Journal of Computers Vol. 33 No. 5, October 2022, pp. 61-72
doi: 10.53106/199115992022103305006

Some Support Vector Regression Machines with Given Empirical Risks Partly

Lin-Kai Luo[1,2*], Chao-Jie Xu[1], Ling-Jun Ye[1], Hong Peng[1,2]

[1] Department of Automation, Xiamen University, Xiamen, China
luolk@xmu.edu.cn
[2] Xiamen Key Laboratory of Intelligent Analysis and Decision of Big Data, Xiamen, China

Received 19 November 2021; Revised 3 February 2022; Accepted 3 March 2022

Abstract. There are often some prior requirements about empirical risk in regression problems. To meet these requirements, this paper firstly proposes two novel support vector regression machine models in which part of empirical risks are given. One is a support vector regression machine in which partial empirical risks are given (PSVR), and the other is a model in which unilateral partial empirical risks are given (UPSVR). For the samples with given empirical risk levels, PSVR meets the requirements by some inequality constraints about empirical risk levels, while for the other samples without empirical risk requirement, PSVR uses the same strategy as the tradition support vector regression (SVR) to meet the requirement of empirical risk. UPSVR is similar to PSVR, except that the inequality constrains of empirical risks are unilateral. Secondly, the dual problems and the solving methods of PSVR and UPSVR are given. Finally, the effectiveness and superiority of PSVR and UPSVR are verified by the experiments on four artificial datasets. Both PSVR and UPSVR achieve better regression performance than the traditional models respectively. At the same time, PSVR is less sensitive to the trade-off coefficient C between empirical risk and confident risk compared with SVR. Thus, PSVR can select parameter C faster and more conveniently. PSVR and UPSVR are the extensions of the traditional models. When the set of samples with given empirical risks is empty, they degenerate into the traditional models. PSVR and UPSVR are suitable for the scene with prior requirements of empirical risk.

Keywords: support vector regression machine, partial empirical risks, confidence risk, unilateral empirical risk

1 Introduction

The support vector machine (SVM) proposed by Vapnik [1] is based on the principle of structural risk minimization (SRM). Structural risk is composed of empirical risk and confidence risk. Empirical risk refers to the performance of the model on training data, such as the misclassification rate of classification problems or the fitting error of regression problems, while the confidence risk refers to the performance of the model on test data or the generalization ability of the model. The SRM principle makes SVM obtain an excellent generalization ability. The training of SVM is realized by solving a convex quadratic programming problem (QPP) that is simple and quick. SVM has gained a wide attention and numerous applications with these advantages [2-5]. Support vector regression (SVR) [5] is the version of SVM in regression problem and maintains all the advantages of SVM in classification problem.

There are three approaches for the realization of SRM. The first one is to minimize the weighted sum of empirical risks and confidence risk. The standard SVM [1, 7] is realized with this approach. The second is to minimize the empirical risks under the conditions of a given level of confidence risk, such as [8-9]. The last is to minimize the confidence risk under the conditions of given empirical risks level, such as [10-12].

One disadvantage of the first approach is that it is very time-consuming to determine the trade-off coefficient C between empirical risk and confidence risk. The commonly used method to determine C is K-fold cross-validation, which is extremely time-consuming when the training set is large. Another disadvantage of this approach is that it cannot guarantee the required empirical risk or confidence risk level. The optimization goal of this approach is to minimize the weighted sum of empirical risks and confidence risk, which cannot ensure that the empirical risk or confidence risk meet the required level. For example, the standard SVM cannot meet the a priori requirement that "the empirical risk or confidence risk of a certain training sample cannot exceed 0.01".

One disadvantage of the second approach is that its dual problem is not a convex quadratic programming problem. In addition, it is difficult for users to give a priori requirements of confidence risk in practical application because the confidence risk is relatively abstract. These two shortcomings cause the second approach is rarely seen

* Corresponding Author

in practice.

The third approach is suitable for the scenarios with a given level of empirical risks. Compared with abstract confidence risk, empirical risk such as misclassification rate and fitting error are clear. In practical applications, users often put forward the requirements of empirical risks, for example, the fitting errors of some training samples cannot exceed 0.001. Therefore, it has more application scenarios than the second approach.

Luo et al. [10] proposed several support vector machine models under given overall empirical risk levels in 2006. In 2010, they extended it to regression problem, and proposed a support vector regression model with given empirical risk levels of all samples (ASVR) [11]. Considering that the empirical risk requirements for certain training samples are more common in practice, they further proposed a support vector classification model with partial empirical risks given (P-SVC) [12], which increases the practicability of the model.

In this paper, P-SVC is extended to regression problem, and a support vector regression machine with partial empirical risks given (PSVR) is proposed. Similar to P-SVC, PSVR covers the traditional SVR and ASVR. PSVR degenerates to the traditional SVR when no empirical risk is given, while it becomes ASVR when the empirical risks of all samples are given. PSVR can be applied to the situation where the regression curve is expected to pass through the specific samples within a given fitting error. At the same time, considering that sometimes only the upper or lower bound of empirical risk needs to be met, we further provide a support vector regression machine with unilateral partial empirical risks given (UPSVR).

