Cross-validation in Scattered Data Approximation

Felix Bartel   Ralf Hielscher   Daniel Potts

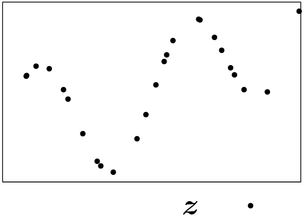
Laboratory “High-dimensional approximation and applications”
Lomonosov Moscow State University

4th of May 2021

Mathematik!
TU Chemnitz

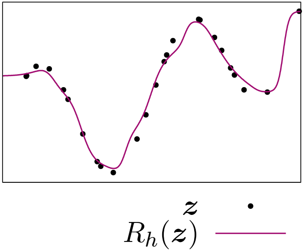
Setting

- data $\boldsymbol{z} = (\boldsymbol{x}_i, y_i)_{i=1}^n \in (\Omega \times Y)^n$
- $(\boldsymbol{x}_i, y_i)$ distributed according to ρ



Setting

- data $\boldsymbol{z} = (\boldsymbol{x}_i, y_i)_{i=1}^n \in (\Omega \times Y)^n$
- $(\boldsymbol{x}_i, y_i)$ distributed according to ρ
- reconstruction algorithm $R_h \colon (\Omega \times Y)^n \to Y^\Omega$



Problem 1: Estimate Risk

Given: data $\boldsymbol{z}$, reconstruction model $R_h(\boldsymbol{z})$

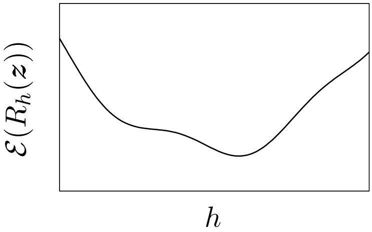
Question: How big is the L_2-reconstruction error of $R_h(\boldsymbol{z})$?

$$\mathcal{E}(R_h(\boldsymbol{z})) = \int_{\Omega \times Y} |(R_h(\boldsymbol{z}))(\boldsymbol{x}) - y|^2 \, \mathrm{d}\rho(\boldsymbol{x}, y)$$

Problem 2: Model selection

Given: data $\boldsymbol{z}$, hypothesis space $\{R_h(\boldsymbol{z}) : h \in H\}$

Question: Which reconstruction model $R_h(\boldsymbol{z})$ to choose?



Purly data-driven method to approximate the risk

- validation set

Motivation

Purly data-driven method to approximate the risk

- validation set
- ”Probably the simplest and most widly used method for estimation prediction error is cross-validation.“ [Hastie 2001]

<table>
<tr><td>test</td><td colspan="4">train</td></tr>
<tr><td>train</td><td>test</td><td colspan="3">train</td></tr>
<tr><td colspan="2">train</td><td>test</td><td colspan="2">train</td></tr>
<tr><td colspan="3">train</td><td>test</td><td>train</td></tr>
<tr><td colspan="4">train</td><td>test</td></tr>
</table>

Leave-one-out cross-validation [Stone 1974]

1. calculate the reconstructions $R_h(\boldsymbol{z}_{-i})$ from the function values $\boldsymbol{z}_{-i} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n)$
2. evaluate the residual of $R_h(\boldsymbol{z}_{-i})$ in the i-th node $\boldsymbol{x}_i$

$$|R_h(\boldsymbol{z}_{-i})(\boldsymbol{x}_i) - y_i|$$

3. calculate the mean value with respect to all nodes

$$\mathrm{CV}(\boldsymbol{z}, \lambda) := \frac{1}{n} \sum_{i=1}^{n} |R_h(\boldsymbol{z}_{-i})(\boldsymbol{x}_i) - y_i|^2$$

Table of contents

Fast computation

- **Deshpande and Girard.** Fast computation of cross-validated robust splines and other non-linear smoothing splines. 2010.
- **Mullin and Sukthankar.** Complete Cross-Validation for Nearest Neighbor Classifiers. 2000.
- **An, Liu, and Venkatesh.** Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. 2007.
- ...
- cross-validation has $4\,000\,000$ hits on Google Scholar
- fast cross-validation has $3\,000\,000$ hits on Google Scholar

