ERCIM'11 18 December 2011

On Support Vector Machines to Estimate Scale Functions

Robert Hable & Andreas Christmann

Robert Hable
Department of Mathematics
University of Bayreuth

Nonparametric Regression

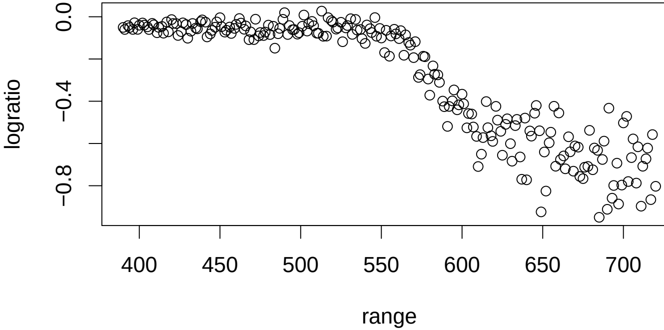
$$Y = f_0(X) + \varepsilon$$

Nonparametric Regression

$$Y = f_0(X) + \varepsilon$$

Example:

LIDAR = Light Detection And Ranging (data set with $n = 221$)

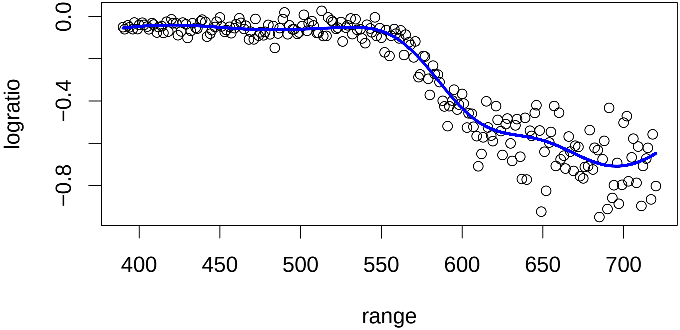


Nonparametric Regression

$$Y = f_0(X) + \varepsilon$$

Example:

LIDAR = Light Detection And Ranging (data set with $n = 221$)

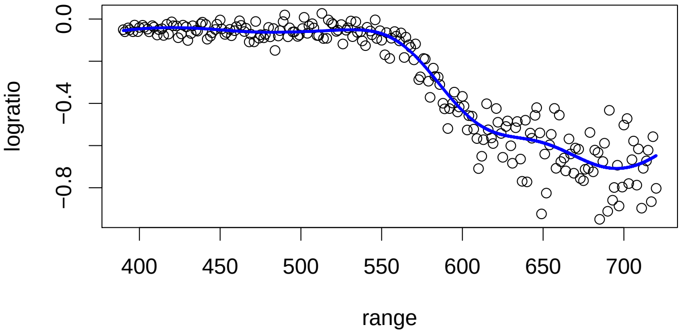


Nonparametric Regression with Heteroscedastic Errors

$$Y = f_0(X) + g_0(X)\varepsilon$$

Example:

LIDAR = Light Detection And Ranging (data set with $n = 221$)



Scale functions

Heteroscedastic model:

$$Y = f_0(X) + g_0(X)\varepsilon$$

Note that

- heteroscedasticity occurs in practice
- regularized kernel methods (SVM) estimate location f_0
  - in a non-parametric way
  - also in case of heteroscedasticity

Question:

How to estimate the scale function g_0 ?

Estimation of Scale

$Y,\ Y_1, \ldots,\ Y_n \ \sim_{\text{i.i.d.}}\ P_{\mathcal{Y}}$

We want to estimate: **Scale**(Y)

- **Variance:** $\text{mean}\big((Y - \text{mean}(Y))^2\big)$
  - estimate $\text{mean}(Y)$ by $\hat{\mu}_n$ and put $\tilde{Y}_i := (Y_i - \hat{\mu}_n)^2$
  - estimate $\text{mean}(\tilde{Y}_i)$

Estimation of Scale

$Y,\ Y_1, \ldots, \ Y_n \ \sim_{\text{i.i.d.}} \ P_{\mathcal{Y}}$

We want to estimate: **Scale**(Y)

