Qualitative Robustness of Support Vector Machines

Robert Hable, Andreas Christmann

Robert Hable
Department of Mathematics
University of Bayreuth

Machine Learning

- Input variables:

$$X_1,\ X_2,\ \ldots,\ X_n \quad \text{and} \quad X_0$$

  Output variables:

$$Y_1,\ Y_2,\ \ldots,\ Y_n \quad \text{and} \quad Y_0$$

- Observe data

$$(x_1, y_1), \ldots, (x_n, y_n)$$

  and "learn" the influence of the input variables X_i on the output variables Y_i .
- Observe another x_0 and try to predict unobserved y_0 .
- Predictor: function $f : x \mapsto f(x) = y$

Examples:

- Nonparametric Regression

$$Y_i \;=\; f_0(X_i) + \varepsilon_i\,, \qquad i \in \{1, \ldots, n\}$$

and

$$(X_i, Y_i) \;\sim\; P \qquad \text{i.i.d.}$$

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

Examples:

- Nonparametric Regression

$$Y_i \;=\; f_0(X_i) + \varepsilon_i\,, \qquad i \in \{1, \ldots, n\}$$

and

$$(X_i, Y_i) \;\sim\; P \qquad \text{i.i.d.}$$

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

- Classification
- Autoregression

Support Vector Machines (SVM)

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" predictor $f : x \mapsto f(x) = y$.

Support Vector Machines (SVM)

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

Support Vector Machines (SVM)

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

- 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

Support Vector Machines (SVM)

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

- 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 $\hat{f}_n : \mathcal{X} \to \mathbb{R}$

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

Support Vector Machines

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

- 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 $\hat{f}_n : \mathcal{X} \to \mathbb{R}$

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

Support Vector Machines

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

- 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 Machines

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

- Loss function

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

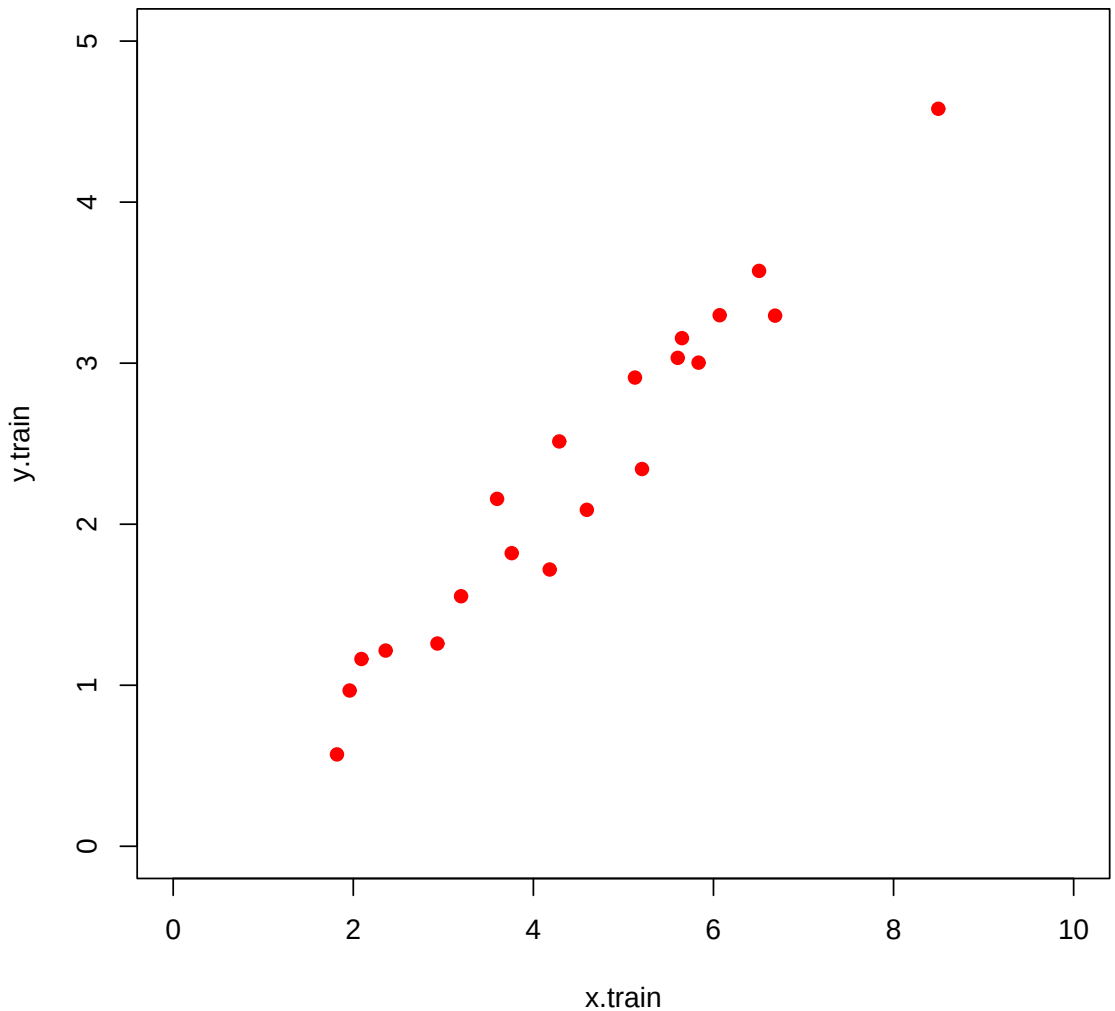
  $L(y, t)$: loss caused by estimation $t = \hat{f}_n(x)$ if y is true
- empirical Risk of an estimate $\hat{f}_n : \mathcal{X} \to \mathbb{R}$