Penalized least squares estimation

$$R_h(\boldsymbol{z}) = \sum_{\boldsymbol{k}\in I} \hat{g}(\boldsymbol{k})\varphi_{\boldsymbol{k}}(\cdot)$$

where

$$\begin{aligned}\hat{\boldsymbol{g}} &= \arg\min_{\hat{\boldsymbol{g}}} \|\boldsymbol{F}\hat{\boldsymbol{g}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 + \lambda \|\hat{\boldsymbol{g}}\|_{\hat{\boldsymbol{W}}}^2 \\ &= \left(\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{F} + \lambda\hat{\boldsymbol{W}}\right)^{-1} \boldsymbol{F}^*\boldsymbol{W}\boldsymbol{y}.\end{aligned}$$

for

- $\boldsymbol{F} = (\varphi_{\boldsymbol{k}}(\boldsymbol{x}_i))_{i=1,\dots,n,\boldsymbol{k}\in I}$
- $\boldsymbol{y} = (y_i)_{i=1}^n$
- $\boldsymbol{W} = \operatorname{diag}(w_i)_{i=1}^n$
- $\hat{\boldsymbol{W}} = \operatorname{diag}(\hat{w}(\boldsymbol{k}))_{\boldsymbol{k}\in I}$

Fast computation

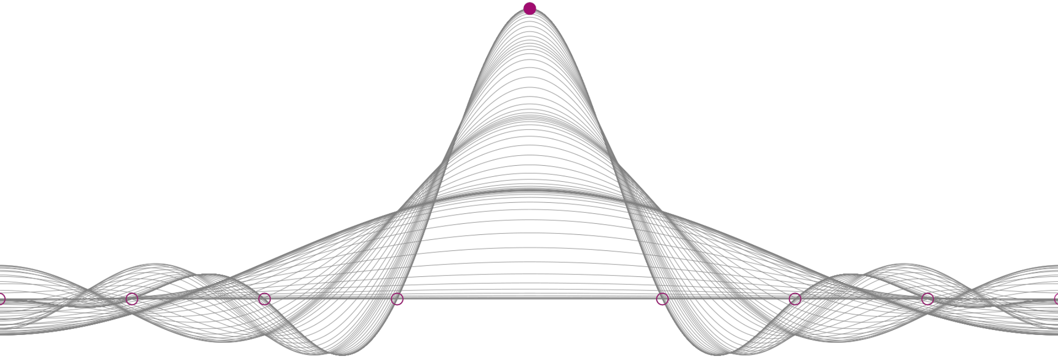
Evaluating the reconstruction for full data, we obtain

$$\begin{aligned} \boldsymbol{F}\hat{\boldsymbol{g}}_\lambda &= \boldsymbol{F}\left(\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{F} + \lambda\hat{\boldsymbol{W}}\right)^{-1}\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{y} \\ &= \boldsymbol{H}_\lambda \boldsymbol{y}. \end{aligned}$$

Theorem [Golub, Heath, Whaba 1979]

$$\mathrm{CV}(\boldsymbol{z}, \lambda) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{[\boldsymbol{H}_\lambda \boldsymbol{y} - \boldsymbol{y}]_i}{1 - h_{i,i}}\right|^2$$

with $h_{i,i}$ being the diagonal entries of $\boldsymbol{H}_\lambda$.



$$h_{i,i} = \boldsymbol{e}_i^\mathsf{T} \boldsymbol{H}_\lambda \boldsymbol{e}_i$$

Fast computation

Theorem [Tasche, Weyrich 1996]

For $\Omega = \mathbb{T}^d$, exponential functions

$$\varphi_{\boldsymbol{k}} = \exp(2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x}),$$

the full grid

$$\boldsymbol{X} = \left\{ \frac{\boldsymbol{m}}{N} : \boldsymbol{m} \in \mathbb{Z}^d \cap \prod_{t=1}^{d} [0, N) \right\}$$

the diagonal entries can be computed via

$$h_{i,i} = \frac{1}{N^d} \sum_{\boldsymbol{k} \in I} \frac{1}{1 + \lambda \hat{w}(\boldsymbol{k})}.$$

Fast computation

Theorem (Diagonal entries $h_{i,i}$ of the hat matrix) [B, Hielscher, Potts, 2019]

Let I be an index set and $\boldsymbol{X}$, $\boldsymbol{W}$ form an exact quadrature rule for $\exp(2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x})$ with $\boldsymbol{k}\in\mathcal{D}(I)=\{\boldsymbol{k}_1-\boldsymbol{k}_2:\boldsymbol{k}_1,\boldsymbol{k}_2\in I\}$. Then

$$h_{i,i}=w_i\sum_{\boldsymbol{k}\in I}\frac{1}{1+\lambda\hat{w}(\boldsymbol{k})}.$$

