GOCPS 2010, Leipzig 3 March 2010

Qualitatively Robust Nonparametric Regression By Support Vector Machines

Robert Hable, Andreas Christmann

Robert Hable
Department of Mathematics
University of Bayreuth

Nonparametric Regression

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

with

- Y : output variable (observable)
- X : input variable (observable)
- f_0 : regression function (unknown)
- ε : error term (not observable)

Goal: Estimation of the unknown regression function f_0

Qualitative Robustness

Qualitative Robustness – Parametric Statistics

Parametric Model

$$(P_\theta)_{\theta \in \Theta}\,, \qquad \Theta \subset \mathbb{R}^d$$

and

$$Z_i \;\sim\; P_{\theta_0} \quad \text{i.i.d.}\,, \qquad \theta_0 \in \Theta$$

Estimator

$$S_n \;:\; \mathcal{Z}^n \;\to\; \Theta\,, \qquad (z_1, \ldots, z_n) \;\mapsto\; S_n(z_1, \ldots, z_n)$$

Qualitative Robustness – Parametric Statistics

Parametric Model

$$(P_\theta)_{\theta\in\Theta}\,, \qquad \Theta\subset\mathbb{R}^d$$

and

$$Z_i \;\sim\; P_{\theta_0} \quad \text{i.i.d.}\,, \qquad \theta_0\in\Theta$$

Estimator

$$S_n \;:\; \mathcal{Z}^n \;\to\; \Theta\,, \qquad (z_1,\ldots,z_n) \;\mapsto\; S_n(z_1,\ldots,z_n)$$

Robustness: different notions

$$\left.\begin{array}{l}\text{small model violations}\\ \text{small errors in the data}\end{array}\right\}\ \text{should not change the result too much}$$

Qualitative Robustness – Parametric Statistics

Estimator

$$S_n \ : \ \mathcal{Z}^n \to \Theta\,, \qquad (z_1,\ldots,z_n) \mapsto S_n(z_1,\ldots,z_n)$$

Qualitative Robustness: Hampel (1968)

- 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 – Parametric Statistics

Estimator

$$S_n \;:\; \mathcal{Z}^n \;\to\; \Theta\,, \qquad (z_1,\dots,z_n) \;\mapsto\; S_n(z_1,\dots,z_n)$$

Qualitative Robustness: Hampel (1968)

- 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 result too much"

  P_θ : model distribution

  Q : real distribution

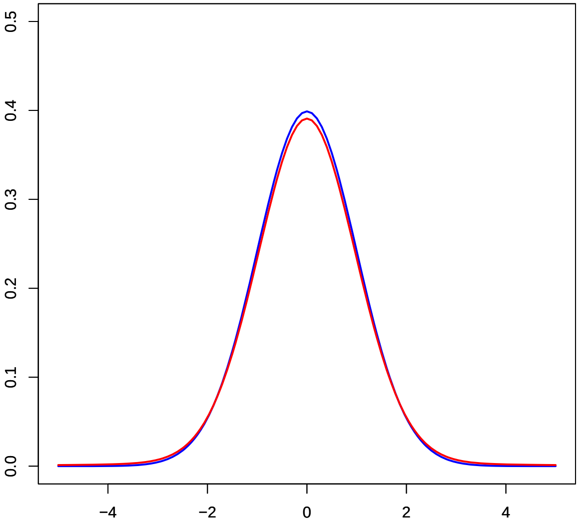
$$Q \text{ close to } P_\theta \qquad \longrightarrow \qquad S_n(Q^m) \text{ close to } S_n(P_\theta)$$

$S_n(P_\theta)$: assumed distribution of the estimator

$S_n(Q^m)$: real distribution of the estimator

(distribution of the estimator $=$ performance of the estimator)

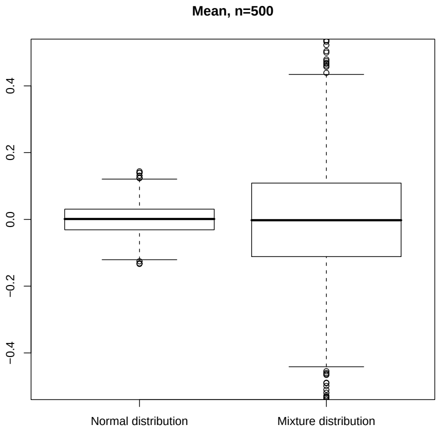
Qualitative Robustness – Example



Qualitative Robustness – Example

"mean" applied in 1000 runs

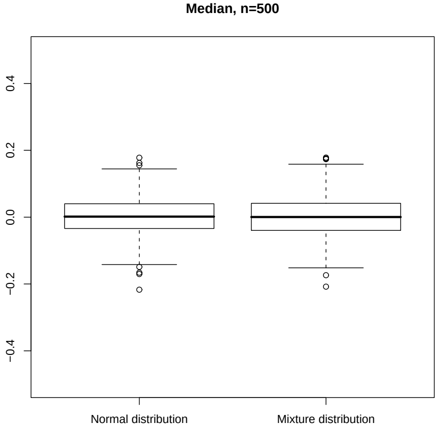
each run consists of a sample with 500 data points