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

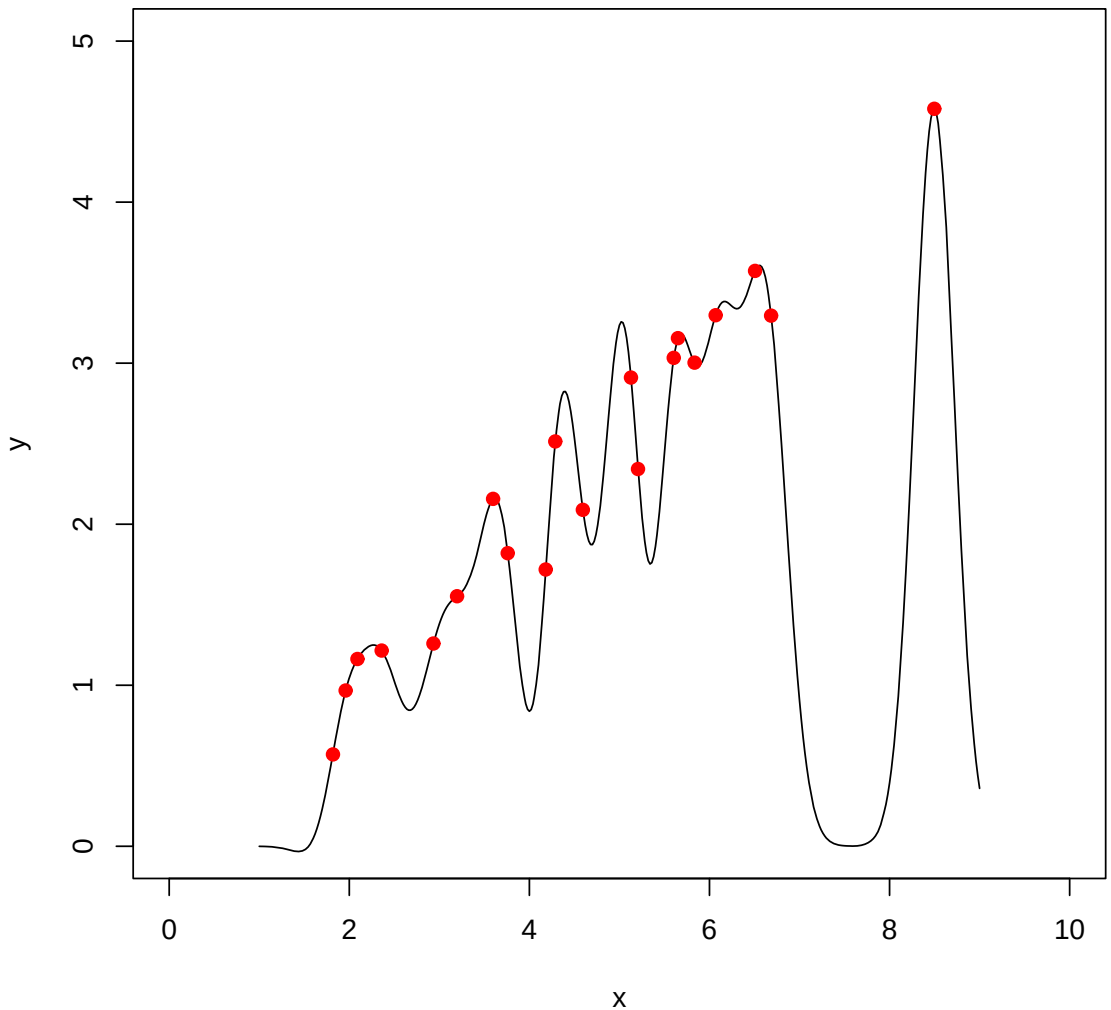
- RKHS H (certain Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$)
- Support vector machine