Fast computation

Theorem (Diagonal entries $h_{i,i}$ of the hat matrix) [B, Hielscher, Potts, 2019]

Let I be an index set and $\boldsymbol{X}$, $\boldsymbol{W}$ form an exact quadrature rule for $\exp(2\pi\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x})$ with $\boldsymbol{k}\in\mathcal{D}(I)=\{\boldsymbol{k}_1-\boldsymbol{k}_2:\boldsymbol{k}_1,\boldsymbol{k}_2\in I\}$. Then

$$h_{i,i}=w_i\sum_{\boldsymbol{k}\in I}\frac{1}{1+\lambda\hat{w}(\boldsymbol{k})}.$$

Proof:

Using quadrature, we have $\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{F}=\boldsymbol{I}$ and thus

$$\begin{aligned}h_{i,i}&=\left(\boldsymbol{F}\left(\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{F}+\lambda\hat{\boldsymbol{W}}\right)^{-1}\boldsymbol{F}^*\boldsymbol{W}\right)_{i,i}\\&=\left(\boldsymbol{F}\,\mathrm{diag}\left(\frac{1}{1+\lambda\hat{w}(\boldsymbol{k})}\right)_{\boldsymbol{k}\in I}\boldsymbol{F}^*\boldsymbol{W}\right)_{i,i}=w_i\sum_{\boldsymbol{k}\in I}\frac{1}{1+\lambda\hat{w}(\boldsymbol{k})}.\end{aligned}$$

■

Algorithm: Computation of the cross-validation score

Input:

- $\boldsymbol{X}$ • $\boldsymbol{W} = \operatorname{diag}(w_i)_{i=1}^n$ • $\hat{\boldsymbol{W}}$ • $\boldsymbol{y} = (y_i)_{i=1}^n$ • λ

Output:

- $\mathrm{CV}(\boldsymbol{z}, \lambda)$

1. Compute $\hat{\boldsymbol{g}} := \left(\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{F} + \lambda\hat{\boldsymbol{W}}\right)^{-1} \boldsymbol{F}^*\boldsymbol{W}\boldsymbol{y}$
2. Compute $\boldsymbol{g} := \boldsymbol{H}_\lambda \boldsymbol{y} = \boldsymbol{F}\hat{\boldsymbol{g}}$
3. Compute $h_{i,i} := w_i \sum_{\boldsymbol{k}\in I} \frac{1}{1+\lambda\hat{w}(\boldsymbol{k})}$ for $1 \le i \le n$
4. Evaluate $\mathrm{CV}(\boldsymbol{z}, \lambda) := \frac{1}{n}\sum_{i=1}^{n} \left|\frac{g_i - y_i}{1 - h_{i,i}}\right|^2$

Fast computation

domain Ω	basis $\varphi_{\boldsymbol{k}}$	diagonal elements $h_{i,i}$
$\mathbb{T}^d$	$\exp(2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x})$	$w_i \sum_{\boldsymbol{k} \in I} \frac{1}{1+\lambda \hat{w}(\boldsymbol{k})}$
$[0,1]$	$\cos(k \arccos x)$	$\frac{w_i}{2} \left(\frac{2}{\pi + \lambda \hat{w}_0} + \sum_{k=1}^{K-1} \frac{\cos(2k \arccos x_i) + 1}{\pi/2 + \lambda \hat{w}_k} \right)$
$\mathbb{S}^2$	$Y_{k,l}(\vartheta, \varphi)$	$\frac{w_i}{4\pi} \sum_{k=0}^{K} \frac{2k+1}{1+\lambda \hat{w}_k}$
$\mathrm{SO}(3)$	$D_k^{ll'}(\boldsymbol{R})$	$w_i \sum_{k=0}^{K} \frac{2k+1}{\frac{8\pi^2}{2k+1} + \lambda \hat{w}_k}$

[B., Hielscher, Potts 2019]

Fast matrix-vector multiplication

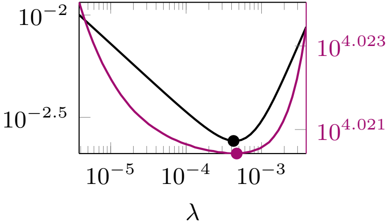
domain Ω	basis $\varphi_{\boldsymbol{k}}$	fast algorithm
$\mathbb{T}^d$	$\exp(2\pi \mathrm{i} \boldsymbol{k} \cdot \boldsymbol{x})$	FFT, LFFT, NFFT
$[0, 1]$	$\cos(k \arccos x)$	FCT, NFCT
$\mathbb{S}^2$	$Y_{k,l}(\vartheta, \varphi)$	NFSFT
$\mathrm{SO}(3)$	$D_k^{ll'}(\boldsymbol{R})$	NSOFT

Thus, $\mathrm{CV}(\boldsymbol{z}, \lambda)$ computable in $\mathcal{O}(n \log n + |I|)$.