- **Variance:** $\text{mean}\big((Y - \text{mean}(Y))^2\big)$
  - estimate $\text{mean}(Y)$ by $\hat{\mu}_n$ and put $\tilde{Y}_i := (Y_i - \hat{\mu}_n)^2$
  - estimate $\text{mean}(\tilde{Y}_i)$
- **MAD:** $\text{median}\big(|Y - \text{median}(Y)|\big)$
  - estimate $\text{median}(Y)$ by $\hat{\mu}_n$ and put $\tilde{Y}_i := |Y_i - \hat{\mu}_n|$
  - estimate $\text{median}(\tilde{Y}_i)$

Estimation of Scale

$Y,\ Y_1, \ldots, Y_n \ \sim_{\text{i.i.d.}} \ P_{\mathcal{Y}}$

We want to estimate: **Scale**(Y)

- **Variance:** $\text{mean}\big((Y - \text{mean}(Y))^2\big)$
  - estimate mean(Y) by $\hat{\mu}_n$ and put $\tilde{Y}_i := (Y_i - \hat{\mu}_n)^2$
  - estimate mean$(\tilde{Y}_i)$
- **MAD:** $\text{median}\big(|Y - \text{median}(Y)|\big)$
  - estimate median(Y) by $\hat{\mu}_n$ and put $\tilde{Y}_i := |Y_i - \hat{\mu}_n|$
  - estimate median$(\tilde{Y}_i)$
- **IQR:** upper.quartile(Y) − lower.quartile(Y)
  - estimate upper and lower quartiles by $\hat{Q}_3$ and $\hat{Q}_1$
  - calculate $\hat{Q}_3 - \hat{Q}_1$

Estimation of Scale Functions

$(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n) \sim_{\text{i.i.d.}} P$

We want to estimate: **Scale**$(Y | X = \cdot)$

- **Variance function:** $\text{mean}\big((Y - \text{mean}(Y|X=\cdot))^2 \,\big|\, X = \cdot\big)$
  - estimate $\text{mean}(Y|X=\cdot)$ by $\hat{f}_n$ and put $\tilde{Y}_i := (Y_i - \hat{f}_n(X_i))^2$
  - estimate $\text{mean}(\tilde{Y}_i | X_i = \cdot)$
- **MAD function:** $\text{median}\big(|Y - \text{median}(Y|X=\cdot)| \,\big|\, X = \cdot\big)$
  - estimate $\text{median}(Y|X=\cdot)$ by $\hat{f}_n$ and put $\tilde{Y}_i := |Y_i - \hat{f}_n(X_i)|$
  - estimate $\text{median}(\tilde{Y}_i | X_i = \cdot)$
- **IQR function:** $\text{upper.quartile}(Y|X=\cdot) - \text{lower.quartile}(Y|X=\cdot)$
  - estimate upper and lower quartile functions by $\hat{f}_3$ and $\hat{f}_1$
  - calculate $\hat{f}_3 - \hat{f}_1$

Estimation of Scale Functions

Summing up:

- Estimation of scale can be done by estimating
  mean, median, or quartiles
- Estimation of scale functions can be done by estimating
  conditional mean, median, or quartile functions

Estimation of Scale Functions

Summing up:

- Estimation of scale can be done by estimating
  mean, median, or quartiles
- Estimation of scale functions can be done by estimating
  conditional mean, median, or quartile functions

and

- Regularized kernel-methods (SVMs) estimate
  conditional mean, median, or quartile functions

Hence: Regularized kernel-methods should
be able to estimate scale functions

Regularized Kernel-Methods (SVMs)

$$Y_i = f_0(X_i) + g_0(X_i)\varepsilon_i\,, \qquad (X_i, Y_i) \sim P \quad \text{i.i.d.}, \qquad i \in \{1, \ldots, n\}$$

Goal: Estimation of $f_0 : \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}$

Regularized Kernel-Methods (SVMs)

$$Y_i \;=\; f_0(X_i) + g_0(X_i)\varepsilon_i\,, \qquad (X_i, Y_i) \;\sim\; P \quad \text{i.i.d.}, \qquad i \in \{1, \ldots, n\}$$

Goal: Estimation of $f_0 : \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}$

- Loss function

$$L \;:\; \mathcal{Y} \times \mathbb{R} \;\to\; [0, \infty)$$

$L(y, t):$ loss caused by estimation $t = \hat{f}_n(x)$ if y is true

Regularized Kernel-Methods (SVMs)

$$Y_i \;=\; f_0(X_i) + g_0(X_i)\varepsilon_i\,, \qquad (X_i, Y_i) \;\sim\; P \quad \text{i.i.d.}, \qquad i \in \{1, \ldots, n\}$$