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

Overfitting



Overfitting



Support Vector Machines

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

- 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 $f : \mathcal{X} \to \mathbb{R}$

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

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

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

Support Vector Machines

Data $(x_1, y_1), \ldots, (x_n, y_n)$ from

$$(X_i, Y_i) \sim P \qquad \text{i.i.d.}$$

Find a "good" estimator of the regression function $f_0 : x \mapsto f_0(x)$.

- Loss function

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

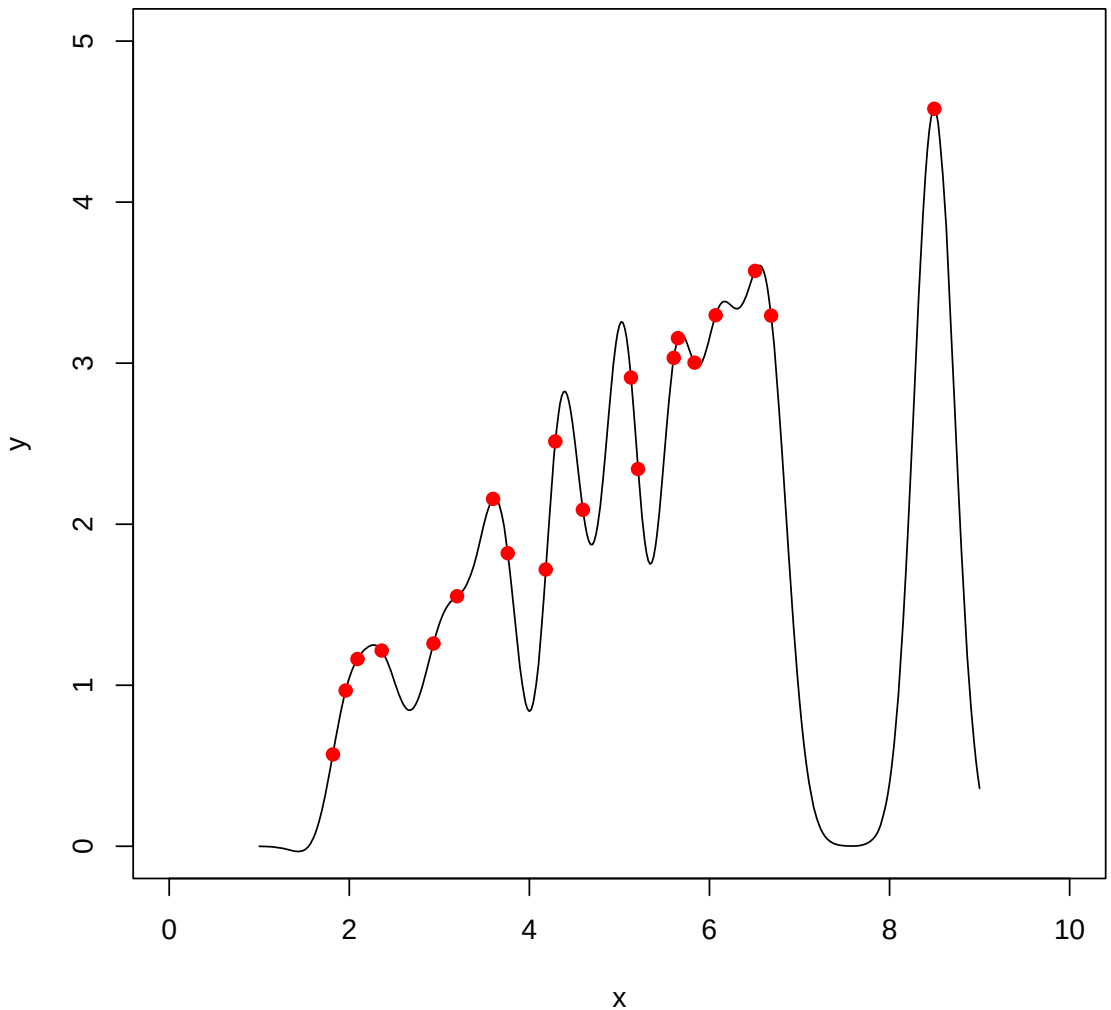
  $L(y, t)$: loss caused by estimation $t = \hat{f}_n(x)$ if y is true
- empirical Risk of an estimate $f : \mathcal{X} \to \mathbb{R}$