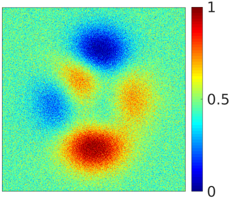
Numerical example

- 1024×1024 nodes
- $10\,\%$ Gaussian noise

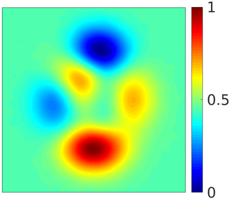
L$_2$-error $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ and
cross-validation score $\mathrm{CV}(\boldsymbol{z}, \lambda)$



noisy data



reconstruction

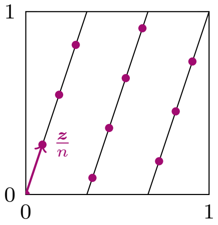


Fast computation

Example: rank-1 lattices [Kämmerer, Potts, Volkmer 2015]

- $\boldsymbol{X} = \Lambda(\boldsymbol{z}, n) = \left\{ \boldsymbol{x} = \frac{1}{n}(m\boldsymbol{z} \bmod n\boldsymbol{1}) \in \mathbb{T}^d : m = 0, \ldots, n-1 \right\}$
- Given I, there exist algorithms that find $\boldsymbol{z}$ and n such that $\boldsymbol{F}^*\boldsymbol{W}\boldsymbol{F} = \boldsymbol{I}$ for $\boldsymbol{W} = \mathrm{diag}(1/n)_{i=1}^n$.
- Multiplication with $\boldsymbol{F}$ can be carried out with the LFFT in $\mathcal{O}(n \log n)$.

$\boldsymbol{z} = (1,3)^\top, n = 11$



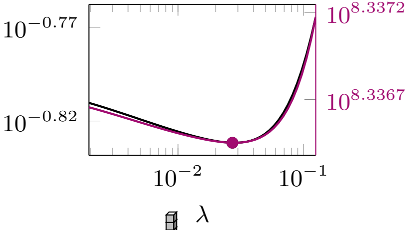
Numerical example

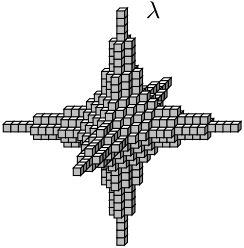
- $f = \bigotimes_{j=1}^{7} B_2$ with 5 % Gaussian noise
- $\hat{w}(\boldsymbol{k}) = \prod_{j=1}^{7} \max\left(|n_j|^2, 1\right)$
- index set

$$I_{16}^{7,\mathrm{hc}} := \left\{ \boldsymbol{k} \in \mathbb{Z}^7 : \prod_{j=1}^{7} \max(1, |n_j|) \leq 16 \right\}$$

- $\boldsymbol{X}$: 1 105 193 rank-1 lattice nodes

L$_2$-error $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ and cross-validation score $\mathrm{CV}(\boldsymbol{z}, \lambda)$





Scattered nodes

By [Tropp 2012], [Oliviera 2010], or [Rauhut 2010] we obtain for uniformly random nodes

$$\boldsymbol{F}^* \boldsymbol{W} \boldsymbol{F} \approx \boldsymbol{I}$$

with high probability, which motivates

Definition

The approximated cross-validation score $\tilde{\mathrm{CV}}(\lambda)$ is defined by

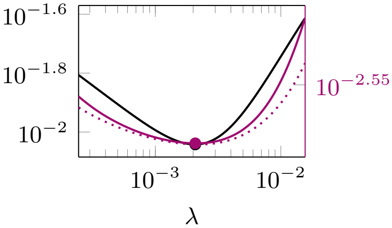
$$\tilde{\mathrm{CV}}(\lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{[\boldsymbol{H}\boldsymbol{y} - \boldsymbol{y}]_i^2}{(1 - \tilde{h}_{i,i})^2} \quad \text{for} \quad \tilde{h}_{i,i} = w_i \sum_{\boldsymbol{k} \in I} \frac{1}{1 + \lambda \hat{w}(\boldsymbol{k})} \approx h_{i,i}.$$

