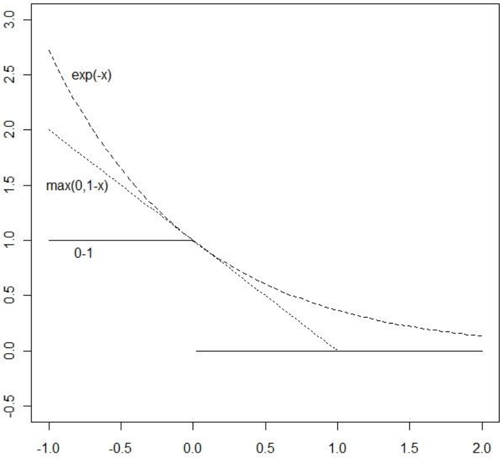
$$Q_n(f) = \frac{1}{n(n-1)} \sum_{1 \leqslant i \neq j \leqslant n} \phi\left[\mathrm{sign}(Y_i - Y_j)\, f(X_i, X_j)\right]$$

Regresja porządkowa

- $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ - niezależne kopie Z
- Minimalizacja ryzyka empirycznego

$$Q_n(f) = \frac{1}{n(n-1)} \sum_{1 \leqslant i \neq j \leqslant n} \phi\left[\mathrm{sign}(Y_i - Y_j)\, f(X_i, X_j)\right]$$

$$\arg\min_{f \in \mathcal{F}} Q_n(f), \qquad \mathcal{F} \subset \mathbb{F}$$


exp(-x)
max(0,1-x)
0-1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0

Kara LASSO

- $\psi_1, \ldots, \psi_m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ - funkcje bazowe

Kara LASSO

- $\psi_1, \ldots, \psi_m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ - funkcje bazowe
- $$\mathcal{F} = \left\{ f_\theta(x, x') = \sum_{k=1}^{m} \theta_k \psi_k(x, x') : \theta \in \Theta \subset \mathbb{R}^m \right\}$$

Kara LASSO

- $\psi_1, \ldots, \psi_m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ - funkcje bazowe
- $$\mathcal{F} = \left\{ f_\theta(x, x') = \sum_{k=1}^{m} \theta_k \psi_k(x, x') : \theta \in \Theta \subset \mathbb{R}^m \right\}$$
- $\mathcal{X} = \mathbb{R}^m$
$$\psi_k(x, x') = x_k - x'_k$$

Kara LASSO

- $\psi_1, \ldots, \psi_m : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ - funkcje bazowe
- $$\mathcal{F} = \left\{ f_\theta(x, x') = \sum_{k=1}^{m} \theta_k \psi_k(x, x') : \theta \in \Theta \subset \mathbb{R}^m \right\}$$
- $\mathcal{X} = \mathbb{R}^m$
$$\psi_k(x, x') = x_k - x'_k$$
$$f_\theta(x, x') = \theta^T (x - x')$$

Kara LASSO

- $$Q_n(f_\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} \phi \left[\mathrm{sign}(Y_i - Y_j) \, f_\theta(X_i, X_j) \right]$$

Kara LASSO

- $$Q_n(f_\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} \phi \left[\mathrm{sign}(Y_i - Y_j)\, f_\theta(X_i, X_j) \right]$$

- $$f_{\hat{\theta}} = \arg \min_{\theta} Q_n(f_\theta) + \lambda_n \sum_{k=1}^{m} |\theta_k|$$

Kara LASSO

- $$Q_n(f_\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} \phi \left[\mathrm{sign}(Y_i - Y_j)\, f_\theta(X_i, X_j) \right]$$

- $$f_{\hat{\theta}} = \arg \min_{\theta} Q_n(f_\theta) + \lambda_n \sum_{k=1}^{m} |\theta_k|$$

- Tarigan, van de Geer (2006), van de Geer (2008), Bickel, Ritov, Tsybakov (2009)

Założenia

1. funkcja straty ϕ jest wypukła

Założenia

1. funkcja straty ϕ jest wypukła
2. $$\forall_{1 \leqslant k \leqslant m} \, \forall_{x,x'} \quad |\psi_k(x,x')| \leqslant \sqrt{\frac{n}{\log m}}$$