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

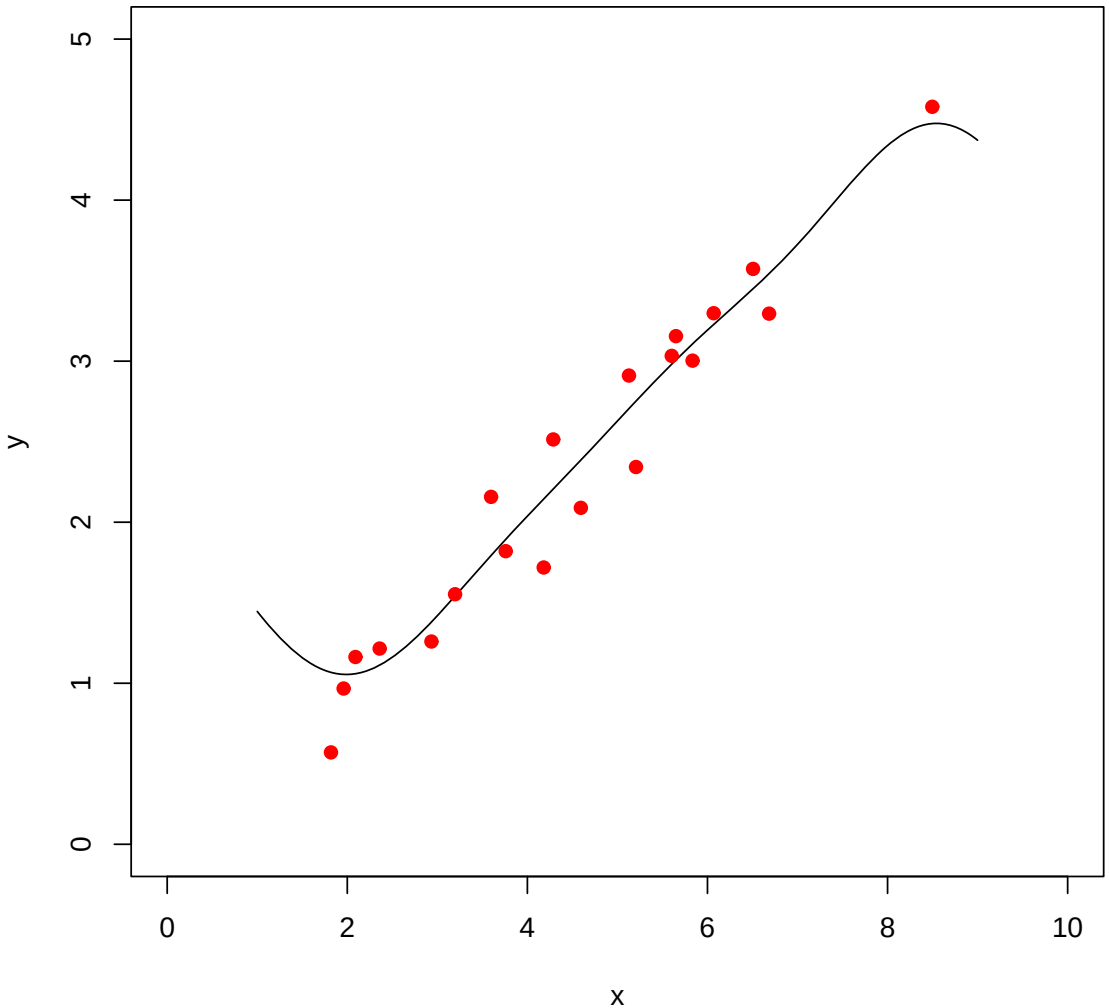
- RKHS H (certain Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$)
- Support vector machine

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

Overfitting



Overfitting



Qualitative Robustness

Small errors in the data should not change the results to much.