Numerical example

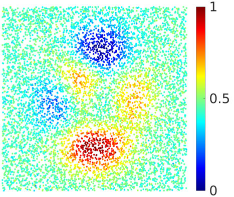
- $8\,192$ uniformly scattered nodes
- $5\,\%$ Gaussian noise

L_2-error $\|\hat{\boldsymbol{g}}_\lambda - \hat{\boldsymbol{f}}\|_2$ and

cross-validation score $\mathrm{CV}(\boldsymbol{z}, \lambda)$



noisy data



reconstruction

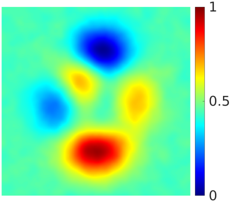


Table of contents

Problem

Given: data $\boldsymbol{z} = (\boldsymbol{x}_i, f(\boldsymbol{x}_i))_{i=1}^n$, reconstruction model $R_h(\boldsymbol{z})$

Question: Relation of

$$\mathcal{E}(R_h(\boldsymbol{z})) = \int_{\Omega\times Y} |(R_h(\boldsymbol{z}))(\boldsymbol{x}) - f(\boldsymbol{x})|^2 \, \mathrm{d}\rho(\boldsymbol{x}, y)$$

and

$$\mathrm{CV}(\boldsymbol{z}, \lambda) = \frac{1}{n} \sum_{i=1}^{n} |R_h(\boldsymbol{z}_{-i})(\boldsymbol{x}_i) - y_i|^2 \, .$$

Bound by Golub, Heath, and Whaba from 1979

For PLSE with $1/n \operatorname{trace} \boldsymbol{H}_\lambda < 1$ they showed

$$\frac{|\mathbb{E}(\mathcal{E}(R_\lambda(\boldsymbol{Z})) - \mathbb{E}(\mathrm{CV}(\boldsymbol{Z},\lambda))|}{\mathbb{E}(\mathrm{CV}(\boldsymbol{Z},\lambda))} \leq \left(\frac{2}{n}\operatorname{trace}\boldsymbol{H}_\lambda + \frac{(\operatorname{trace}\boldsymbol{H}_\lambda)^2}{\operatorname{trace}\boldsymbol{H}_\lambda^2}\right)\frac{1}{(1-1/n\operatorname{trace}\boldsymbol{H}_\lambda)^2}.$$

Bound by Li from 1987

Let

$$h^+ = \arg\min_h \mathcal{E}(R_h(\boldsymbol{z})) \quad \text{and} \quad h^\star = \arg\min_h \mathrm{CV}(\boldsymbol{z}, h).$$

Then under certain assumptions

$$\frac{\mathcal{E}(R_{h^\star}(\boldsymbol{z}))}{\mathcal{E}(R_{h^+}(\boldsymbol{z}))} \to 1 \quad \text{for} \quad n \to \infty.$$

Concentration Inequalities

Theorem (Hoeffding 1963)

Let

- $Z_1, \ldots, Z_n$ *independent rv's with values in* $[0,1]$ *and*
- $m = \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n} Z_i\right\}$.

Then for $\varepsilon > 0$

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z_i - m\right| > \varepsilon\right\} \leq 2\exp\left(-2n\varepsilon^2\right).$$

A function $f\colon \Omega^n \to \mathbb{R}$ is $\boldsymbol{c}$-bounded on $\Xi \subset \Omega^n$ for $\boldsymbol{c} = (c_1, \ldots, c_n)$, iff

$$|f(z_1, \ldots, z_n) - f(z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_n)| \leq c_i$$

for all $(z_1, \ldots, z_n), (z_1', \ldots, z_n') \in \Xi$, and $1 \leq i \leq n$.

Theoretical Foundation

Theorem (McDiarmid 1989)

Let

- $Z_1, \ldots, Z_n$ *be independent rv's with values in* Ω,
- $f : \Omega^n \to \mathbb{R}$ *be* $\boldsymbol{c}$*-bounded on* Ω^n, *and*
- $m = \mathbb{E}\{f(Z_1, \ldots, Z_n)\}$.