Goal: Estimation of $f_0 : \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}$

- Loss function

$$L \;:\; \mathcal{Y} \times \mathbb{R} \;\to\; [0, \infty)$$

  $L(y, t)$: loss caused by estimation $t = \hat{f}_n(x)$ if y is true
- Risk of an estimate $\hat{f}_n : \mathcal{X} \to \mathbb{R}$

$$\int L\big(y, \hat{f}_n(x)\big)\, P\big(d(x, y)\big)$$

Regularized Kernel-Methods (SVMs)

$$Y_i \;=\; f_0(X_i) + g_0(X_i)\varepsilon_i\,, \qquad (X_i, Y_i) \;\sim\; P \quad \text{i.i.d.,} \qquad i \in \{1, \ldots, n\}$$

Goal: Estimation of $f_0 : \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}$

- Loss function

$$L \;:\; \mathcal{Y} \times \mathbb{R} \;\to\; [0, \infty)$$

  $L(y, t)$: loss caused by estimation $t = \hat{f}_n(x)$ if y is true
- Empirical risk of an estimate $\hat{f}_n : \mathcal{X} \to \mathbb{R}$

$$\frac{1}{n} \sum_{i=1}^{n} L\big(y_i, \hat{f}_n(x_i)\big)$$

Regularized Kernel-Methods (SVMs)

$$Y_i \;=\; f_0(X_i)+g_0(X_i)\varepsilon_i\,, \qquad (X_i, Y_i) \;\sim\; P \quad \text{i.i.d.}, \qquad i \in \{1, \dots, n\}$$

Goal: Estimation of $f_0 : \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}$

- Loss function

$$L \;:\; \mathcal{Y} \times \mathbb{R} \;\to\; [0, \infty)$$

  $L(y, t)$: loss caused by estimation $t = \hat{f}_n(x)$ if y is true
- Empirical risk of an estimate $\hat{f}_n : \mathcal{X} \to \mathbb{R}$

$$\frac{1}{n} \sum_{i=1}^{n} L\big(y_i, \hat{f}_n(x_i)\big)$$

- RKHS H (certain Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$)

Regularized Kernel-Methods (SVMs)

$$Y_i \;=\; f_0(X_i)+g_0(X_i)\varepsilon_i\,, \qquad (X_i,Y_i)\;\sim\;P \quad \text{i.i.d.}, \qquad i\in\{1,\dots,n\}$$

Goal: Estimation of $f_0:\mathcal{X}\to\mathcal{Y}\subset\mathbb{R}$

- Loss function

$$L\;:\;\;\mathcal{Y}\times\mathbb{R}\;\to\;[0,\infty)$$

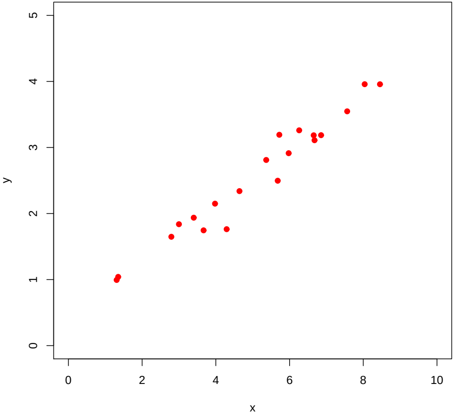
  $L(y,t):$ loss caused by estimation $t=\hat{f}_n(x)$ if y is true
- Empirical risk of an estimate $\hat{f}_n:\mathcal{X}\to\mathbb{R}$

$$\frac{1}{n}\sum_{i=1}^{n}L\big(y_i,\hat{f}_n(x_i)\big)$$

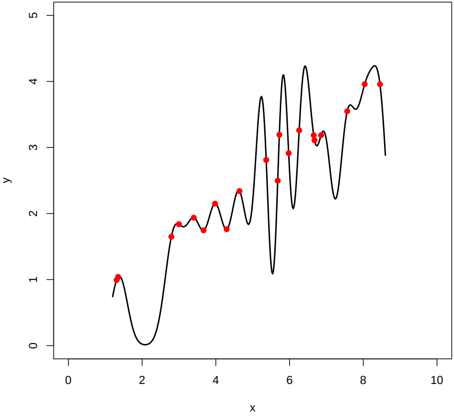
- RKHS H (certain Hilbert space of functions $f:\mathcal{X}\to\mathbb{R}$)
- Support vector machine

$$S_n\big((x_1,y_1),\dots,(x_n,y_n)\big)\;=\;\arg\inf_{f\in H}\;\frac{1}{n}\sum_{i=1}^{n}L\big(y_i,f(x_i)\big)$$