Qualitative Robustness

Small errors in the data should not change the results to much.

- "Small errors in the data"
  - Small errors in many of the data points (rounding etc.)
  - Large errors in a few data points (gross errors, outliers)

Qualitative Robustness

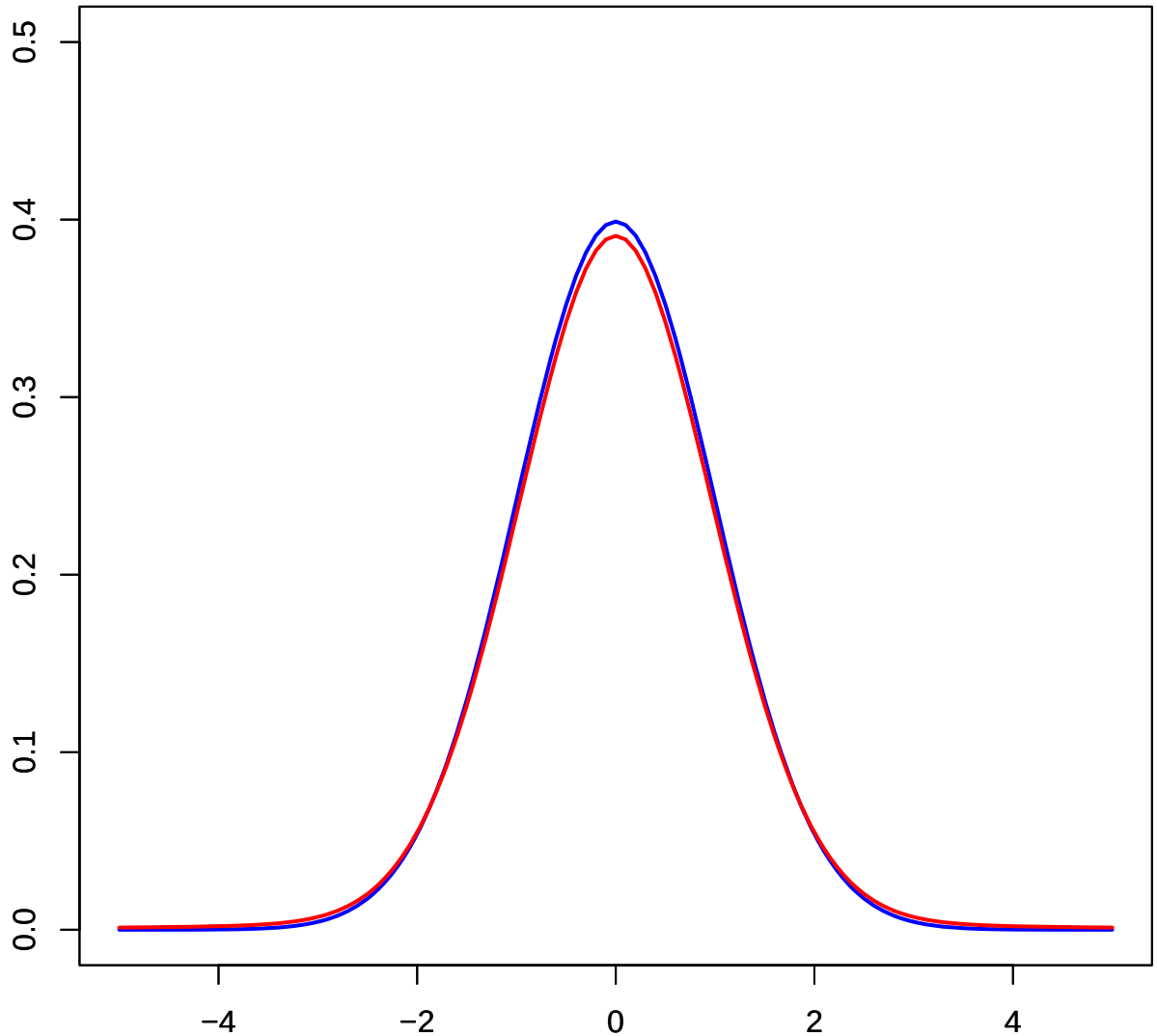
Small errors in the data should not change the results to much.