Then for $\varepsilon > 0$

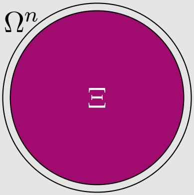
$$\mathbb{P}\{|f(Z_1, \ldots, Z_n) - m| > \varepsilon\} \leq 2 \exp\left(-\frac{2\varepsilon^2}{\|\boldsymbol{c}\|_2^2}\right).$$

Theoretical Foundation

Theorem (Extension of McDiarmid by Combes 2015)

Let

- $Z_1, \ldots, Z_n$ *be independent rv's with values in* Ω,
- $f\colon \Omega^n \to \mathbb{R}$ *be* $\boldsymbol{c}$*-bounded on* $\Xi \subset \Omega^n$, *and*
- $m = \mathbb{E}\left\{f(Z_1, \ldots, Z_n) | (Z_1, \ldots, Z_n) \in \Xi\right\}$, *and*
- $\gamma = 1 - \mathbb{P}\{(Z_1, \ldots, Z_n) \in \Xi\}$.



Then for $\varepsilon > \gamma \|\boldsymbol{c}\|_1$

$$\mathbb{P}\left\{|f(Z_1, \ldots, Z_n) - m| > \varepsilon\right\} \le 2\gamma + 2\exp\left(-\frac{2(\varepsilon - \gamma\|\boldsymbol{c}\|_1)^2}{\|\boldsymbol{c}\|_2^2}\right).$$

General framework

Define $\Xi = \Xi(C_1, C_2)$ as the set of $\boldsymbol{z} = (z_1, \ldots, z_n)$ fulfilling

1. a uniform bound on the reconstruction error

$$\|R_h(\boldsymbol{z}_{-i}) - f\|_\infty \leq C_1$$

for $1 \leq i \leq n$.

General framework

Define $\Xi = \Xi(C_1, C_2)$ as the set of $\boldsymbol{z} = (z_1, \dots, z_n)$ fulfilling

1. a uniform bound on the reconstruction error

$$\|R_h(\boldsymbol{z}_{-i}) - f\|_\infty \leq C_1$$

2. $\boldsymbol{z}_{-i} \mapsto R_h(\boldsymbol{z}_{-i})(\boldsymbol{x})$ $\boldsymbol{c}$-bounded for all $\boldsymbol{x} \in \Omega$ on Ξ with $\boldsymbol{c} = \mathbb{1}C_2$

for $1 \leq i \leq n$.

Theoretical Foundation

Risk functional $\boldsymbol{z} \mapsto \mathcal{E}(R_h(\boldsymbol{z}_{-i}))$ is $\boldsymbol{c}$-bounded on Ξ with $\boldsymbol{c} = 4C_1^2 \mathbb{1}$.

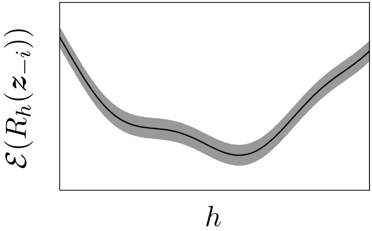
Cross-validation score $\boldsymbol{z} \mapsto \mathrm{CV}(\boldsymbol{z}, h)$ is $\boldsymbol{c}$-bounded on Ξ with $\boldsymbol{c} = 5C_1^2 \mathbb{1}$.

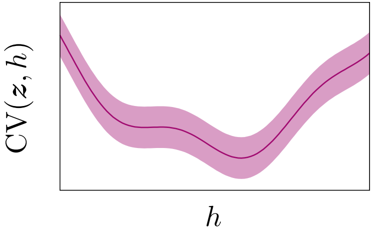
Theoretical Foundation

Risk functional $\boldsymbol{z} \mapsto \mathcal{E}(R_h(\boldsymbol{z}_{-i}))$ is $\boldsymbol{c}$-bounded on Ξ with $\boldsymbol{c} = 4C_1^2\mathbb{1}$.

Cross-validation score $\boldsymbol{z} \mapsto \mathrm{CV}(\boldsymbol{z}, h)$ is $\boldsymbol{c}$-bounded on Ξ with $\boldsymbol{c} = 5C_1^2\mathbb{1}$.

Now we apply Combes extension of McDiarmid.



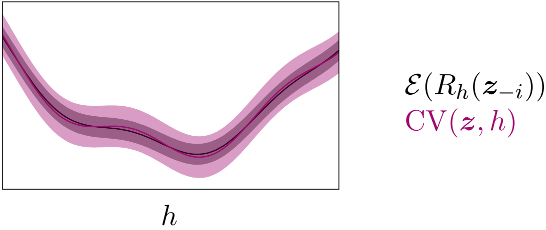


Theoretical Foundation

Lemma

For $\boldsymbol{Z}'$ representing $n-1$ samples and $\boldsymbol{Z}$ representing n samples element-wise distributed according to ρ we have

$$\mathbb{E}_{\boldsymbol{Z}'}\left\{\mathcal{E}(R_h(\boldsymbol{Z}'))\right\} = \mathbb{E}_{\boldsymbol{Z}}\left\{\mathrm{CV}(\boldsymbol{Z}, h)\right\}.$$



Theoretical Foundation

Theorem (B., Hielscher 2021)

Let $n \geq 3$*,*

- $\boldsymbol{Z}$ *be element-wise distributed according to* ρ *with values in* $(\Omega \times Y)^n$*, and*
- $\gamma = 1 - \mathbb{P}\{\boldsymbol{Z} \in \Xi\}$.