The rest of the paper is organized as follows. The related work is summarized in Section 2. The details of our proposed model are provided in Section 3. Experiments are conducted in Section 4. The conclusion and future work are presented in Section 5.

2 Related Work

The regression problem considered is as follows: find a regression function f according to a given training set $T = \{(x_i, y_i) \mid x_i \in \mathbf{X} \subseteq \mathbf{R}^n, y_i \in \mathbf{R}, i = 1,2,...,l\}$, so that $y = f(x)$ can be used to infer the y value of any mode x. The traditional SVR and ASVR are commonly used to solve above regression problem.

2.1 SVR

The traditional SVR model is:

$$\begin{aligned} &\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} q_i(\xi_i + \xi_i^*) \\ &s.t. \quad y_i - w^T\psi(x_i) - b \le \varepsilon + \xi_i^*, \\ &\qquad w^T\psi(x_i) + b - y_i \le \varepsilon + \xi_i \ , \\ &\qquad 0 \le \xi_i, \xi_i^* \ , i = 1,2,\cdots,l. \end{aligned} \tag{1}$$

where w is the weight vector, b is the bias, q_i is the sample weight, ξ_i and ξ_i^* are the slack variables that represents the empirical risk of $\boldsymbol{x}_i$, ε is the bandwidth of the insensitive band, C is the trade-off coefficient between empirical risk and confidence risk, ψ denotes the nonlinear mapping from the input space to high-dimensional Hilbert space.

The dual model of (1) is:

$$\begin{aligned} &\min_{\alpha,\alpha^*} \frac{1}{2}\sum_{i=1,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i, x_j) + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i \\ &s.t. \quad \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \\ &\qquad 0 \le \alpha_i, \alpha_i^* \le Cq_i, i = 1,2,\cdots,l. \end{aligned} \tag{2}$$

where α_i and α_i^* are the multiplier variables, $k(x_i, x_j) = <\psi(x_i), \psi(x_j)>$ is the kernel function that represents the inner product of the samples in the high-dimensional Hilbert space. The matrix form of (2) is

$$\begin{aligned} &\min_{\alpha,\alpha^*} \frac{1}{2}\left[\alpha^{*T}, \alpha^T\right] Q \begin{bmatrix} \alpha^* \\ \alpha \end{bmatrix} + \left[\varepsilon e^T - y^T, \varepsilon e^T + y^T\right] \begin{bmatrix} \alpha^* \\ \alpha \end{bmatrix} \\ &s.t. \quad z^T \begin{bmatrix} \alpha^* \\ \alpha \end{bmatrix} = 0, \\ &\qquad 0 \le \alpha_i^*, \alpha_i \le Cq_i,\ i = 1,2,\cdots,l. \end{aligned} \tag{3}$$

where

$$Q = \begin{bmatrix} K & -K \\ -K & K \end{bmatrix}_{2l \times 2l}, K = [k(x_i, x_j)]_{l \times l}$$

is the inner product matrix of the samples in the high-dimensional Hilbert space, $\boldsymbol{e} = [1,1,\cdots,1]_{1\times l}^T$, $z = [1,1,\cdots,1,-1,-1,\cdots,-1]_{1\times 2l}^T$. If the kernel function is positive definite, then model (2) is a convex quadratic programming problem, which can be solved quickly with the sequential minimal optimization (SMO). If $\bar{\alpha}, \bar{\alpha}^*$ is an optimal solution of model (2), the regression function is

$$y = \bar{w}^T \psi(x) + \bar{b} = \sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) k(x, x_i) + \bar{b}$$

where

$$\bar{b} = y_i - \sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) k(x_i, x_j) + \varepsilon, \quad if\ 0 < \bar{\alpha}_j \le Cq_i, \text{ or}$$

$$\bar{b} = y_i - \sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) k(x_i, x_j) - \varepsilon, \quad if\ 0 < \bar{\alpha}_j^* \le Cq_i.$$

2.2 ASVR

ASVR minimizes the structural risk by giving the empirical risks of all samples without the trade-off coefficient *C*. It can be expressed by

$$\begin{aligned} &\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 \\ &s.t. \quad y_i - w^T \psi(x_i) - b \le \varepsilon + \xi_i^*, \\ &\qquad w^T \psi(x_i) + b - y_i \le \varepsilon + \xi_i, \\ &\qquad 0 \le \xi, \xi_i^* \le \rho_i,\ i = 1,2,\cdots,l. \end{aligned} \tag{4}$$

where $\rho_i \ge 0$ represents the given empirical risk level for sample *i*.

3 Proposed Models

In this section, two new models, PSVR and UPSVR, are proposed. Their dual problems and the solving methods are also given.