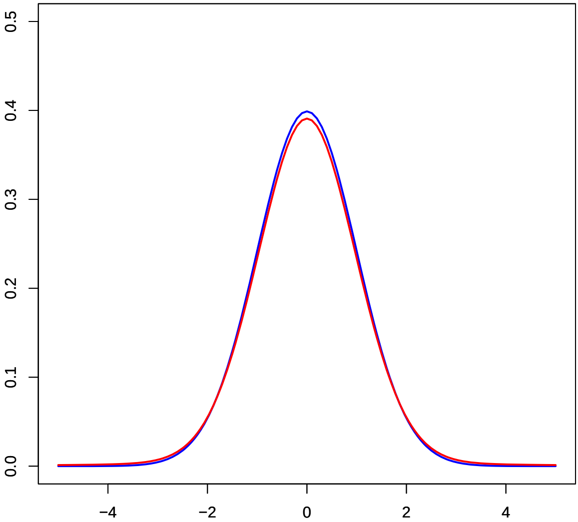
Qualitative Robustness – Example

"median" applied in 1000 runs

each run consists of a sample with 500 data points



Qualitative Robustness – Example



Qualitative Robustness – Parametric Statistics

Estimator

$$S_n \;:\; \mathcal{Z}^n \to \Theta\,, \qquad (z_1, \ldots, z_n) \mapsto S_n(z_1, \ldots, z_n)$$

Qualitative Robustness: Hampel (1968)

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

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

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

Qualitative Robustness – Parametric Statistics

Estimator

$$S_n \;:\; \mathcal{Z}^n \to \Theta\,, \qquad (z_1,\dots,z_n) \mapsto S_n(z_1,\dots,z_n)$$

Qualitative Robustness: Hampel (1968)

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

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

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

Prokhorov distance between probability measures Q and P:

$$d_{\text{Pro}}(Q,P) = \inf\big\{\delta > 0 \,\big|\, Q(A) \le P(A^\delta) + \delta \;\forall\, A\big\}$$

$$A^\delta = \big\{z \,|\, \inf_{z'\in A} d(z,z') < \delta\big\}$$

Qualitatively Robust Nonparametric Regression

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 $f_0 : \mathcal{X} \to \mathcal{Y}$

Estimator:

$$S_n \;:\; (\mathcal{X} \times \mathcal{Y})^n \;\to\; H$$

H: a large set of functions $f : \mathcal{X} \to \mathcal{Y}$

Qualitatively Robust Nonparametric Regression

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 $f_0 : \mathcal{X} \to \mathcal{Y}$

Estimator:

$$S_n \;:\; (\mathcal{X} \times \mathcal{Y})^n \;\to\; H$$

H: a large set of functions $f : \mathcal{X} \to \mathcal{Y}$

Robustness:

small model violations
small errors in the data
} should not change the result too much

Qualitatively Robust Nonparametric Regression

Nonparametric Regression

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

and

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

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

Estimator:

$$S_n \ : \ (\mathcal{X} \times \mathcal{Y})^n \ \to \ H$$

H : a large set of functions $f : \mathcal{X} \to \mathcal{Y}$

Robustness:

small model violations
small errors in the data
} should not change the result too much

Support Vector Machines

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

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

Support Vector Machines

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

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

Support Vector Machines

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

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

- 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

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

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

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

Support Vector Machines

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

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 Machines

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

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

- Loss function

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

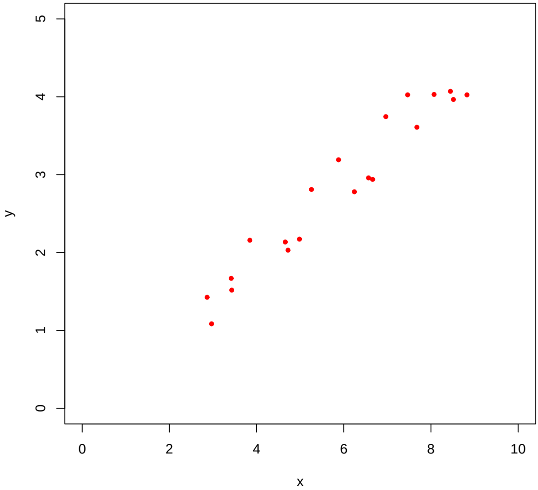
  $L(y,t)$: loss caused by estimation $t = \hat{f}_n(x)$ if y is true
- empirical Risk of an estimate $\hat{f}_n : \mathcal{X} \to \mathbb{R}$

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

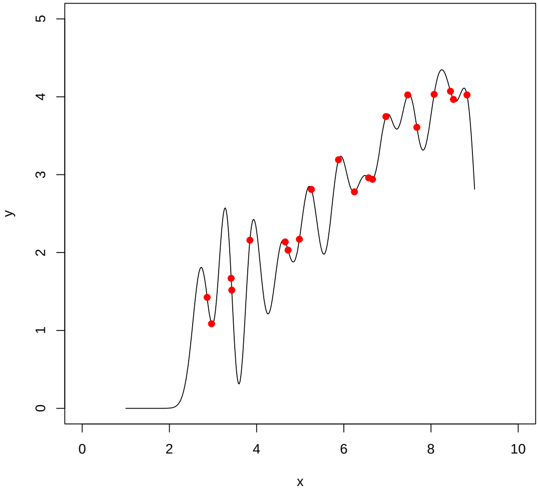
- RKHS H (certain Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$)
- Support vector machine

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

Overfitting



Overfitting



Support Vector Machines

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

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 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

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

Support Vector Machines

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

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

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

Qualitative Robustness and SVMs

Support Vector Machines

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

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

Qualitative Robustness:

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

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

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

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

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

can be represented by a functional

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

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

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

can be represented by a functional

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

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

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

can be represented by a functional

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

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

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

can be represented by a functional

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

and if

the functional S is continuous,

Qualitative Robustness and SVMs

Cuevas (1988):
If a sequence of estimators

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

can be represented by a functional

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

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

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

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

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

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

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

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

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

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

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

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:

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

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. *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