Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Deterministic Error Analysis of Support Vector Regression and Related Regularized Kernel Methods

by

Christian Rieger, and Barbara Zwicknagl

Preprint no.: 152 2006

Deterministic Error Analysis of Support Vector Regression and Related Regularized Kernel Methods

Christian Rieger* and Barbara Zwicknagl†

December 15, 2006

Abstract

This paper introduces a new technique for the analysis of kernel-based regression problems. The basic tools are sampling inequalities which apply to all machine learning problems involving penalty terms induced by kernels related to Sobolev spaces. They lead to explicit deterministic results concerning the worst case behaviour of ϵ- and ν-SVRs. Using these, we show how to adjust regularization parameters to get best possible approximation orders for regression. The results are illustrated by some numerical examples.

Keywords: Sampling inequality, radial basis functions, approximation theory, reproducing kernel Hilbert space, Sobolev space

1 Introduction

Support Vector (SV) machines and related kernel-based algorithms are modern learning systems motivated by results of statistical learning theory [9]. The concept of SV machines is to provide a prediction function which is accurate on the given training data and which is sparse in the sense that it can be written in terms of a typically small subset [6] of all examples, called the support vectors. Therefore, SV regression and classification algorithms are closely related to regularized problems from classical approximation theory [3], and techniques from functional analysis were applied to derive probabilistic error bounds for SV regression [2].

*Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Lotzestr. 16-18, 37083 Göttingen, Germany

†Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany

The purpose of this paper is to provide a theoretical framework to derive deterministic error bounds for some popular SV machines. We show how a sampling inequality by [11] can be used to bound the worst-case generalization error for the ν- and the ϵ-regression without making any statistical assumptions on the inaccuracy of the training data. In contrast to the literature, our error bounds explicitly depend on the pointwise noise in the data. Thus they can be used for any subsequent probabilistic analysis modelling certain assumptions on the noise distribution.

The paper is organized as follows. In section 2 we review some facts about regularized approximation problems in Hilbert spaces with reproducing kernels and outline the connection to classical SV regression (SVR) problems. We provide a deterministic error analysis for the ν- and the ϵ-SVR for both exact and inexact training data. Our analytical results showing optimal convergence order in Sobolev spaces are confirmed by numerical experiments.

2 Regularized Problems

We suppose K to be a positive definite kernel on some domain $\Omega \subset \mathbb{R}^d$ and denote the native space, which is the unique associated reproducing kernel Hilbert space, with $\mathcal{N}_K := \mathcal{N}_K(\Omega)$ [10]. In the following we always think of native Hilbert spaces as Sobolev spaces with the usual inner product. We consider the following learning or recovery problem. We assume that we are given (possibly only approximate) function values $y_1, \ldots, y_N \in \mathbb{R}$ of an unknown function $f \in \mathcal{N}_K$ on some scattered points $x_1, \ldots, x_N \in \Omega$, i.e., $f(x_j) \approx y_j$ for $j = 1, \ldots, N$. To control accuracy and complexity of the reconstruction simultaneously, we use the optimization problem

$$\min_{\substack{f \in \mathcal{N}_K \\ (\epsilon \in \mathbb{R}^+)}} \frac{1}{N} \sum_{j=1}^{N} V_\epsilon\left(|f(x_j) - y_j|\right) + \frac{1}{2C} \|f\|_{\mathcal{N}_K}^2 \ , \tag{1}$$

where $C > 0$ is a positive parameter and and V_ϵ denotes a positive function which is parametrized by a positive real number ϵ. We point out that V_ϵ need not be a classical loss function. Therefore we shall give some proofs of results which are well-known [7] in the case of V_ϵ being a loss function. By (ϵ) we denote that there might be a primal variable ϵ included in (1) or not.

Theorem 2.1 (Representer theorem) *If $(f^*, (\epsilon^*))$ is a solution of the optimization problem (1), then there exists a vector $w \in \mathbb{R}^N$ such that*

$$f^*(\cdot) = \sum_{j=1}^{N} w_j K(x_j, \cdot) \ ,$$

i.e., $f^* \in span\{K(x_1, \cdot), \ldots, K(x_N, \cdot)\}$.