3.1 PSVR

Let $I=\{1,2,\cdots,l\}$ be the index set of all samples, $I_1 \subseteq I$ be the index set of samples with given empirical risk levels, and $I_2 = I \setminus I_1$ be the index set of samples without given empirical risk level. The prime model of PSVR is

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \ & \frac{1}{2}\|w\|^2 + C\sum_{i\in I_2}(\xi_i + \xi_i^*) \\ s.t. \ & y_i - w^T\psi(x_i) - b \le \varepsilon + \xi_i^*, \\ & w^T\psi(x_i) + b - y_i \le \varepsilon + \xi_i \ , \\ & 0 \le \xi_i, \xi_i^*, \quad i=1,2,\cdots,l, \\ & \xi_i, \xi_i^* \le \rho_i, \quad if \ i \in I_1. \end{aligned} \tag{5}$$

where $\rho_i \ge 0$ represents the given empirical risk level of $\boldsymbol{x}_i$. If a sample is given by an empirical risk level, PSVR meets this requirement by $\xi_i, \xi_i^* \le \rho_i$. Otherwise, PSVR still uses the second item of the objective function to meet the requirement of empirical risks, which is the same as the traditional SVR.

One can see that PSVR has two extreme cases. When $I_1 = \phi$, PSVR is degraded to the traditional SVR in which no sample is given empirical risk. When $I_2 = \phi$, PSVR is the same as ASVR in which all samples are given empirical risks. Therefore, PSVR is an extension of the traditional SVR and ASVR.

The following theorem 1 gives the dual problem of PSVR model and a calculation formula of regression function. The proof is omitted due to limited space.

Theorem 1: The dual model of (5) is

$$\begin{aligned} \min_{\alpha,\alpha^*,\beta,\beta^*} \ & \frac{1}{2}\sum_{i=1,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i,x_j) + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i + \sum_{i\in I_1}\rho_i(\beta_i^* + \beta_i) \\ s.t. \ & \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \\ & 0 \le \alpha_i \le \beta_i, 0 \le \alpha_i^* \le \beta_i^*, \quad if \quad i \in I_1, \\ & 0 \le \alpha_i \le C, 0 \le \alpha_i^* \le C, \quad if \quad i \in I_2. \end{aligned} \tag{6}$$

where $\alpha_i, \alpha_i^*, \beta_i, , \beta_i^*$ are the multiplier variables. If $\overline{\alpha}, \overline{\alpha}^*, \overline{\beta}, \overline{\beta}^*$ is an optimal solution of model (6), the regression function can be provided by

$$y = \overline{\boldsymbol{w}}^T\psi(\boldsymbol{x}) + \overline{b} = \sum_{i=1}^{l}(\overline{a}_i^* - \overline{a}_i)k(\boldsymbol{x},\boldsymbol{x}_i) + \overline{b}.$$

where

$$\overline{b} = y_i - \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)k(\mathbf{x}_i,\mathbf{x}_j) - \rho_j - \varepsilon, if \ j \in I_1, 0 < \overline{\alpha}_j^* \le \overline{\beta}_j^*, \text{ or}$$

$$\overline{b} = y_i - \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)k(\mathbf{x}_i,\mathbf{x}_j) + \rho_j + \varepsilon, if \ j \in I_1, 0 < \overline{\alpha}_j \le \overline{\beta}_j, \text{ or}$$

$$\overline{b} = y_i - \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)k(\mathbf{x}_i,\mathbf{x}_j) - \varepsilon, \quad if \ j \in I_2, 0 < \overline{\alpha}_j^* \le C, \text{ or}$$

$$\overline{b} = y_i - \sum_{i=1}^{l}(\overline{\alpha}_i^* - \overline{\alpha}_i)k(\mathbf{x}_i,\mathbf{x}_j) + \varepsilon, \quad if \ j \in I_2, 0 < \overline{\alpha}_j \le C.$$

The model (6) can be solved by the method of solving model (3) through some conversions.

Theorem 2: Let

$$\overline{\rho}_i = \begin{cases} \rho_i, if\ i \in I_1 \\ 0,\ if\ i \in I_2 \end{cases}, \quad \overline{C}_i = \begin{cases} +\infty, if\ i \in I_1 \\ C,\ \ if\ i \in I_2 \end{cases}. \tag{7}$$

If $\overline{\alpha}, \overline{\alpha}^*, \overline{\beta}, \overline{\beta}^*$ is an optimal solution of (6), then

$$\beta_i = \alpha_i, \beta_i^* = \alpha_i^*, if\ i \in I_1. \tag{8}$$

The proof is omitted due to limited space.

According to (7) and (8), (6) can be rewritten as

$$\begin{aligned} &\min_{\alpha,\alpha^*} \frac{1}{2}\left[\alpha^{*T}, \alpha^T\right] Q \begin{bmatrix} \alpha^* \\ \alpha \end{bmatrix} + \left[\varepsilon e^T - y^T + \overline{\rho}^T, \varepsilon e^T + y^T + \overline{\rho}^T\right] \begin{bmatrix} \alpha^* \\ \alpha \end{bmatrix} \\ &s.t. \quad z^T \begin{bmatrix} \alpha^* \\ \alpha \end{bmatrix} = 0, \\ &\qquad 0 \le \alpha_i^*, \alpha_i \le \overline{C}_i, i = 1,2,\cdots,l. \end{aligned} \tag{9}$$

The model (9) has the same structure as the model (3). Therefore, it can be solved by the method of (3).