Then for $\varepsilon > 2\gamma \max\{5nC_1^2, (\|(R_h(\cdot))(\cdot)\|_\infty + \|f\|_\infty)^2\}$ *we have*

$$\mathbb{P}\left\{|\mathrm{CV}(\boldsymbol{Z}, h) - \mathcal{E}(R_h(\boldsymbol{Z}_{-1}))| > \varepsilon\right\}$$
$$\leq 2\gamma + 2\exp\left(-\left(\frac{\varepsilon}{12\sqrt{n}C_1^2} - \sqrt{2n}\gamma\right)^2\right).$$

Shepard's model

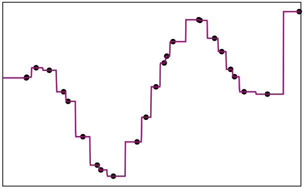
Shepard's model for $\Omega = \mathbb{T}$, $Y = \mathbb{R}$, and $\boldsymbol{z} = (x_i, f(x_i))_{i=1}^n$

$$R_h(\boldsymbol{z}) = \frac{\sum_{i=1}^m K_h(\cdot, x_i) f(x_i)}{\sum_{i=1}^m K_h(\cdot, x_i)}$$

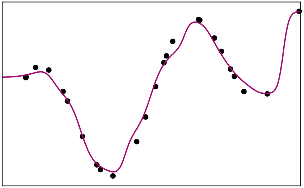
with

$$K_h(x_1, x_2) = \max\{0, 1 - h|x_1 - x_2|\}$$

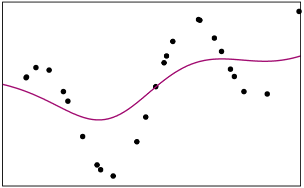
small h



medium h



big h



Theoretical Foundation

We need estimates for

- $\|(R_h(\cdot))(\cdot)\|_\infty$,
- $\|R_h(\boldsymbol{z}) - f\|_\infty$ for $\boldsymbol{z} \in \Xi$, and
- $\gamma = 1 - \mathbb{P}\{\boldsymbol{Z} \in \Xi\}$.

Theoretical Foundation

We need estimates for

- $\|(R_h(\cdot))(\cdot)\|_\infty \leq \|f\|_\infty$,
- $\|R_h(\boldsymbol{z}) - f\|_\infty$ for $\boldsymbol{z} \in \Xi$, and
- $\gamma = 1 - \mathbb{P}\{\boldsymbol{Z} \in \Xi\}$.

Theoretical Foundation

We need estimates for

- $\|(R_h(\cdot))(\cdot)\|_\infty \leq \|f\|_\infty$,
- $\|R_h(\boldsymbol{z}) - f\|_\infty$ for $\boldsymbol{z} \in \Xi$, and
- $\gamma = 1 - \mathbb{P}\{\boldsymbol{Z} \in \Xi\}$.

Theorem (B., Hielscher 2021)

Let $x_1, \ldots, x_n$ be uniformly random, $K_h(x, \cdot)$ be supported on $[x - 1/h, x + 1/h]$, and f be Lipschitz with constant L. Then $\|R_h(\boldsymbol{z}) - f\|_\infty \leq L/h$ with probability

$$1 - \gamma \geq \sum_{k=0}^{\lfloor h \rfloor} (-1)^{k+1} \binom{n}{k} \left(1 - \frac{k}{2h}\right)^{n-1}$$

Theoretical Foundation

Theorem (B., Hielscher 2021)

Let

- f *be Lipschitz with constant* L,
- $\boldsymbol{z} = (x_i, f(x_i))_{i=1}^n$ *be samples with* $x_1, \ldots, x_n \in \mathbb{T}$ *uniformly distributed,*
- $K_h(x, \cdot)$ *be supported on* $[x - 1/h, x + 1/h]$, *and*
- $\gamma = \sum_{k=1}^{\lfloor h \rfloor} (-1)^{k+1} \binom{n}{k} \left(1 - \frac{k}{2h}\right)^{n-1}$.