Proof: Every $f \in \mathcal{N}_K$ can be decomposed into two parts

$$f = f_{||} + f_\perp \ ,$$

where $f_{||}$ is contained in the linear span of $K(x_1, \cdot), \dots, K(x_N, \cdot)$ and $f_\perp$ is contained in the orthogonal complement, i.e., $\langle f_{||}, f_\perp \rangle_{\mathcal{N}_K} = 0$. By the reproducing property of the kernel K in the native space we have

$$f(x_j) = \langle f_{||} + f_\perp, K(x_j, \cdot) \rangle_{\mathcal{N}_K} = \langle f_{||}, K(x_j, \cdot) \rangle_{\mathcal{N}_K} \ .$$

Using this identity (1) can be rewritten as

$$\min_{\substack{f = f_{||} + f_\perp \\ (\epsilon \in \mathbb{R}^+)}} \frac{1}{N} \sum_{j=1}^{N} V_\epsilon \left(\left| \langle f_{||}, K(x_j) \rangle - y_j \right| \right) + \frac{1}{2C} \left\| f_{||} \right\|_{\mathcal{N}_K}^2 + \frac{1}{2C} \left\| f_\perp \right\|_{\mathcal{N}_K}^2 \ .$$

Therefore a solution $(f^*, (\epsilon^*))$ of the optimization problem (1) satisfies $f_\perp^* = 0$, i.e. $f^* \in \text{span}\{K(x_1, \cdot), \dots, K(x_N, \cdot)\}$. □

We shall use the representer theorem to reformulate infinite-dimensional optimization problems of the form (1) in a finite-dimensional setting.

3 Support Vector Regression

As a first optimization problem of the form (1) we will consider the ν-SVR in Hilbert space formulation. The function

$$V_\epsilon(x) = |x|_\epsilon + \epsilon\nu$$

is related to Vapnik's ϵ-intensive loss function

$$|x|_\epsilon = \begin{cases} 0 & if\, |x| \le \epsilon \\ |x| - \epsilon & if\, |x| > \epsilon \end{cases} ,$$

but has an additional term with a positive parameter ν. The associated optimization problem takes the form

$$\min_{\substack{f \in \mathcal{N}_K \\ \epsilon \in \mathbb{R}^+}} \frac{1}{N} \sum_{j=1}^{N} |f(x_j) - y_j|_\epsilon + \epsilon\nu + \frac{1}{2C} \|f\|_{\mathcal{N}_K}^2 \ . \tag{2}$$

Theorem 3.1 *The optimization problem (2) possesses a solution* (f^*, ϵ^*).

Proof: The problem (2) is equivalent to the optimization problem

$$\min_{\substack{f\in\mathcal{N}_K\\ \delta\in\mathbb{R}}} \frac{1}{N}\sum_{j=1}^{N} |f(x_j)-y_j|_{\delta^2} + \delta^2\nu + \frac{1}{2C}\|f\|^2_{\mathcal{N}_K} \ . \tag{3}$$

If we set $\mathcal{H} := \mathcal{N}_K \times \mathbb{R}$ we can define an inner product on $\mathcal{H}$ by

$$\langle h_1, h_2\rangle_{\mathcal{H}} := \langle f_1, f_2\rangle_{\mathcal{N}_K} + 2C\nu \langle r_1, r_2\rangle_{\mathbb{R}}$$

for $h_j = (f_j, r_j)$, $j = 1, 2$. Since $\mathbb{R}$ can be identified canonically with the space of all constant functions $\mathbb{R} \to \mathbb{R}$, the Hilbert space $\mathcal{H}$ has the reproducing kernel $\tilde{K} := \left(K, \frac{1}{2C\nu}\mathbf{1}\right)$ where $\mathbf{1}$ denotes the constant function which maps everything to 1, i.e. $\tilde{K}((x,r),(y,s)) = K(x,y) + 1/(2C\nu)$. With this notation (3) can be rewritten as

$$\min_{(f,\delta)\in\mathcal{H}} Q^y(I_X(f,\delta)) + \frac{1}{2C}\|(f,\delta)\|^2_{\mathcal{H}} \ , \tag{4}$$

where

$$I_X(f,\delta) := (f(x_1),\ldots,f(x_N),\delta)^T \in \mathbb{R}^{N+1}$$

and

$$Q^y : \mathbb{R}^{N+1} \to \mathbb{R}, \quad Q^y((x,\delta)) = \frac{1}{N}\sum_{j=1}^{N} |x_j - y_j|_{\delta^2} \ .$$

Since Q^y is continuous on $\mathbb{R}^{N+1}$ for all $y \in \mathbb{R}^N$, the problem (4) possesses a solution [4, Lemma 1]. □

If we introduce the slack variables ξ, ξ^*, the representer theorem gives us an equivalent well known [8] finite-dimensional problem.