3.2 UPSVR

Different from PSVR, UPSVR only considers the requirement of bilateral empirical risk. The requirement of bilateral empirical risk can be the upper bound or the lower bound of empirical risk. Without loss of generality, taking the upper bound as an example, the prime model of UPSVR is

$$\begin{aligned} &\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=I} \xi_i + C\sum_{i=I_2} \xi_i^* \\ &s.t. \quad y_i - w^T\psi(x_i) - b \le \varepsilon + \xi_i^*, i = 1,2,\cdots,l, \\ &\qquad w^T\psi(x_i) + b - y_i \le \varepsilon + \xi_i, \ i = 1,2,\cdots,l, \\ &\qquad 0 \le \xi_i, \xi_i^*, \quad i = 1,2,\cdots,l, \\ &\qquad \xi_i^* \le \rho_i, \quad if\ i \in I_1. \end{aligned} \tag{10}$$

There are two differences between PSVR and UPSVR: 1) the constrain $\xi_i,\ \xi_i^* \le \rho_i$ in PSVR is changed to $\xi_i^* \le \rho_i$. 2) the objective function of UPSVR adds an item of $C\sum_{i\in I_1}\xi_i$. For the sample with given upper empirical risk level, the given requirement is met by the constrain $\xi_i^* \le \rho_i$.

The following theorem 3 gives the dual problem of the UPSVR model and a calculation formula of regression function. The proof is omitted due to limited space.

Theorem 3: The dual model of (10) is

$$\begin{aligned} &\min_{\alpha,\alpha^*,\beta^*} \frac{1}{2}\sum_{i=1,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i,x_j) + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i + \sum_{i\in I_1}\rho_i\beta_i^*; \\ &s.t. \quad \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0, \quad i = 1,2,\cdots,l, \\ &\qquad 0 \le \alpha_i \le C, \quad i = 1,2,\cdots,l, \\ &\qquad 0 \le \alpha_i^* \le \beta_i^*, \quad if \quad i \in I_1, \\ &\qquad 0 \le \alpha_i^* \le C, \quad if \quad i \in I_2, \end{aligned} \tag{11}$$

If $\overline{\alpha},\overline{\alpha}^*,\overline{\beta}^*$ is an optimal solution of model (11), the regression function is

$$y=\overline{w}^T\psi(x)+\overline{b}=\sum_{i=1}^{l}(\overline{\alpha}_i^*-\overline{\alpha}_i)k(x,x_i)+\overline{b}.$$

where

$$\overline{b}=y_i-\sum_{i=1}^{l}(\overline{\alpha}_i^*-\overline{\alpha}_i)k(x_i,x_j)+\varepsilon,\qquad if\ j\in I\ ,0<\overline{\alpha}_j\le C,\text{or}$$
$$\overline{b}=y_i-\sum_{i=1}^{l}(\overline{\alpha}_i^*-\overline{\alpha}_i)k(x_i,x_j)-\rho_j-\varepsilon,if\ j\in I_1,0<\overline{\alpha}_j^*\le\overline{\beta}_j^*,\text{or}$$
$$\overline{b}=y_i-\sum_{i=1}^{l}(\overline{\alpha}_i^*-\overline{\alpha}_i)k(x_i,x_j)-\varepsilon,\qquad if\ j\in I_2,0<\overline{\alpha}_j^*\le C.$$

Model (11) can be solved by the method of solving model (3) through following conversion.

Theorem 4: If $\overline{\alpha},\overline{\alpha}^*,\overline{\beta}^*$ is an optimal solution of model (11), then

$$\beta_i^*=\alpha_i^*,if\ i\in I_1. \tag{12}$$

According to (7) and (12), (11) can be rewritten as

$$\begin{aligned}&\min_{\alpha,\alpha^*}\frac{1}{2}\left[\alpha^{*T},\alpha^T\right]Q\begin{bmatrix}\alpha^*\\ \alpha\end{bmatrix}+\left[\varepsilon e^T-y^T+\overline{\rho}^T,\varepsilon e^T+y^T\right]\begin{bmatrix}\alpha^*\\ \alpha\end{bmatrix}\\ &s.t.\quad z^T\begin{bmatrix}\alpha^*\\ \alpha\end{bmatrix}=0,\\ &\qquad 0\le\alpha_i\le C,0\le\alpha_i^*\le\overline{C}_i,i=1,2,\cdots,l.\end{aligned} \tag{13}$$

The model (13) has the same structure as the model (3) and can be solved by the method of (3).