Overfitting



Overfitting



Regularized Kernel-Methods (SVMs)

$$Y_i = f_0(X_i) + g_0(X_i)\varepsilon_i\,, \qquad (X_i, Y_i) \sim P \quad \text{i.i.d.,} \qquad i \in \{1, \dots, n\}$$

Goal: Estimation of $f_0 : \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}$

- Loss function

$$L \;:\; \mathcal{Y} \times \mathbb{R} \;\to\; [0, \infty)$$

  $L(y, t)$: loss caused by prediction t if y is the true value
- Empirical risk of an estimate $f : \mathcal{X} \to \mathbb{R}$

$$\frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(x_i)\big)$$

- RKHS H (certain Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$)
- Support vector machine

$$S_n\big((x_1, y_1), \dots, (x_n, y_n)\big) \;=\; \arg\inf_{f \in H} \; \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(x_i)\big)$$

Regularized Kernel-Methods (SVMs)

$$Y_i \; = \; f_0(X_i) + g_0(X_i)\varepsilon_i\,, \qquad (X_i, Y_i) \;\sim\; P \quad \text{i.i.d.}, \qquad i \in \{1, \dots, n\}$$

Goal: Estimation of $f_0 : \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}$

- Loss function

$$L \;:\; \mathcal{Y} \times \mathbb{R} \;\to\; [0, \infty)$$

  $L(y,t)$: loss caused by prediction t if y is the true value

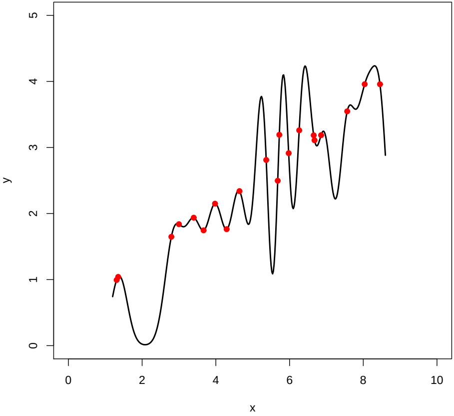
- Empirical risk of an estimate $f : \mathcal{X} \to \mathbb{R}$

$$\frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big)$$

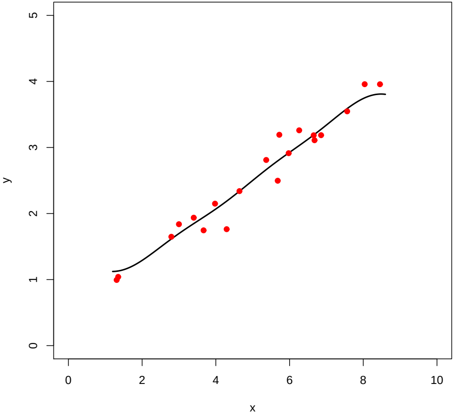
- RKHS H (certain Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$)
- Support vector machine

$$S_n\big((x_1, y_1), \dots, (x_n, y_n)\big) \;=\; \arg\inf_{f \in H} \; \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) + \lambda \|f\|_H^2$$

Overfitting



Overfitting



Estimation of Scale Functions

Summing up:

- Estimation of scale can be done by estimating
  mean, median, or quartiles
- Estimation of scale functions can be done by estimating
  conditional mean, median, or quartile functions

and

- Regularized kernel-methods (SVMs) estimate
  conditional mean, median, or quartile functions

Hence: Regularized kernel-methods should
be able to estimate scale functions

Estimation of the Conditional Mean Function

$$S_n\big((x_1, y_1), \ldots, (x_n, y_n)\big) \;=\; \arg\inf_{f \in H} \; \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda \|f\|_H^2$$

Estimation of the **conditional mean function**

$$x \;\mapsto\; \text{mean}(Y|X = x)$$

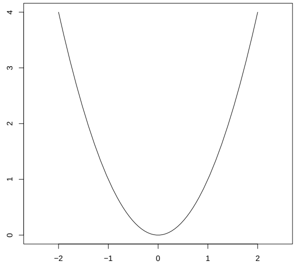
Estimation of the Conditional Mean Function

$$S_n\big((x_1, y_1), \dots, (x_n, y_n)\big) \;=\; \arg\inf_{f \in H} \; \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda \|f\|_H^2$$