$$\begin{aligned}
\min_{\substack{\mathbf{w}\in\mathbb{R}^{\mathbf{N}}\\ \xi^*,\xi\in\mathbb{R}^{\mathbf{N}}\\ \epsilon\in\mathbb{R}^+}} \ & \frac{1}{2}\mathbf{w}^T\mathbf{K}\mathbf{w} + C\cdot\left(\nu\epsilon + \frac{1}{N}\sum_{j=1}^{N}(\xi_j + \xi_j^*)\right)\\
\text{subject to} \quad & (\mathbf{K}\mathbf{w})_j - y_j \ \leq\ \epsilon + \xi_j \ ,\\
& (-\mathbf{K}\mathbf{w})_j + y_j \ \leq\ \epsilon + \xi_j^* \ ,\\
& \xi_j^*, \xi_j \geq 0\ , \qquad \epsilon \geq 0\ ,
\end{aligned} \tag{5}$$

where

$$\mathbf{K} = \Big(K(x_i, x_j)\Big)_{j=1\ldots N}$$

denotes the Gram matrix of the kernel K. We will use this equivalent problem for implementation and our numerical tests.

A particularly interesting problem arises if we skip the parameter ν and let ϵ be fixed. Then (5) takes the form

$$\begin{aligned} \min_{\substack{\mathbf{w}\in\mathbb{R}^{\mathbf{N}}\\ \xi^*,\xi\in\mathbb{R}^{\mathbf{N}}}} & \frac{1}{2}\mathbf{w}^T\mathbf{K}\mathbf{w} + C \cdot \frac{1}{N}\sum_{j=1}^{N}\left(\xi_j + \xi_j^*\right) \\ \text{subject to} \quad & (\mathbf{K}\mathbf{w})_j - y_j \leq \epsilon + \xi_j \ , \\ & (-\mathbf{K}\mathbf{w})_j + y_j \leq \epsilon + \xi_j^* \ , \\ & \xi_j^*, \xi_j \geq 0 \ . \end{aligned} \tag{6}$$

This problem is well known as ϵ-SVR [8]. Similarily to the ν-SVR this problem can be formulated as a regularized minimization problem in a Hilbert space [2], namely

$$\min_{f\in\mathcal{N}_K} \frac{1}{N}\sum_{j=1}^{N}\left|f(x_j) - y_j\right|_\epsilon + \frac{1}{2C}\|f\|_{\mathcal{N}_K}^2 \ . \tag{7}$$

Like the ν-SVR, this optimization problem possesses a solution [4, Lemma 1].

4 A Sampling inequality

We shall employ a special case of a *sampling inequality* from [11]. It requires the following assumptions which we need from now on. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary that satisfies an interior cone condition. Suppose further that K is a radial basis function such that the native Hilbert space of K is a Sobolev space, i.e., $\mathcal{N}_K = W_2^\tau(\Omega)$. Here we assume that $\lfloor \tau - 1 \rfloor > d/2$. Furthermore, let $X = \{x_1, \ldots, x_N\} \subset \Omega$ be a discrete set with sufficiently small fill distance

$$h_{X,\Omega} := \sup_{x\in\Omega}\min_{x_j\in X}\|x - x_j\|_2 \ .$$

We shall use the following result from [11].

Theorem 4.1 *Let τ be a positive real number with $\lfloor \tau - 1 \rfloor > \frac{d}{2}$ and $1 \leq q \leq \infty$. Then there exists a positive constant $C > 0$ such that for all discrete sets $X \subset \Omega$ with sufficiently small fill distance $h_{X,\Omega}$ the inequality*

$$\|u\|_{L_q(\Omega)} \leq C \cdot \left(h^{\tau - d(1/2-1/q)_+}\|u\|_{W_2^\tau(\Omega)} + \|u|_X\|_{\ell_\infty(X)}\right)$$

holds for all $u \in W_2^\tau(\Omega)$.

We shall apply this theorem to the difference $f - f^*$ of the function $f \in W_2^\tau(\Omega)$ to be recovered and a solution $f^* \in W_2^\tau(\Omega)$ of the regression problem. In our applications we shall focus on the two main cases $q = \infty$ and $q = 2$. It will turn out that we get optimal convergence rates in the noiseless case. In presence of noise the resulting error will explicitly be bounded in terms of the noise in the data.

5 ν-SVR with exact data

In this section we assume that our given data is exact, i.e.,

$$f(x_j) = y_j \quad \text{for } j = 1, \ldots, N, \tag{8}$$

where $f \in W_2^\tau(\Omega)$.