To satisfy the requirement of bilateral empirical risk, the traditional SVR can increase the one-sided weight of empirical risk. For the sake of simplicity, we denote it as USVR. The prime model of USVR is

$$\begin{aligned}&\min_{w,b,\xi,\xi^*}\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(q_i\xi_i+q_i^*\xi_i^*)\\ &s.t.\quad y_i-w^T\psi(x_i)-b\le\varepsilon+\xi_i^*,\\ &\qquad w^T\psi(x_i)+b-y_i\le\varepsilon+\xi_i\ ,\\ &\qquad 0\le\xi_i,\xi_i^*\ ,i=1,2,\cdots,l.\end{aligned} \tag{14}$$

The dual model of (14) is

$$\begin{aligned}&\min_{\alpha,\alpha^*}\frac{1}{2}\sum_{i=1,j=1}^{l}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)k(x_i,x_j)+\varepsilon\sum_{i=1}^{l}(\alpha_i^*-\alpha_i)-\sum_{i=1}^{l}(\alpha_i^*-\alpha_i)y_i;\\ &s.t.\quad\sum_{i=1}^{l}(\alpha_i^*-\alpha_i)=0\\ &\qquad 0\le\alpha_i\le Cq_i,i=1,2,\cdots,l.\\ &\qquad 0\le\alpha_i^*\le Cq_i^*,i=1,2,\cdots,l.\end{aligned} \tag{15}$$

If $\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\alpha}}^*$ are an optimal solution of model (15), the regression function is

$$y = \bar{w}^T \psi(x) + \bar{b} = \sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) k(x, x_i) + \bar{b}$$

where

$$\bar{b} = y_i - \sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) k(x_i, x_j) + \varepsilon, \quad if \;\; 0 < \bar{\alpha}_j \le C q_i, \text{or}$$
$$\bar{b} = y_i - \sum_{i=1}^{l} (\bar{\alpha}_i^* - \bar{\alpha}_i) k(x_i, x_j) - \varepsilon, \quad if \;\; 0 < \bar{\alpha}_j^* \le C q_i^*,$$

The solving method of model (15) is the same as model (2).

4 Experiments and Results

In order to verify the effectiveness of PSVR and UPSVR, we conducted two numerical experiments. The first one compares the regression performance of PSVR and the traditional SVR, and the evaluation criterion is the absolute error between the regression value and the actual value. The second experiment compares the regression performance of UPSVR and USVR, and the evaluation criterion is the absolute error between the regression value and the actual value under the condition the regression value is greater than the actual value.

4.1 Datasets

Two experiments share the same datasets. All data in the datasets are synthetic where the true regression functions are known. Table 1 lists the basic information of the datasets. The samples are randomly generated with the true regression functions. The ratio of training set and test set is 4:1. Gaussian noise with mean zero and standard deviation 0.1 is added to each training samples.

Table 1. Datasets

Data	Function expression	Variable domain	Number of samples	Number of features
Data1	$f(x) = \dfrac{\sin x}{x}$	$x \in [-4\pi, 4\pi] \setminus \{0\}$	100	1
Data2	$f(x) = \sin\left(\dfrac{9\pi}{0.35x+1}\right)$	$x \in [0,10]$	100	1
Data3	$f(x_1, x_2) = e^{x_1 \sin(\pi x_2)}$	$x_1, x_2 \in [-1,1]$	400	2
Data4	$f(x_1, x_2) = \dfrac{(5-x_2)^2}{3*(5-x_1)^2 + (5-x_2)^2}$	$x_1, x_2 \in [0,10]$	400	2

4.2 Experimental Program

We select the radial basis function (RBF) $k(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\sigma \mid \boldsymbol{x} - \boldsymbol{x}' \mid^2)$ as kernel function in all models. We set the insensitive margin $\varepsilon = 0.01$. The kernel parameter σ and the trade-off coefficient C are selected by 5-fold

cross validation on the training set from the candidate set $\{2^{-8},2^{-7},...,2^{8}\}$. Once the optimal σ and C are determined, they are used to establish the final regression model on the full training set.

Since the prior knowledge of the synthetic dataset is known, we select the 20% samples of training set that are the closest to the actual values as the specific points (SP). For SP, we will provide the requirement of empirical risk.

For SP, the requirement of empirical risk is set by $\rho_i = 0.001$ in PSVR and UPSVR. To obtain the corresponding empirical risk in the tradition SVRs, we set $q_i = 2$ for SP, $q_i = 1$ for the other samples in SVR, and $q_i^* = 2$ for SP, $q_i = q_i^* = 1$ for the other samples in USVR.

All values of samples are scaled in [0, 1]. All methods are coded in Python and Scikit-learn [13] is used to solve the traditional SVR.

4.3 Experimental Results

Six measures are employed to evaluate the performance of models. They are root mean squared error (RMSE), mean absolute error (MAE), the ratio between the sum of squared error and the sum of squared deviation (SSE/SST), the ratio between the interpretable sum of squared deviation and the sum of squared deviation (SSR/SST), symmetric mean absolute percentage error (SMAPE) and mean absolute scaled error (MASE) [14]. In addition, the ratio between the number of samples whose predicted value is greater than the actual value and the total number of samples (Grate) is also calculated in the second experiment.