Estimation of the **conditional mean function**

$$x \;\mapsto\; \text{mean}(Y|X = x)$$

by use of **SVM** with **least-squares loss**: $L(y, t) = (y - t)^2$



Estimation of the Conditional Median Function

$$S_n\big((x_1,y_1),\ldots,(x_n,y_n)\big) \;=\; \arg\inf_{f\in H}\; \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda\|f\|_H^2$$

Estimation of the **conditional median function**

$$x \;\mapsto\; \text{median}(Y|X=x)$$

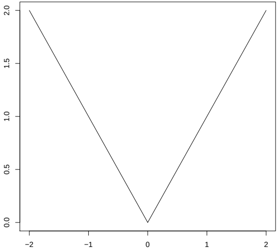
Estimation of the Conditional Median Function

$$S_n\big((x_1, y_1), \ldots, (x_n, y_n)\big) \;=\; \arg\inf_{f \in H} \; \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda \|f\|_H^2$$

Estimation of the **conditional median function**

$$x \;\mapsto\; \text{median}(Y|X = x)$$

by use of **SVM** with **absolute distance**: $L(y, t) = |y - t|$



Estimation of Conditional Quantile Functions

$$S_n\big((x_1, y_1), \ldots, (x_n, y_n)\big) \;=\; \arg\inf_{f \in H} \; \frac{1}{n} \sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda \|f\|_H^2$$

Estimation of the **conditional** τ**-quantile function**

$$x \;\mapsto\; \text{quantile}_\tau(Y|X = x)$$

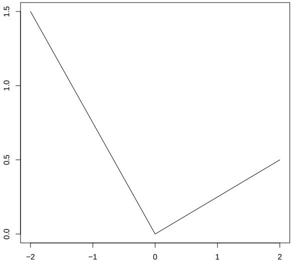
Estimation of Conditional Quantile Functions

$$S_n\big((x_1,y_1),\ldots,(x_n,y_n)\big) \;=\; \arg\inf_{f\in H}\; \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda \|f\|_H^2$$

Estimation of the **conditional τ-quantile function**

$$x \;\mapsto\; \text{quantile}_\tau(Y|X=x)$$

by use of **SVM** with **pinball loss**: $L(y,t) = \begin{cases} (1-\tau)(t-y) & \text{if } y < t \\ \tau(y-t) & \text{if } y \geq t \end{cases}$



Estimation of Scale Functions

$(X,Y), (X_1,Y_1), \ldots, (X_n,Y_n) \sim_{\text{i.i.d.}} P$

We want to estimate: $\textbf{Scale}(Y|X=\cdot)$

- **Variance function:** $\text{mean}\big((Y - \text{mean}(Y|X=\cdot))^2 \big| X=\cdot\big)$
  - estimate $\text{mean}(Y|X=\cdot)$ by $\hat{f}_n$ and put $\tilde{Y}_i := (Y_i - \hat{f}_n(X_i))^2$
  - estimate $\text{mean}(\tilde{Y}_i|X_i=\cdot)$
- **MAD function:** $\text{median}\big(|Y - \text{median}(Y|X=\cdot)| \,\big|\, X=\cdot\big)$
  - estimate $\text{median}(Y|X=\cdot)$ by $\hat{f}_n$ and put $\tilde{Y}_i := |Y_i - \hat{f}_n(X_i)|$
  - estimate $\text{median}(\tilde{Y}_i|X_i=\cdot)$
- **IQR function:** $\text{upper.quartile}(Y|X=\cdot) - \text{lower.quartile}(Y|X=\cdot)$
  - estimate upper and lower quartile functions by $\hat{f}_3$ and $\hat{f}_1$
  - calculate $\hat{f}_3 - \hat{f}_1$

MAD function

MAD function: $g_0 = \text{median}\big(|Y - \text{median}(Y|X=\cdot)| \,\big|\, X=\cdot\big)$

MAD function

MAD function: $g_0 = \text{median}\big(|Y - \text{median}(Y|X=\cdot)| \,\big|\, X=\cdot\big)$

- estimate $\text{median}(Y|X=\cdot)$ by SVM $f_{\text{SVM},n}$ with absolute deviation loss