- "Small errors in the data"
  - Small errors in many of the data points (rounding etc.)
  - Large errors in a few data points (gross errors, outliers)
- "should not change the results too much"
  i.e.: the distribution of the estimator is hardly affected

(distribution of the estimator = performance of the estimator)

Hampel (1968)

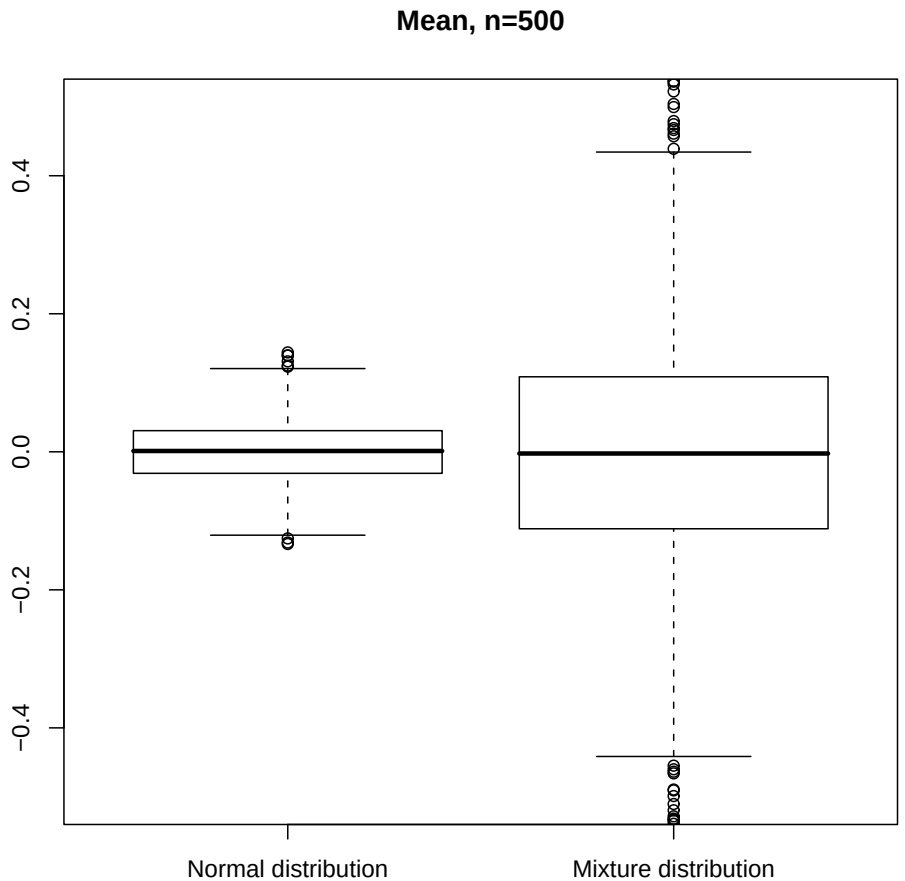
Qualitative Robustness – Example



Qualitative Robustness – Example

"mean" applied in 1000 runs

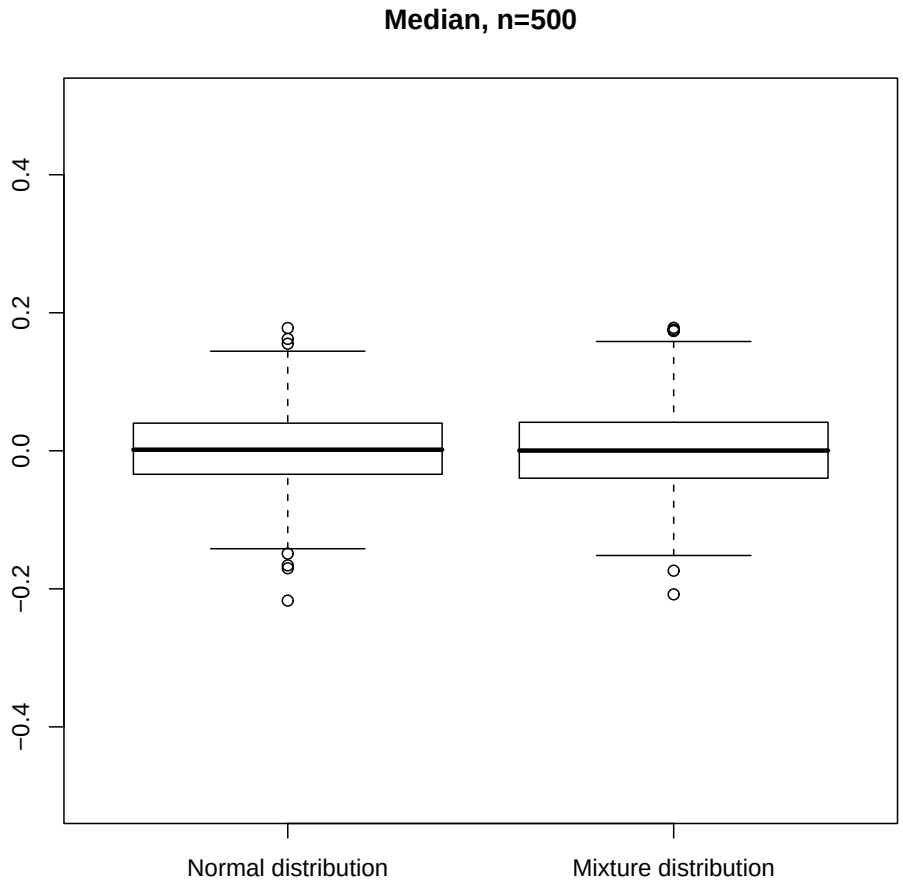
each run consists of a sample with 500 data points