4.3.1 The Compassion Between the Traditional SVR and PSVR

As an example, Fig. 1 shows the regression results of the traditional SVR and PSVR on Data1. From Fig. 1, we see that PSVR can satisfy the given empirical risk levels of SPs and the prediction values of PSVR on the training set is closer to the real values than the traditional SVR does. Further, the prediction values of PSVR on the test set is better than the tradition SVR.

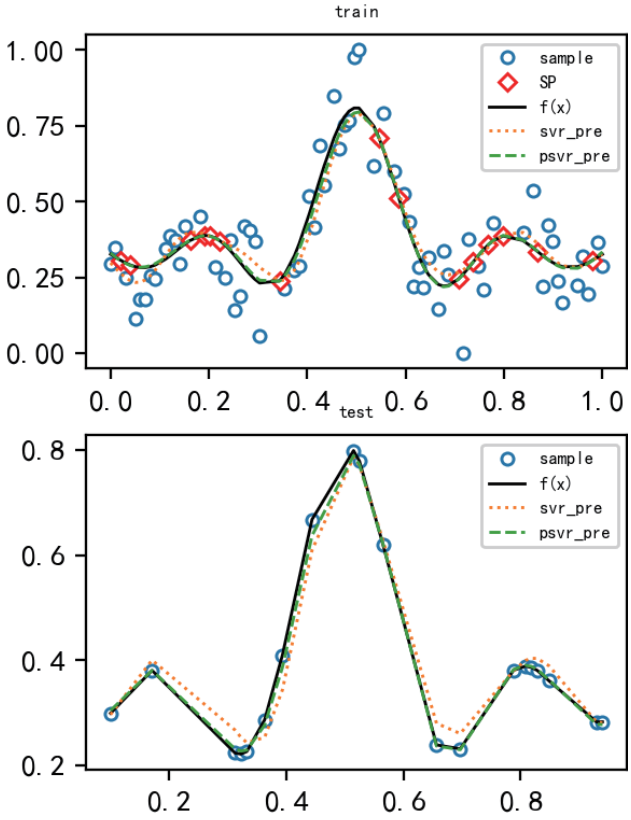

Fig. 1. The regression results of PSVR and SVR on Data1

Table 2 shows the performance comparison of traditional SVR and PSVR on the four datasets. For the six measures, PSVR obtains relative improvements of 34.06%, 39.12%, 49.71%, -7.31%, 31.58% and 38.27% respectively, compared with the traditional SVR on the four datasets. It shows that PSVR achieves a better regression performance than traditional SVR when the empirical risks of SPs are given.

Table 3 lists the results of *C*-sensitivity in PSVR and the traditional SVR on the four datasets, in which AVG and STD represent the mean and standard deviation respectively when *C* varies from $\{2^{-8},2^{-7},...,2^{8}\}$. Except SSR/SST, the AVG and STD of all measures in PSVR are significantly better than the traditional SVR, which means the sensibility for *C* in PSVR is lower than the traditional SVR.

Table 2. The regression results of PSVR and SVR on the four datasets

Data	Model	RMSE	MAE	SSE/SST	SSR/SST	SMAPE	MASE
Data1	PSVR	**0.0097**	**0.0062**	**0.0031**	0.9530	**0.0160**	**0.0710**
	SVR	0.0292	0.0226	0.0275	**0.8745**	0.0680	0.2613
Data2	PSVR	**0.0281**	**0.0169**	**0.0127**	0.8700	**0.0542**	**0.1077**
	SVR	0.0537	0.0368	0.0404	**0.7895**	0.0728	0.2185
Data3	PSVR	**0.0247**	**0.0170**	**0.0181**	**0.9595**	**0.0635**	**0.1245**
	SVR	0.0285	0.0221	0.0242	0.9991	0.0773	0.1613
Data4	PSVR	**0.0443**	**0.0345**	**0.0737**	0.9809	**0.0810**	**0.4208**
	SVR	0.0484	0.0370	0.0881	**0.8601**	0.0866	0.4511

Table 3. The results of *C*-sensitivity in PSVR and SVR on the four datasets

Data	Model	AVG/STD	*C* varies in {2^(-8),2^(-7),…,2^(8)}					
			RMSE	MAE	SSE/SST	SSR/SST	SMAPE	MASE
Data1	PSVR	AVG	**0.0155**	**0.0102**	**0.0084**	0.9598	**0.0291**	**0.1078**
		STD	**0.0088**	**0.0063**	**0.0089**	**0.036**	**0.0185**	**0.0683**
	SVR	AVG	0.0536	0.0417	0.1692	**0.7415**	0.1116	0.4709
		STD	0.0493	0.0396	0.2613	0.3589	0.0969	0.4652
Data2	PSVR	AVG	**0.0529**	**0.0358**	**0.0461**	0.8469	**0.0739**	**0.2174**
		STD	**0.0195**	**0.0132**	**0.0295**	**0.0765**	**0.0215**	**0.077**
	SVR	AVG	0.0958	0.0664	0.1687	**0.6449**	0.1461	0.4125
		STD	0.0404	0.0384	0.1727	0.2418	0.0968	0.2515
Data3	PSVR	AVG	**0.063**	**0.0453**	**0.117**	0.9767	**0.0998**	**0.3881**
		STD	**0.0086**	**0.011**	**0.0331**	**0.2262**	**0.0281**	**0.0941**
	SVR	AVG	0.0829	0.0668	0.2206	**0.8623**	0.1426	0.5723
		STD	0.0276	0.0228	0.1632	0.4294	0.0413	0.1951
Data4	PSVR	AVG	**0.0386**	**0.0283**	**0.0503**	0.887	**0.083**	**0.2116**
		STD	**0.0112**	**0.0108**	**0.0282**	**0.0687**	**0.0327**	**0.0812**
	SVR	AVG	0.0773	0.0536	0.2313	**0.6534**	0.1461	0.4017
		STD	0.038	0.0223	0.2381	0.3494	0.0509	0.1668