Lemma 5.1 *Under the assumptions (8) we get*

$$\begin{aligned} \|f^*\|_{\mathcal{N}_K} &\leq \|f\|_{\mathcal{N}_K} \\ \|f^*|_X - y\|_{\ell_\infty(X)} &\leq \frac{N}{2C}\|f\|^2_{\mathcal{N}_K} + \epsilon^* \cdot (1 - N\nu) \,. \end{aligned}$$

Proof: We denote the objective function of (2) by

$$H^y_{C,\nu}(f,\epsilon) := \frac{1}{N}\sum_{j=1}^{N} |f(x_j) - y_j|_\epsilon + \nu\epsilon + \frac{1}{2C}\|f\|^2_{\mathcal{N}_K} \,, \tag{9}$$

and the interpolant to f with respect to X and K with I_f, that is $I_f|_X = y$. With this notation we have

$$\frac{1}{2C}\|f^*\|^2_{\mathcal{N}_K} \leq H^y_{C,\nu}(f^*,\epsilon^*) \leq H^y_{C,\nu}(I_f,0) \leq \frac{1}{2C}\|I_f\|^2_{\mathcal{N}_K} \leq \frac{1}{2C}\|f\|^2_{\mathcal{N}_K}$$

since $\|I_f\|_{\mathcal{N}_K} \leq \|f\|_{\mathcal{N}_K}$ [10] which implies the first claim. Furthermore we have for $i = 1, \dot{,} N$

$$\begin{aligned} |f^*(x_i) - y_i| &\leq \sum_{j=1}^{N} |f^*(x_j) - y_j|_{\epsilon^*} + \epsilon^* \leq N H^y_{C,\nu}(f^*,\epsilon^*) + \epsilon^* \cdot (1 - N\nu) \\ &\leq N H^y_{C,\nu}(I_f,0) + \epsilon^*(1 - N\nu)_+ \leq \frac{N}{2C}\|I_f\|^2_{\mathcal{N}_K} + \epsilon^*(1 - N\nu) \\ &\leq \frac{N}{2C}\|f\|^2_{\mathcal{N}_K} + \epsilon^*(1 - N\nu) \end{aligned}$$

which finishes the proof. □

With Theorem 4.1 we find immediately the following result.

Theorem 5.2 *We suppose $f \in W_2^\tau(\Omega)$ with $f(x_i) = y_i$. Let (f^*, ϵ^*) be a solution of (2) . Then there is a constant $\tilde{C} > 0$ such that the approximation error can be bounded by*

$$\|f - f^*\|_{L_q(\Omega)} \leq \tilde{C}\left(2h^{\tau - d(1/2 - 1/q)_+}\|f\|_{W_2^\tau(\Omega)} + \frac{N}{2C}\|f\|^2_{W_2^\tau(\Omega)} + (1 - N\nu) \cdot \epsilon^*\right).$$

Proof: Combining Lemma 5.1 and Theorem 4.1 leads to

$$\begin{aligned}\|f-f^*\|_{L_q(\Omega)} &\le \tilde{C}\left(h^{\tau-d(1/2-1/q)_+}\|f-f^*\|_{W_2^\tau(\Omega)}+\|y-f^*|_X\|_{\ell_\infty(X)}\right)\\ &\le \tilde{C}\left(h^{\tau-(d/2-d/q)_+}(\|f\|_{W_2^\tau(\Omega)}+\|f^*\|_{W_2^\tau(\Omega)})+\|y-f^*|_X\|_{\ell_\infty(X)}\right)\\ &\le \tilde{C}\left(2h^{\tau-(d/2-d/q)_+}\|f\|_{W_2^\tau(\Omega)}+\frac{N}{2C}\|f\|^2_{W_2^\tau(\Omega)}+(1-N\nu)\epsilon^*\right)\end{aligned}$$

□

At first glance the term containing ϵ^* seems to be odd because it could be uncontrollable. But according to [1] we can at least assume ϵ^* to be bounded by

$$\epsilon^* \le \frac{1}{2}\left(\max_{i=1\ldots N} y_i - \min_{i=1\ldots N} y_i\right) .$$

If this inequality is not satisfied, the problem (5) possesses only the trivial solution $s \equiv 0$ which is not interesting. Furthermore, the factor $(1-N\nu)$ controls this term. If we choose $\nu \ge \frac{1}{N}$, the additional term vanishes or is negative. In case of a nontrivial solution this condition is no restriction at all since ν is a lower bound on the fraction of support vectors [7] and $\nu = 1/N$ means to have at least one support vector. But we can use the results from Lemma 5.1 to derive a more explicit upper bound on $\epsilon^* = \epsilon^*(C,\nu,f)$.

$$0 \le \|f^\star|_X - y\|_{\ell_\infty(X)} \le \frac{N}{2C}\|f\|^2_{\mathcal{N}_K} + \epsilon^*(C,\nu,f)(1-N\nu)$$

If we assume $\nu > 1/N$, this leads to

$$\epsilon^*(C,\nu,f) \le \frac{N}{2C(N\nu-1)}\|f\|^2_{\mathcal{N}_K}$$

Note that these bounds can not be used for a better parameter choice, since we would need to rearrange this inequality and solve for C or ν. This is only possible if there were lower bounds on ϵ^* as well.
Moreover, the parameter C appears in our error bound as a factor $\frac{N}{2C}$ which implies that we expect convergence only in the case $C \to \infty$. In this case ϵ^* will be small, as can be deduced from (5).