MAD function

MAD function: $g_0 = \text{median}\big(|Y - \text{median}(Y|X=\cdot)| \,\big|\, X=\cdot\big)$

- estimate $\text{median}(Y|X=\cdot)$ by SVM $f_{\text{SVM},n}$ with absolute deviation loss
- put $\tilde{Y}_i := |Y_i - f_{\text{SVM},n}(X_i)|$

MAD function

MAD function: $g_0 = \text{median}\big(|Y - \text{median}(Y|X=\cdot)|\,\big|\,X=\cdot\big)$

- estimate $\text{median}(Y|X=\cdot)$ by SVM $f_{\text{SVM},n}$ with absolute deviation loss
- put $\tilde{Y}_i := |Y_i - f_{\text{SVM},n}(X_i)|$
- use non-i.i.d. data $(X_1, \tilde{Y}_1), \ldots, (X_n, \tilde{Y}_n)$ and estimate $\text{median}(\tilde{Y}_i|X_i=\cdot)$ by SVM $\tilde{g}_{\text{SVM},n}$ with absolute deviation loss

MAD function

MAD function: $g_0 = \text{median}\big(|Y - \text{median}(Y|X=\cdot)|\,\big|\,X=\cdot\big)$

- estimate $\text{median}(Y|X=\cdot)$ by SVM $f_{\text{SVM},n}$ with absolute deviation loss
- put $\tilde{Y}_i := |Y_i - f_{\text{SVM},n}(X_i)|$
- use non-i.i.d. data $(X_1, \tilde{Y}_1), \ldots, (X_n, \tilde{Y}_n)$ and estimate $\text{median}(\tilde{Y}_i|X_i=\cdot)$ by SVM $\tilde{g}_{\text{SVM},n}$ with absolute deviation loss
- put $g_{\text{SVM},n}(x) = \max\{\tilde{g}_{\text{SVM},n}(x), 0\}$

MAD function

MAD function: $g_0 \; = \; \text{median}\big(|Y - \text{median}(Y|X = \cdot)| \,\big|\, X = \cdot\big)$

- estimate $\text{median}(Y|X = \cdot)$ by SVM $f_{\text{SVM},n}$ with absolute deviation loss
- put $\tilde{Y}_i := |Y_i - f_{\text{SVM},n}(X_i)|$
- use non-i.i.d. data $(X_1, \tilde{Y}_1), \ldots, (X_n, \tilde{Y}_n)$ and estimate $\text{median}(\tilde{Y}_i|X_i = \cdot)$ by SVM $\tilde{g}_{\text{SVM},n}$ with absolute deviation loss
- put $g_{\text{SVM},n}(x) = \max\{\tilde{g}_{\text{SVM},n}(x), 0\}$

Risk of function g:

$$\mathcal{R}_P(g) = \int \Big| |y - f_0(x)| - g(x) \Big| \, P\big(d(x, y)\big)$$

Then, hopefully,

$$\lim_{n \to \infty} \Big| \mathcal{R}_P\big(g_{\text{SVM},n}\big) - \mathcal{R}_P\big(g_0\big) \Big| \; = \; 0$$

MAD-type Estimation

MAD function: $g_0 \; = \; \mathrm{median}\big(|Y - \mathrm{median}(Y|X=\cdot)| \,\big|\, X = \cdot\big)$

- estimate $\mathrm{median}(Y|X=\cdot)$ by SVM $f_{\mathrm{SVM},n}$ with absolute deviation loss
- put $\tilde{Y}_i := |Y_i - f_{\mathrm{SVM},n}(X_i)|$
- use non-i.i.d. data $(X_1, \tilde{Y}_1), \ldots, (X_n, \tilde{Y}_n)$ and estimate $\mathrm{median}_\varepsilon(\tilde{Y}_i|X_i=\cdot)$ by SVM $\tilde{g}_{\mathrm{SVM},n}$ with ε-smoothed absolute deviation loss
- put $g_{\mathrm{SVM},n}(x) = \max\{\tilde{g}_{\mathrm{SVM},n}(x), 0\}$

Risk of function g:

$$\mathcal{R}_P(g) = \int \Big| |y - f_0(x)| - g(x) \Big| \, P\big(d(x,y)\big)$$

Then, at least,

$$\limsup_{n\to\infty} \Big| \mathcal{R}_P\big(g_{\mathrm{SVM},n}\big) - \mathcal{R}_P\big(g_0\big) \Big| \; \leq \; \varepsilon$$