4.3.2 The Compassion Between USVR and UPSVR

Fig. 2 shows the regression results of USVR and UPSVR on Data1. For SP, UPSVR can meet the requirements of given unilateral empirical risks, so that the regression curve is above the SP, while USVR cannot meet this requirement.

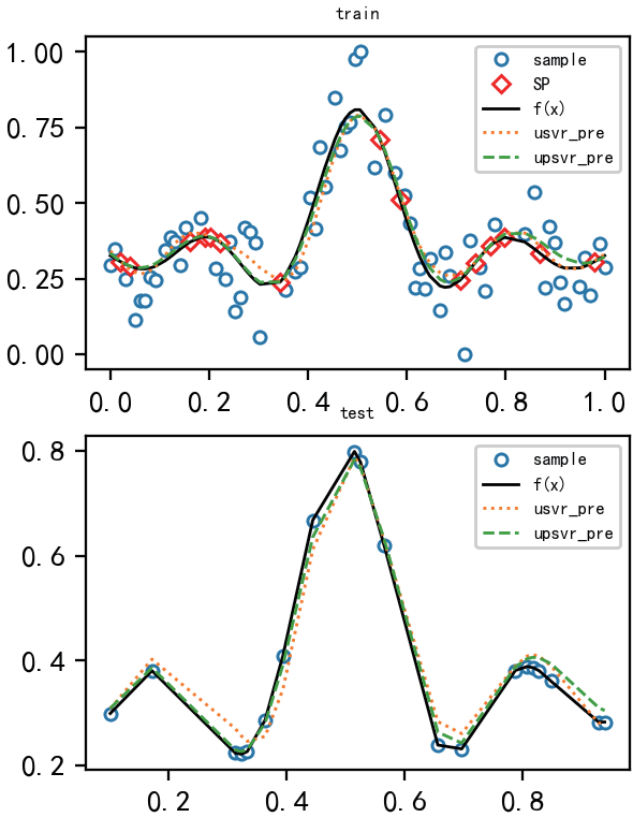

Fig. 2. The regression results of UPSVR and USVR on Data1

Table 4 lists the comparison results of regression performance between USVR and UPSVR on the four datasets. For the seven measures, UPSVR obtains the average improvements of 18.00%, 25.93%, 37.52%, 2.30%, 31.94%, 28.22% and 17.17% respectively compared with USVR on the four datasets. It shows that UPSVR has better generalization ability than USVR.

Table 5 shows the results of *C*-sensitivity in UPSVR and USVR on the four datasets. Except Grate, UPSVR is equivalent to USVR in all measures, which indicates that UPSVR and USVR are both sensitive to the parameter *C*. The reason is that UPSVR will make the regression function shift to one side of the real function so that the unilateral empirical risk can be satisfied. When *C* is small, UPSVR may cause a large deviation.

Table 4. The regression results of UPSVR and USVR on the four datasets

Data	Model	RMSE	MAE	SSE/SST	SSR/SST	SMAPE	MASE	Grate
data1	UPSVR	**0.0252**	**0.0147**	**0.0145**	0.9573	**0.0465**	**0.1432**	**0.7500**
	USVR	0.0303	0.0249	0.0296	**0.8728**	0.0748	0.2874	0.6500
data2	UPSVR	**0.0379**	**0.0285**	**0.0209**	**0.9975**	**0.1426**	**0.1701**	**0.6500**
	USVR	0.0515	0.0400	0.0384	1.0427	0.2501	0.2388	0.5500
data3	UPSVR	**0.0293**	**0.0193**	**0.0292**	**1.0095**	**0.0575**	**0.1514**	**0.7000**
	USVR	0.0344	0.0239	0.0404	1.0280	0.0803	0.1874	0.5375
data4	UPSVR	**0.0586**	**0.0473**	**0.0994**	**1.0780**	**0.1031**	**0.4053**	**0.5375**
	USVR	0.0681	0.0555	0.1339	1.2353	0.1266	0.4753	0.5125