Corollary 5.3 *In case of quasi-uniform exact data we can choose the poblem parameters as*

$$C = \frac{N\|f\|_{W_2^\tau(\Omega)}}{2h^\tau} \approx h^{-(\tau+d)}\|f\|_{W_2^\tau(\Omega)} \text{ and } \nu \ge \frac{1}{N} ,$$

to get

$$\|f-f^*\|_{L_2(\Omega)} \le \tilde{C}h^\tau\|f\|_{W_2^\tau(\Omega)} , \tag{10}$$

or as

$$C = \frac{N\|f\|_{W_2^\tau(\Omega)}}{2h^{\tau-d/2}} \approx h^{-(\tau+d/2)}\|f\|_{W_2^\tau(\Omega)} \text{ and } \nu \geq \frac{1}{N} ,$$

to get

$$\|f - f^*\|_{L_\infty(\Omega)} \leq \tilde{C}h^{\tau-d/2}\|f\|_{W_2^\tau(\Omega)} . \tag{11}$$

Therefore the solution of the ν-SVR leads to the optimal approximation order in the Sobolev space [5] with respect to the fill distance h. These optimal rates are also attained by classical interpolation in the native Hilbert Space [10]. But the ν-SVR allows for much more flexibility and less complicated solutions. Our numerical results will confirm these convergence rates.

6 ν-SVR with inexact data

In this section we allow the given data to be corrupted by some additive error $r = (r_1, \ldots, r_N)$, i.e.,

$$f(x_j) = y_j + r_j \quad \text{for } j = 1, \ldots, N, \tag{12}$$

where is $f \in W_2^\tau(\Omega)$. There are no assumptions concerning the error distribution.

Lemma 6.1 *Under the assumption (12) we have for all* $\epsilon \geq 0$

$$\begin{aligned} \|f^*\|_{\mathcal{N}_K} &\leq \sqrt{\frac{2C}{N}\sum_{j=1}^N |r_j|_\epsilon + 2C\nu\epsilon + \|f\|^2_{\mathcal{N}_K}} \text{ and} \\ \|f^* - y\|_{\ell_\infty(X)} &\leq \sum_{j=1}^N |r_j|_\epsilon + \nu N\epsilon + (1 - N\nu)\epsilon^* + \frac{N}{2C}\|f\|^2_{\mathcal{N}_K}. \end{aligned}$$

Proof: Again, we denote the interpolant to f with respect to X and K by I_f and use $H^y_{C,\nu}$ as defined in (9). Then we have

$$\frac{1}{2C}\|f^*\|^2_{\mathcal{N}_K} \leq H^y_{C,\nu}(f^*, \epsilon^*) \leq H^y_{C,\nu}(I_f, \epsilon) \leq \frac{1}{N}\sum_{j=1}^N |r_j|_\epsilon + \nu\epsilon + \frac{1}{2C}\|f\|^2_{\mathcal{N}_K}$$

which implies

$$\|f^*\|_{\mathcal{N}_K} \leq \sqrt{\frac{2C}{N}\sum_{j=1}^N |r_j|_\epsilon + 2C\nu\epsilon + \|f\|^2_{\mathcal{N}_K}} .$$

Moreover we have for all $i = 1, \dot{;} N$

$$\begin{aligned} |f^*(x_i) - y_i| &\leq \sum_{j=1}^{N} |f^*(x_j) - y_j|_{\epsilon^*} + \epsilon^* \\ &\leq NH^y_{C,\nu}(f^*, \epsilon^*) + (1 - N\nu)\epsilon^* \\ &\leq \sum_{j=1}^{N} |r_j|_\epsilon + \nu N\epsilon + (1 - N\nu)\epsilon^* + \frac{N}{2C}\|f\|^2_{\mathcal{N}_K}. \end{aligned}$$

□

Again we can use the results from Lemma 6.1 to derive a more explicit upper bound on $\epsilon^* = \epsilon^*(C, \nu, f, \epsilon)$. Note that ϵ^* depends now also on the free parameter ϵ.

$$0 \leq \|f^\star|_X - y\|_{\ell_\infty(X)} \leq \frac{N}{2C}\|f\|^2_{\mathcal{N}_K} + \epsilon^*(C, \nu, f, \epsilon)(1 - N\nu) + \sum_{j=1}^{N} |r_j|_\epsilon + \nu N\epsilon$$

If we assume $\nu > 1/N$, this leads to

$$\epsilon^*(C, \nu, f, \epsilon) \leq \frac{1}{N\nu - 1}\left(\frac{N}{2C}\|f\|^2_{\mathcal{N}_K} + \sum_{j=1}^{N} |r_j|_\epsilon + \nu N\epsilon\right)$$

Using the sampling inequality as in the case of exact data leads to the following result on L_q-norms.