IQR-type Estimation

IQR function: $g_0 = f^*_{0.25} - f^*_{0.75}$
where $f^*_\tau := \text{quantile}_\tau(Y|X=\cdot)$

- estimate $\text{quantile}_{0.25}(Y|X=\cdot)$
  by SVM $f_{\text{SVM.1},n}$ with 0.25-pinball loss $L_{0.25}$
- estimate $\text{quantile}_{0.75}(Y|X=\cdot)$
  by SVM $f_{\text{SVM.2},n}$ with 0.75-pinball loss $L_{0.75}$
- put $g_{\text{SVM},n}(x) = f_{\text{SVM.2},n}(x) - f_{\text{SVM.1},n}(x)$

IQR-type Estimation

IQR function: $g_0 = f^*_{0.25} - f^*_{0.75}$
where $f^*_\tau := \text{quantile}_\tau(Y|X=\cdot)$

- estimate $\text{quantile}_{0.25}(Y|X=\cdot)$ by SVM $f_{\text{SVM.1},n}$ with 0.25-pinball loss $L_{0.25}$
- estimate $\text{quantile}_{0.75}(Y|X=\cdot)$ by SVM $f_{\text{SVM.2},n}$ with 0.75-pinball loss $L_{0.75}$
- put $g_{\text{SVM},n}(x) = f_{\text{SVM.2},n}(x) - f_{\text{SVM.1},n}(x)$

Risk of function f:

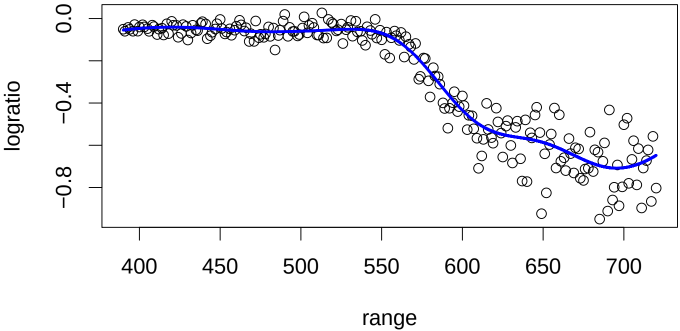
$$\mathcal{R}_{P,\tau}(f) = \int L_\tau\big(y, f(x)\big)\, P\big(d(x,y)\big)$$

Then,

$$\lim_{n\to\infty} \left| \Big(\mathcal{R}_{P,0.25}(f_{\text{SVM.1},n}) + \mathcal{R}_{P,0.75}(f_{\text{SVM.2},n})\Big) - \Big(\mathcal{R}_{P,0.25}(f^*_{0.25}) + \mathcal{R}_{P,0.75}(f^*_{0.75})\Big) \right| = 0$$

Example: LIDAR data set

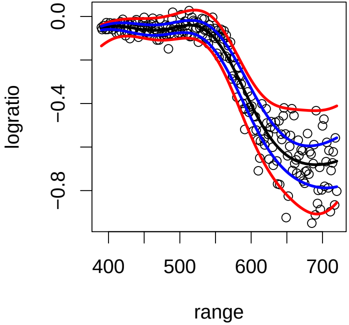
LIDAR = Light Detection And Ranging (data set with $n = 221$)