Table 5. The results of *C*-sensitivity in UPSVR and USVR on the four datasets

Data	Model	AVG/STD	*C* varies in {2^(-8),2^(-7),…,2^(8)}						
			RMSE	MAE	SSE/SST	SSR/SST	SMAPE	MASE	Grate
data1	UPSVR	AVG	0.0855	**0.0653**	0.2988	0.9926	**0.1671**	**0.6341**	**0.8269**
		STD	0.076	0.0598	0.4156	**0.0968**	0.1392	0.5805	0.1564
	USVR	AVG	**0.0803**	0.0657	**0.2321**	**0.6439**	0.1675	0.6376	0.5615
		STD	**0.061**	**0.0523**	**0.2653**	0.3529	**0.1229**	**0.5079**	**0.0812**
data2	UPSVR	AVG	0.1083	0.0714	0.2155	0.6627	**0.1487**	0.4239	**0.55**
		STD	0.0605	0.0496	0.2669	**0.1848**	0.1069	0.2946	0.124
	USVR	AVG	**0.1036**	**0.0701**	**0.186**	**0.6252**	0.1498	**0.4161**	0.4231
		STD	**0.0504**	**0.046**	**0.1954**	0.2479	**0.1022**	**0.2734**	**0.1187**
data3	UPSVR	AVG	**0.088**	**0.0682**	**0.2498**	1.1507	**0.1618**	**0.5835**	**0.5212**
		STD	0.0299	0.0228	0.1659	**0.4261**	0.0642	0.1952	0.0655
	USVR	AVG	0.0939	0.0744	0.2789	**1.1086**	0.1729	0.6373	0.4625
		STD	**0.0289**	**0.022**	**0.1641**	0.5257	**0.0599**	**0.188**	**0.0282**
data4	UPSVR	AVG	**0.072**	**0.0539**	**0.1751**	0.8218	**0.1601**	**0.4034**	**0.5404**
		STD	**0.0207**	0.0153	**0.0986**	**0.1717**	0.0482	0.1149	0.1151
	USVR	AVG	0.0827	0.0579	0.2313	**0.7343**	0.164	0.4333	0.4692
		STD	0.024	**0.0133**	0.1359	0.2961	**0.0403**	**0.0993**	**0.0349**

5 Conclusion

In order to meet the requirements of given empirical risk level, we propose two new support vector regression models, PSVR and UPSVR, and give their dual problems and the solving methods. PSVR is an extension of the traditional SVR and ASVR. PSVR degenerates into the traditional SVR when no empirical risk is given. When the empirical risks of all training samples are given, PSVR degenerates to ASVR. UPSVR can meet the upper or lower bounds of the empirical risk when the unilateral empirical risk is given. Experiments on four artificial datasets verify the effectiveness and the superiority of the proposed models. Given the empirical risks, PSVR and UPSVR can achieve better regression performance. The sensitivity experiments of the trade-off coefficient *C* show that PSVR is less sensitive to *C* than the traditional SVR, which means that *C* can be selected faster and more conveniently in PSVR. However, UPSVR is still sensitive to *C*.

References

[1] V.N. Vapnik, An overview of statistical learning theory, IEEE transactions on neural networks 10(5)(1999) 988-999.
[2] C. Schuldt, I. Laptev, B. Caputo, Recognizing human actions: a local SVM approach, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004.
[3] L.-F. Huang, C.-L. Ye, C. Feng, H.-B. Li, Y.-M Zhang, A Detection System of Android Malware Based on SVM Algorithm, Journal of Computers 30(4)(2019) 151-158.
[4] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, Y. Li, A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease, Neurocomputing 320(2018) 195-202.
[5] R. Tu, Z. Zhu, Y. Bai, Improved Pedestrian Detection Algorithm Based on HOG and SVM, Journal of Computers 31(4) (2020) 211-221.
[6] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and computing 14(3)(2004) 199-222.
[7] D. Meyer, F. Leisch, K. Hornik, The support vector machine under test, Neurocomputing 55(1-2)(2003) 169-186.
[8] B. Scholkopf, M.K. Warmuth (Eds.), Learning Theory and Kernel Machines. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2003 (pp. 243-257).
[9] C. Zhang, Y. Tian, N. Deng, The new interpretation of support vector machines on statistical learning theory, Science in China Series A: Mathematics 53(1)(2010) 151-164.
[10]L.-K. Luo, C.-D. Lin, H. Peng, Z.-J. Wang, Some new support vector machine models under given empirical risk, in: Proc. 2006 4th IEEE International Conference on Industrial Informatics, 2006.
[11]L.-K. Luo, L.-J. Ye, H. Peng, F. Yang, Support vector regression machines with given empirical risk, Journal of Huazhong

University of Science and Technology (Natural Science Edition) 38(10)(2010) 47-51.
[12]L.-K. Luo, L.-J. Ye, Q. Zhou, H. Peng, A Support Vector Classification Model with Partial Empirical Risks Given, in: Proc. 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), 2015.
[13]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python, the Journal of machine Learning research 12(85)(2011) 2825-2830.
[14]M. Tanveer, K. Shubham, M. Aldhaifallah, S.S. Ho, An efficient regularized K-nearest neighbor based weighted twin support vector regression, Knowledge-Based Systems 94(2016) 70-87.