Theorem 6.2 *Under the assumption (12) we have for all* $\epsilon > 0$

$$\begin{aligned} \|f - f^*\|_{L_q(\Omega)} \leq \tilde{C}\Bigg(h^{\tau-(d/2-d/q)_+}\left(\|f\|_{W_2^\tau(\Omega)} + \sqrt{\frac{2C}{N}\sum_{j=1}^{N}|r_j|_\epsilon + 2C\nu\epsilon + \|f\|^2_{W_2^\tau(\Omega)}}\right) \\ + \sum_{j=1}^{N} |r_j|_\epsilon + \nu N\epsilon + \epsilon^*(1 - N\nu) + \frac{N}{2C}\|f\|^2_{W_2^\tau(\Omega)}\Bigg) \ . \end{aligned}$$

□

We now want to assume that the data errors do not exceed the data itself. For this we suppose

$$\|r\|_{\ell_\infty(X)} \leq \delta \leq \|f\|_{W_2^\tau(\Omega)} \ , \qquad (13)$$

for a $\delta > 0$.

Corollary 6.3 *If we choose*

$$
\begin{aligned}
C &= \frac{N\|f\|^2_{W_2^\tau(\Omega)}}{2\delta} \approx \frac{h^{-d}}{\delta}\|f\|^2_{W_2^\tau(\Omega)} \\
\epsilon &= \delta, \quad \nu = \frac{1}{N}
\end{aligned}
$$

we get

$$\|f - f^*\|_{L_2(\Omega)} \leq \tilde{C}\left(h^\tau \|f\|_{W_2^\tau(\Omega)} + \delta\right). \tag{14}$$

or

$$\|f - f^*\|_{L_\infty(\Omega)} \leq \tilde{C}\left(h^{\tau-d/2} \|f\|_{W_2^\tau(\Omega)} + \delta\right). \tag{15}$$

for any non-trivial solution.

7 ϵ-SVR with exact data

Since our arguments for the ν-SVR apply similarily to the ϵ-SVR, we skip over details and just state the results. Note that in this case the non-negative parameter ϵ is fixed in contrast to the free variable in the ν-SVR.

Lemma 7.1 *Under the assumption (8) we get*

$$
\begin{aligned}
\|f^*\|_{\mathcal{N}_K} &\leq \|f\|_{\mathcal{N}_K} \\
\|f^*|_X - y\|_{\ell_\infty(X)} &\leq \frac{N}{2C}\|f\|^2_{\mathcal{N}_K} + \epsilon\,.
\end{aligned}
$$

Again this leads to the following result on continuous L_q-norms.

Theorem 7.2 *Under the assumption (8) we get*

$$\|f - f^*\|_{L_q(\Omega)} \leq \tilde{C}\left(2h^{\tau-(1/2-1/q)_+}\|f\|_{W_2^\tau(\Omega)} + \frac{N}{2C}\|f\|^2_{W_2^\tau(\Omega)} + \epsilon\right)\,. \tag{16}$$

Applying the same arguments as in the ν-SVR case we obtain the following corollary.

Corollary 7.3 *If we choose*

$$C = \frac{N\|f\|_{W_2^\tau(\Omega)}}{2h^\tau} \text{ respectively } C = \frac{N\|f\|_{W_2^\tau(\Omega)}}{2h^{\tau-d/2}}$$

the inequality (16) turns into

$$\|f - f^*\|_{L_2(\Omega)} \leq \tilde{C}\left(3h^\tau \|f\|_{W_2^\tau(\Omega)} + \epsilon\right) \tag{17}$$

or

$$\|f - f^*\|_{L_\infty(\Omega)} \leq \tilde{C}\left(3h^{\tau-d/2} \|f\|_{W_2^\tau(\Omega)} + \epsilon\right)\,. \tag{18}$$

The role of the parameter C is similar to the one in case of the ν-SVR. Unlike the ν-SVR we are free to choose the parameter ϵ. We see that exact data implies that we should choose $\epsilon \approx 0$. The case $C \to \infty$ and $\epsilon \to 0$ leads to exact interpolation, and the well known error bounds from [10] are attained.