Założenia

1. funkcja straty ϕ jest wypukła
2. $$\forall_{1 \leqslant k \leqslant m} \, \forall_{x,x'} \quad |\psi_k(x, x')| \leqslant \sqrt{\frac{n}{\log m}}$$
3. $(\mathbb{F}, || \cdot ||)$

Założenia

1. funkcja straty ϕ jest wypukła
2. $$\forall_{1 \leqslant k \leqslant m} \forall_{x,x'} \quad |\psi_k(x,x')| \leqslant \sqrt{\frac{n}{\log m}}$$
3. $(\mathbb{F}, ||\cdot||)$

$$\exists_{B>0} \quad \forall_{f_\theta \in \mathcal{F}} \quad ||f_\theta - \bar{f}||^2 \leqslant B \left[Q(f_\theta) - Q(\bar{f}) \right]$$

Założenia - warunek zgodności (WZ)

$$S \subset \{1, \ldots, m\} \text{ oraz } S' = \{1, \ldots, m\} \backslash S$$

Założenia - warunek zgodności (WZ)

$S \subset \{1, \ldots, m\}$ oraz $S' = \{1, \ldots, m\} \setminus S$

$\theta_S : \quad (\theta_S)_k = \theta_k \, \mathbb{I}(k \in S), \quad k = 1, \ldots, m$

Założenia - warunek zgodności (WZ)

$S \subset \{1, \ldots, m\}$ oraz $S' = \{1, \ldots, m\} \backslash S$

$\theta_S : \quad (\theta_S)_k = \theta_k \, \mathbb{I}(k \in S), \quad k = 1, \ldots, m$

$\theta = \theta_S + \theta_{S'}$

Założenia - warunek zgodności (WZ)

$S \subset \{1, \ldots, m\}$ oraz $S' = \{1, \ldots, m\} \backslash S$
$\theta_S : \quad (\theta_S)_k = \theta_k \, \mathbb{I}(k \in S), \quad k = 1, \ldots, m$
$\theta = \theta_S + \theta_{S'}$

- Zbiór S spełnia warunek zgodności, o ile istnieje stała $A(S) > 0$ taka, że

$$|\theta_S|_1^2 \leqslant \frac{||f_\theta||^2 \, |S|}{A(S)}$$

dla wszystkich $\theta \in \mathbb{R}^m$ spełniających $|\theta_{S'}|_1 \leqslant 3|\theta_S|_1$.

Założenia - warunek zgodności (WZ)

- $\psi(x, x') = [\psi_1(x, x'), \ldots, \psi_m(x, x')]^T$

Założenia - warunek zgodności (WZ)

- $\psi(x,x') = [\psi_1(x,x'), \ldots, \psi_m(x,x')]^T$
- $\Sigma = \mathbf{E}\,\psi(X,X')\psi^T(X,X')$

Założenia - warunek zgodności (WZ)

- $\psi(x, x') = [\psi_1(x, x'), \ldots, \psi_m(x, x')]^T$
- $\Sigma = \mathbf{E}\,\psi(X, X')\psi^T(X, X')$
- $||f_\theta||^2 := \mathbf{E} f_\theta^2(X, X') = \theta^T \Sigma \theta$

Założenia - warunek zgodności (WZ)

- $\psi(x, x') = [\psi_1(x, x'), \ldots, \psi_m(x, x')]^T$
- $\Sigma = \mathbf{E}\, \psi(X, X')\psi^T(X, X')$
- $||f_\theta||^2 := \mathbf{E} f_\theta^2(X, X') = \theta^T \Sigma \theta$

- Najmniejsza wartość własna ρ jest dodatnia

Założenia - warunek zgodności (WZ)

- $\psi(x, x') = [\psi_1(x, x'), \ldots, \psi_m(x, x')]^T$
- $\Sigma = \mathbf{E}\, \psi(X, X')\psi^T(X, X')$
- $||f_\theta||^2 := \mathbf{E} f_\theta^2(X, X') = \theta^T \Sigma \theta$