Then we have for $\varepsilon > \gamma \max\{10nL^2/h^2, 8\|f\|_\infty^2\}$

$$\mathbb{P}\left\{|\mathrm{CV}(\boldsymbol{Z}, h) - \mathcal{E}(R_h(\boldsymbol{Z}_{-1}))| > \varepsilon\right\}$$
$$\leq 2\gamma + 2\exp\left(-\left(\frac{h^2\varepsilon}{12\sqrt{n}L^2} - \sqrt{2n}\gamma\right)^2\right).$$

Györfi, Kohler, Krzyżak, and Walk in 2002

Using

- binary kernels $K_h : \Omega \times \Omega \to \{0, 1\}$
- Ξ all possible samples

they showed

$$\gamma = 0, \quad C_1 = 2\|f\|_\infty, \quad C_2 \sim \frac{1}{n}$$

and obtained

$$\mathbb{P}\left\{|\mathrm{CV}(\boldsymbol{Z}, h) - \mathcal{E}(R_h(\boldsymbol{Z}_{-1}))| > \varepsilon\right\} \lesssim \exp\left(-\left(\sqrt{n}\varepsilon\right)^2\right).$$

Györfi, Kohler, Krzyżak, and Walk in 2002

Using

- binary kernels $K_h : \Omega \times \Omega \to \{0, 1\}$
- Ξ all possible samples

they showed

$$\gamma = 0, \quad C_1 = 2\|f\|_\infty, \quad C_2 \sim \frac{1}{n}$$

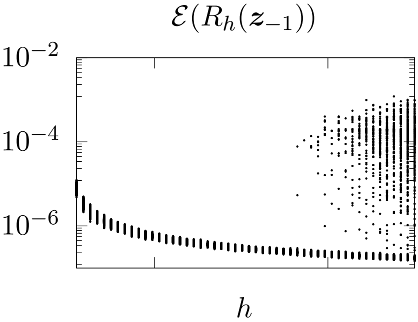
and obtained

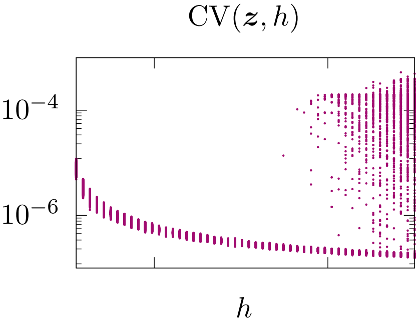
$$\mathbb{P}\left\{|\mathrm{CV}(\boldsymbol{Z}, h) - \mathcal{E}(R_h(\boldsymbol{Z}_{-1}))| > \varepsilon\right\} \lesssim \exp\left(-\left(\sqrt{n}\varepsilon\right)^2\right).$$

Our result assuming $h \sim n$:

$$\mathbb{P}\left\{|\mathrm{CV}(\boldsymbol{Z}, h) - \mathcal{E}(R_h(\boldsymbol{Z}_{-1}))| > \varepsilon\right\} \lesssim \gamma + \exp\left(-\left(n^{3/2}\varepsilon + \sqrt{n}\gamma\right)^2\right).$$

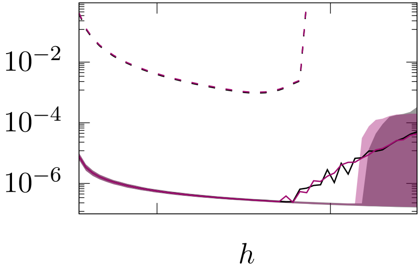
Numerical example





Numerical example

$\mathrm{CV}(\boldsymbol{z}, h)$ and $\mathcal{E}(R_h(\boldsymbol{z}_{-1}))$



$|\,\mathrm{CV}(\boldsymbol{z}, h) - \mathcal{E}(R_h(\boldsymbol{z}_{-1}))|$

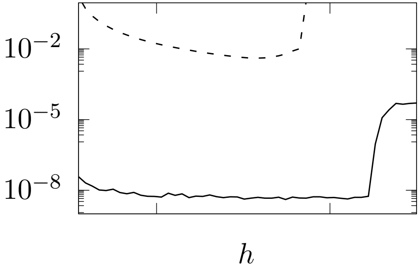


Table of contents

1. Fast computation (for PLSE) ✓
2. Theoretical foundation (for Shepard's model) ✓