8 ϵ-SVR with inexact data

For inaccurate data we can proceed as above.

Lemma 8.1 *Under the assumption (12) we have*

$$\begin{aligned} \|f^*\|_{\mathcal{N}_K} &\leq \sqrt{\|f\|^2_{\mathcal{N}_K} + \frac{2C}{N}\sum_{i=1}^N |r_i|_\epsilon} \\ \|f^*|_X - y\|_{\ell_\infty(X)} &\leq \frac{N}{2C}\|f\|^2_{\mathcal{N}_K} + \sum_{i=1}^N |r_i|_\epsilon + \epsilon\ . \end{aligned}$$

These bounds shall now be plugged into the sampling inequality.

Theorem 8.2 *Under the assumption (12) we have*

$$\begin{aligned} \|f - f^*\|_{L_q(\Omega)} \leq \ & \tilde{C}\left(2h^{\tau - d(1/2-1/q)_+}\left(\|f\|_{W_2^\tau(\Omega)} + \sqrt{\|f\|^2_{W_2^\tau(\Omega)} + \frac{2C}{N}\sum_{i=1}^N |r_i|_\epsilon}\right)\right. \\ & + \left.\frac{N}{2C}\|f\|^2_{W_2^\tau(\Omega)} + \sum_{i=1}^N |r_i|_\epsilon + \epsilon\right)\ . \end{aligned}$$

Finally, we get these convergence orders, for our specific choice of the parameters.

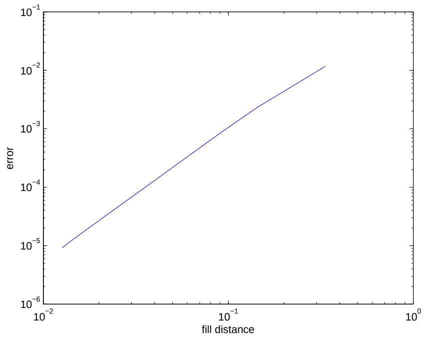
Corollary 8.3 *Again we assume that the error satisfies (12). If we then choose* $\epsilon = \delta$ *and* $C = h^{-\tau-d}/2$ *we find for quasi-uniform data*

$$\|f - f^*\|_{L_2(\Omega)} \leq \tilde{C}\left(h^\tau \|f\|_{W_2^\tau(\Omega)} + \delta\right) \quad (19)$$

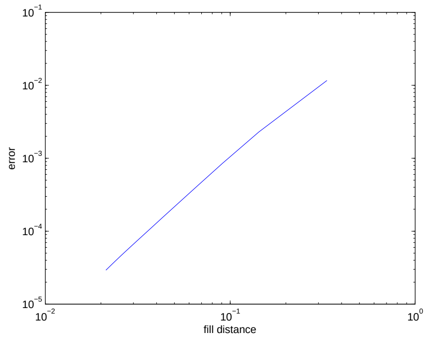
$$\|f - f^*\|_{L_\infty(\Omega)} \leq \tilde{C}\left(h^{\tau - d/2}\|f\|_{W_2^\tau(\Omega)} + \delta\right)\ . \quad (20)$$

9 Numerical results

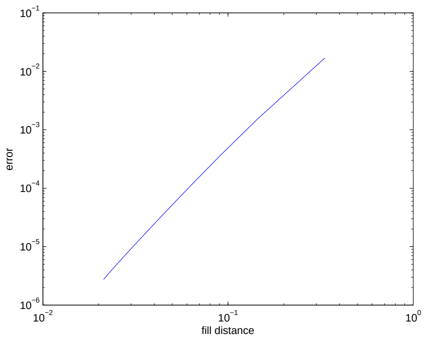
In this section we present some numerical examples to confirm our analytic results. To be able to determine convergence rates with good accuracy, we consider only univariate examples. Our first example deals with the approximation orders for the ν- and the ϵ-SVR, respectively. To test the approximation order we applied the ν-machine to a special $W_2^2([0,1])$ function, namely $f(x) = (x - 0.5)_+^{1.5+eps}$ where eps denotes the relative machine precision in the sense of MATLAB. We used the radial kernel function $K(x) = (1 - \|x\|)_+^3 (3\|x\| + 1)$ (a Wendland function) which has $W_2^2([0,1])$ as its associated native space [10]. In case of the ϵ-SVR we employed another radial kernel function (also a Wendland function) $K(x) = (1 - \|x\|)_+^5 (8\|x\|^2 + 5\|x\| + 1)$ which leads to the Hilbert space $W_3^2([0,1])$. As a test function we used $f(x) = (x - 0.5)^{2.5+eps}$. We chose the parameters according to corollaries 5.3 and 7.3, respectively. The double logarithmic plots show the expected approximation orders in the fill distance h.