- Najmniejsza wartość własna ρ jest dodatnia

$$\forall_{\theta \in \mathbb{R}^m} \qquad |\theta|_2^2 \leqslant \frac{||f_\theta||^2}{\rho}$$

Wyrocznia

- $S_\theta = \{1 \leqslant k \leqslant m : \theta_k \neq 0\}$

Wyrocznia

- $S_\theta = \{1 \leqslant k \leqslant m : \theta_k \neq 0\}$
- $\Theta_1 = \{\theta \in \Theta : S_\theta \text{ spełnia (WZ)}\}$

Wyrocznia

- $S_\theta = \{1 \leqslant k \leqslant m : \theta_k \neq 0\}$
- $\Theta_1 = \{\theta \in \Theta : S_\theta \text{ spełnia (WZ)}\}$
- $$\theta^* = \arg\min_{\theta \in \Theta_1} \left\{ Q(f_\theta) - Q(\bar{f}) \quad + \quad \frac{\lambda_n^2 \, |S_\theta|}{A^2(\theta)} \right\}$$

Nierówności z wyrocznią

- Z prawdopodobieństwem $\geqslant 1 - \frac{1}{m^2}$

$$Q(f_{\hat{\theta}}) - Q(\bar{f}) \leqslant C_1 \left[Q(f_{\theta^*}) - Q(\bar{f}) \quad + \quad \frac{\lambda_n^2 \, |S_{\theta^*}|}{A^2(\theta^*)} \right]$$

Nierówności z wyrocznią

- Z prawdopodobieństwem $\geqslant 1 - \frac{1}{m^2}$

$$Q(f_{\hat{\theta}}) - Q(\bar{f}) \leqslant C_1 \left[Q(f_{\theta^*}) - Q(\bar{f}) \quad + \quad \frac{\lambda_n^2 \, |S_{\theta^*}|}{A^2(\theta^*)} \right]$$

$$\lambda_n = C_2 \sqrt{\frac{\log m}{n}}$$

Nierówności z wyrocznią

- Z prawdopodobieństwem $\geqslant 1 - \frac{1}{m^2}$

$$Q(f_{\hat{\theta}}) - Q(\bar{f}) \leqslant C_1 \left[Q(f_{\theta^*}) - Q(\bar{f}) \;\; + \;\; \frac{\lambda_n^2 \, |S_{\theta^*}|}{A^2(\theta^*)} \right]$$

$$\lambda_n = C_2 \sqrt{\frac{\log m}{n}}$$

- $m \sim n^d$ dla $d \geqslant 1$

Nierówności z wyrocznią

- Z prawdopodobieństwem $\geqslant 1 - \frac{1}{m^2}$

$$\left|\hat{\theta} - \theta^*\right|_1 \leqslant C_1 \left[Q(f_{\theta^*}) - Q(\bar{f}) \;\; + \;\; \frac{\lambda_n^2 \, |S_{\theta^*}|}{A^2(\theta^*)}\right]$$

Nierówności z wyrocznią

- Z prawdopodobieństwem $\geqslant 1 - \frac{1}{m^2}$

$$\left|\hat{\theta} - \theta^*\right|_1 \leqslant C_1 \left[Q(f_{\theta^*}) - Q(\bar{f}) \ + \ \frac{\lambda_n^2 \left|S_{\theta^*}\right|}{A^2(\theta^*)} \right]$$

- $$\left|\hat{\theta} - \theta^*\right|_1 \leqslant \frac{\log m}{n}$$

Bibliografia

Bickel, P. J., Ritov, Y., Tsybakov, A., B. (2009). *Simultaneous analysis of Lasso and Dantzig selector.* Annals of Statistics 37, 1705–1732.

Bühlmann, P., van de Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer.

Tarigan, B., van de Geer, S. (2006). *Classifiers of support vector machine type with* l_1 *penalty.* Bernoulli 12, 1045–1076.

Tibshirani, R. (1996). *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society: Series B 58, 267–288.

van de Geer, S. (2008). *High-dimensional generalized linear models and the Lasso.* Annals of Statistics 36, 614–645.