Example: LIDAR data set

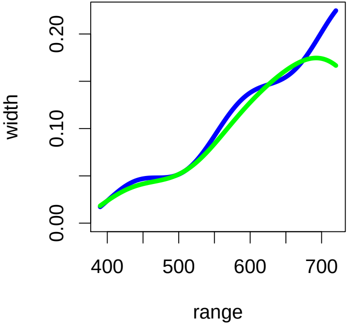
τ-quantile estimation with SVM

$\tau \in \{5\%, 25\%, 50\%, 75\%, 95\%\}$



estimated width of interval I with $\mathbb{P}(Y \in I | X = x) = 0.5$

IQR-type , 2×MAD-type

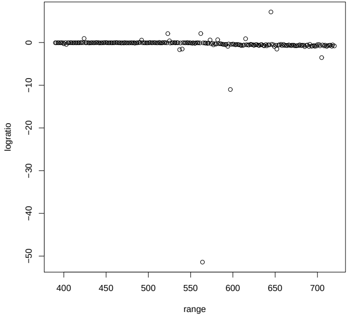


Robustness

Example

- LIDAR data set
- but now with additional 10% Cauchy errors in y-direction

whole data set

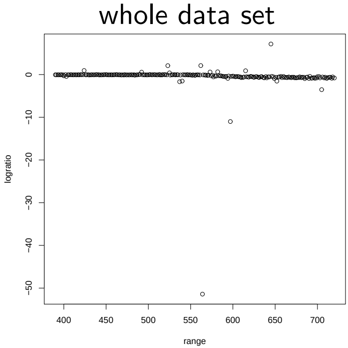


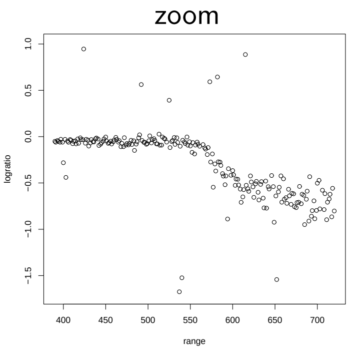
zoom

Robustness

Example

- LIDAR data set
- but now with additional 10% Cauchy errors in y-direction



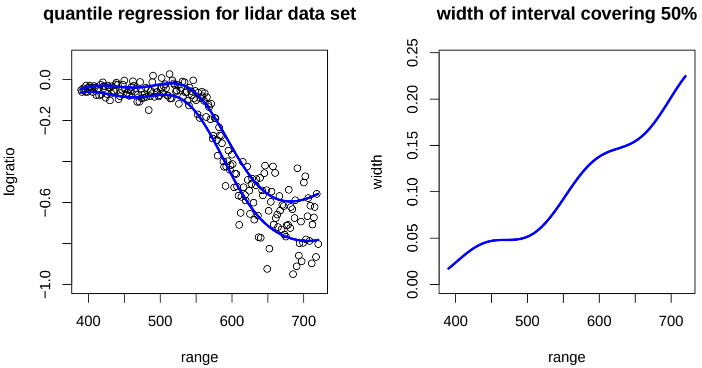


Robustness

Example: IQR-type estimation

- LIDAR data set
- but now with additional 10% Cauchy errors in *y*-direction

original data set

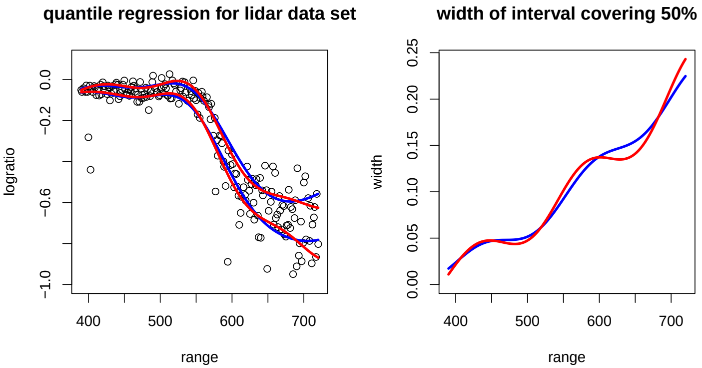


Robustness

Example: IQR-type estimation

- LIDAR data set
- but now with additional 10% Cauchy errors in y-direction

original data set   data set with 10% Cauchy



Robustness: Theoretical Results

MAD-type:

- uniform upper bound for the bias

IQR-type:

- uniform upper bound for the bias
- bounded Bouligand influence function
- qualitative robustness

References

- **Hable and Christmann (2011)**. Estimation of scale functions to model heteroscedasticity by support vector machines. *arXiv:*1111.1830.
- **Steinwart and Christmann (2008)**. *Support vector machines*. Springer, New York.
- **Vapnik (1998)**: *Statistical learning theory*. John Wiley & Sons, New York.
- **Christmann, Van Messem, and Steinwart (2009)**. On Consistency and Robustness Properties of Support Vector Machines for Heavy-Tailed Distributions. *Statistics and Its Interface*, 2, 311-327.
- **Hable and Christmann (2011)**. On qualitative robustness of support vector machines. *Journal of Multivariate Analysis*, 102:993-1007, 2011.
- **Steinwart and Christmann (2011)**. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17, 211-225.

The handout to this talk is also available on my homepage

http://www.staff.uni-bayreuth.de/~btms04/index.html