(a) W_2^2 function with ν-SVR show order 2.3



(b) W_2^3 function with ϵ-SVR show order 3.3

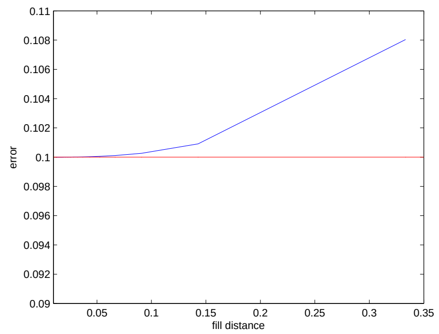


(c) W_2^3 function with ϵ-SVR show order 3.3

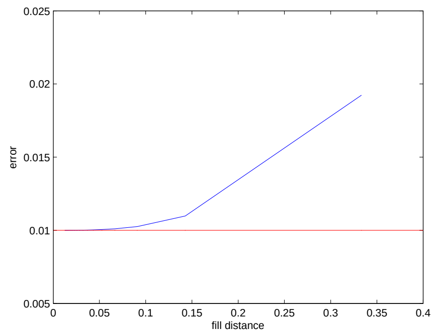
Figure 1: The double logarithmic plots confirm the analytically found approximation orders.

In Figure 9 we present some numerical experiments for erroneous data. The examples show that the approximation error converges to the error level in the limit $h \to 0$ which confirms our analytic results.

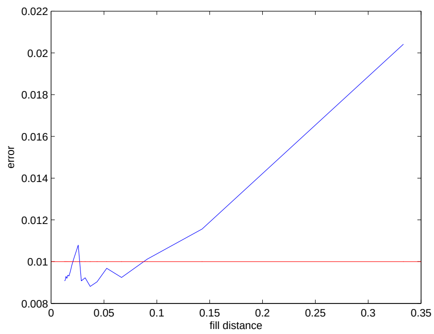
The right hand side in figure 9 shows the approximation error where the data was randomly corrupted by ± 0.01. For the left hand side plot we used data that was corrupted by a positive $+0.1$ error. In both cases we employed the ν-SVR where the free parameters were chosen according to corollary 6.3. The last figure shows the behaviour of the ν-SVR in the case of an normal distributed error with mean zero and variance $\sigma = 0.0001$. In this case the standard deviation is the error level.



(a) positive deterministic error



(b) random sign deterministic error



(c) random error

Acknowledgement

We thank Professor Robert Schaback for helpful discussions and his continued support. CR was supported by the Deutsche Forschungsgemeinschaft through the Graduiertenkolleg 1023 *Identification in Mathematical Models: Synergy of Stochastic and Numerical Methods*. BZ would like to thank the German National Academic Foundation (Studienstiftung des deutschen Volkes) for their support.

References

[1] C.-C. Chang and C.-L. Lin, *Training ν-support vector regression: Theory and algorithms*, Neural Computation, 14(8) (2002), pp. 1959–1977.

[2] T. EVGENIOU, M. PONTIL, AND T. POGGIO, *Regularization networks and support vector machines*, Advances in Computational Mathematics, 13 (2000), pp. 1–50.

[3] F. GIROSI, *An equivalence between sparse approximation and support vector machines*, Neural Computation, 10 (8) (1998), pp. 1455–1480.

[4] C. A. MICCHELLI AND M. PONTIL, *Learning the kernel function via regularization*, Journal of Machine Learning Research, 6 (2005), pp. 1099–1125.

[5] R. SCHABACK AND H. WENDLAND, *Inverse and saturation theorems for radial basis function interpolation*, Mathematics of Computation, 71 (2002), pp. 669–681.

[6] B. SCHÖLKOPF, C. BURGES, AND V.VAPNIK, *Extracting support data for a given task*, in Proceedings, First International Conference on Knowledge Discovery and Data Mining, CA:AAAI Press., Menlo Park, 1995.

[7] B. SCHÖLKOPF AND A. SMOLA, *Learning with kernels - Support Vector Machines, Regularisation, and Beyond*, MIT Press, Cambridge, Massachusetts, 2002.

[8] B. SCHÖLKOPF, R. WILIAMSON, AND P. BARTLETT, *New support vector algorithms*, Neural Computation, 12 (2000), pp. 1207–1245.

[9] V. VAPNIK, *The nature of statistical learning theory*, Springer-Verlag, New York, 1995.

[10] H. WENDLAND, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, 2005.

[11] H. WENDLAND AND C. RIEGER, *Approximate interpolation*, Numerische Mathematik, 101 (2005), pp. 643–662.