Qualitative Robustness – Example

"median" applied in 1000 runs

each run consists of a sample with 500 data points



Qualitative Robustness and SVMs

Support Vector Machines

$$\begin{aligned} S_n \;:\quad (\mathcal{X}\times\mathcal{Y})^n &\to H\,, \\ \big((x_1,y_1),\ldots,(x_n,y_n)\big) &\mapsto \arg\inf_{f\in H}\ \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda\|f\|_H^2 \end{aligned}$$

with H a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$

Qualitative Robustness:

A sequence of estimators $(S_n)_{n\in\mathbb{N}}$ is called qualitatively robust if

$$\forall\, P\ \ \forall\, \epsilon > 0\ \ \exists\, \delta > 0 \text{ such that } \forall\, Q \text{ with } d_{\mathrm{Pro}}(Q,P) < \delta$$

$$\sup_{n\in\mathbb{N}} d_{\mathrm{Pro}}\big(S_n(Q^n), S_n(P^n)\big) < \varepsilon$$

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

$$S_n : \quad (\mathcal{X} \times \mathcal{Y})^n \to H, \qquad n \in \mathbb{N},$$

can be represented by a functional

$$S : \mathcal{M}_1(\mathcal{X} \times \mathcal{Y}) \to H, \qquad P \mapsto S(P),$$

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

$$S_n : \quad (\mathcal{X} \times \mathcal{Y})^n \to H, \qquad n \in \mathbb{N},$$

can be represented by a functional

$$S : \mathcal{M}_1(\mathcal{X} \times \mathcal{Y}) \to H, \qquad P \mapsto S(P),$$

i.e.

$$S_n\big((x_1, y_1), \ldots, (x_n, y_n)\big) = S\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i, y_i)}\right) \quad \forall\, (x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$$

for every $n \in \mathbb{N}$

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

$$S_n : \quad (\mathcal{X} \times \mathcal{Y})^n \to H, \qquad n \in \mathbb{N},$$

can be represented by a functional

$$S : \mathcal{M}_1(\mathcal{X} \times \mathcal{Y}) \to H, \qquad P \mapsto S(P),$$

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

$$S_n : \quad (\mathcal{X} \times \mathcal{Y})^n \to H, \qquad n \in \mathbb{N},$$

can be represented by a functional

$$S : \mathcal{M}_1(\mathcal{X} \times \mathcal{Y}) \to H, \qquad P \mapsto S(P),$$

and if

the functional S is continuous,

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

$$S_n : \quad (\mathcal{X} \times \mathcal{Y})^n \to H, \qquad n \in \mathbb{N},$$

can be represented by a functional

$$S : \mathcal{M}_1(\mathcal{X} \times \mathcal{Y}) \to H, \qquad P \mapsto S(P),$$

and if

the functional S is continuous,

then

the sequence of estimators $(S_n)_{n\in\mathbb{N}}$ is qualitatively robust.

Qualitative Robustness and SVMs

In case of support vector machines

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

put

$$S(P) \;=\; \arg\inf_{f\in H} \int L\big(y, f(x)\big)\, P\big(d(x,y)\big) \;+\; \lambda\|f\|_H^2 \qquad \forall\, P\,.$$

Qualitative Robustness and SVMs

In case of support vector machines

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

put

$$S(P) \;=\; \arg\inf_{f\in H} \int L\big(y, f(x)\big)\, P\big(d(x,y)\big) \;+\; \lambda\|f\|_H^2 \qquad \forall\, P\,.$$

It remains to be proven:

$$P_n \xrightarrow[n\to\infty]{} P_0 \;\text{ weakly} \quad\Rightarrow\quad \Big\|S(P_n) - S(P_0)\Big\|_H \xrightarrow[n\to\infty]{} 0$$

Qualitative Robustness and SVMs

Assumptions:

- $\mathcal{X}$ a Polish space, $\mathcal{Y} \subset \mathbb{R}$ closed
- $L : \mathcal{Y} \times \mathbb{R} \to [0, \infty)$ continuous
  - $t \mapsto L(y, t)$ convex for every $t \in \mathbb{R}$
  - $t \mapsto L(y, t)$, $y \in \mathcal{Y}$, uniformly Lipschitz continuous
- the reproducing kernel

$$k \; : \; \mathcal{X} \times \mathcal{X} \; \to \; \mathbb{R}$$

of the RKHS H is continuous and bounded

Qualitative Robustness and SVMs

Assumptions:

- $\mathcal{X}$ a Polish space, $\mathcal{Y} \subset \mathbb{R}$ closed
- $L : \mathcal{Y} \times \mathbb{R} \to [0, \infty)$ continuous
  - $t \mapsto L(y, t)$ convex for every $t \in \mathbb{R}$
  - $t \mapsto L(y, t)$, $y \in \mathcal{Y}$, uniformly Lipschitz continuous
- the reproducing kernel

$$k \; : \; \mathcal{X} \times \mathcal{X} \; \to \; \mathbb{R}$$

  of the RKHS H is continuous and bounded

Main Result: The functional

$$P \; \mapsto \; S(P) \; = \; \arg\inf_{f \in H} \int L\big(y, f(x)\big) \, P\big(d(x, y)\big) \; + \; \lambda \|f\|_H^2$$

is continuous.

Corollaries

Main Result: The functional

$$P \;\mapsto\; S(P) \;=\; \arg\inf_{f\in H} \int L\big(y, f(x)\big)\, P\big(d(x,y)\big) \;+\; \lambda\|f\|_H^2$$

is continuous.

Corollary 1:

Support vector machines

$$\begin{aligned} S_n \;:\quad & (\mathcal{X}\times\mathcal{Y})^n && \to && H\,, \\ & \big((x_1,y_1),\ldots,(x_n,y_n)\big) && \mapsto && \arg\inf_{f\in H} \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda\|f\|_H^2 \end{aligned}$$

are qualitatively robust.

Corollaries

Main Result: The functional

$$P \;\mapsto\; S(P) \;=\; \arg\inf_{f\in H} \int L\big(y, f(x)\big)\, P\big(d(x,y)\big) \;+\; \lambda\|f\|_H^2$$

is continuous.

Corollary 2:

Support vector machines

$$\begin{aligned} S_n \;:\quad & (\mathcal{X}\times\mathcal{Y})^n && \to && H\,, \\ & \big((x_1,y_1),\ldots,(x_n,y_n)\big) && \mapsto && \arg\inf_{f\in H} \frac{1}{n}\sum_{i=1}^{n} L\big(y_i, f(x_i)\big) \;+\; \lambda\|f\|_H^2 \end{aligned}$$

depend on the data **continuously**.

Corollaries

Main Result: The functional

$$P \;\mapsto\; S(P) \;=\; \arg\inf_{f\in H} \int L\big(y, f(x)\big)\, P\big(d(x,y)\big) \;+\; \lambda\|f\|_H^2$$

is continuous.

Corollary 3:

Strong consistency:

$$S_n \quad \xrightarrow{P-\text{a.s.}} \quad S(P) \qquad \text{for} \quad n \to \infty$$

References

- **A. Cuevas (1988)**: Qualitative robustness in abstract inference. *Journal of Statistical Planning and Inference*, 18:277–289.
- **R. Hable and A. Christmann (2010)**: Qualitative robustness of support vector machines. *Submitted.* `arXiv:0912.0874v1`
- **I. Steinwart and A. Christmann (2008)**: *Support vector machines*. Springer, New York.
- **V. N. Vapnik (1998)**: *Statistical learning theory*. John Wiley & Sons, New York.

The handout to this talk is also available on my homepage

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