

# Robust Support Vector Regression Model in the Presence of Outliers and Leverage Points

Waleed Dhhan<sup>1</sup>, Habshah Midi<sup>2</sup> & Thaera Alameer<sup>3</sup>

<sup>1</sup>Nawroz University (NZU), Duhok, Iraq

<sup>2</sup>Faculty of Science and Institute for Mathematical Research, University Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

<sup>3</sup>Department of Statistics, Al-Mustansiriyah University, Baghdad, Iraq

Correspondence: WaleedDhhan, Nawroz University (NZU), Duhok, Iraq. Tel: 964-782-600-0519. E-mail: w.dhhan@gmail.com

Received: May 30, 2017

Accepted: June 11, 2017

Online Published: July 30, 2017

doi:10.5539/mas.v11n8p92

URL: <https://doi.org/10.5539/mas.v11n8p92>

## Abstract

Support vector regression is used to evaluate the linear and non-linear relationships among variables. Although it is non-parametric technique, it is still affected by outliers, because the possibility to select them as support vectors. In this article, we proposed a robust support vector regression for linear and nonlinear target functions. In order to carry out this goal, the support vector regression model with fixed parameters is used to detect and minimize the effects of abnormal points in the data set. The efficiency of the proposed method is investigated by using real and simulation examples.

**Keywords:** robust regression, support vector regression, fixed parameters, outliers, leverage points

## 1. Introduction

Support Vector Regression (SVR) involves a new class of learning algorithms which is presented by Cortes and Vapnik (1995). It is a universal technique to handle regression problems. Since then, SVR has attracted the interest of researchers due to its excellent performance of solving a variety of learning problems (Ceperic, 2014; Dhhan et al. 2015). Some additional reasons stand of behind the widely use of the SVR such as lower sensitivity to local minima, theoretical guarantees about its performance, and high flexibility to add extra dimensions to the input space, which prevents the increasing of the model complexity (Ceperic, 2014). According to Chuang et al. (2002), there is possibility to select outliers as support vectors. Further, the highest Lagrange multipliers in EQ. (4) mostly belong to the data points which are considered as outliers in the training data and it control the model (Jordaan & Smits, 2004).

In General, samples are always subject to unusual data which is called outliers. Hawkins, (1980) define an outlier as “an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism”. There are various reasons behind the presence of outliers: misplaced decimal points, measurement errors that come from lack of experience, and exceptional phenomena such as earthquakes (Rousseeuw & Leroy, 1987). Robust regression is concerned with developing estimator that is not sensitive to the outliers and to the noise distribution (Calafiore, 2000). The main idea of robust methods is to give low weights for outliers.

In this study, we have proposed Robust SVR based on fixed parameters SVR technique (FP-SVR), and we call it double support vector regression (DSVR). This article has shown that the proposed robust method achieves a higher efficiency in comparison with the standard SVR model.

This paper is organized as follows: A brief description of the support vector regression model is given in Section 2. In section 3, we present the proposed method based on FP-SVR. In Section 4 and 5, we apply the proposed method to real and simulation data sets. Finally, the discussion of the results is given in Section 6.

## 2. Support Vector Regression

In the SVR technique, the input  $x$  vector is first mapped onto a high-dimensional predictor space, which is nonlinearly related to the input space. The idea of SVR model is to employ the kernel function to transform the

nonlinear relationship in the input space to linear form in a high-dimensional feature space (Vapnik 2000). The linear model is given by

$$f(x, w) = w \cdot \Phi(x) + b \tag{1}$$

where  $\Phi(x)$  denotes a non-linear function,  $w$  and  $b$  are the slope and the bias term respectively.

Support vector regression technique aims to estimate the parameter values  $w$  and  $b$  that optimize predicted risk by minimizing the following loss function

$$L_\epsilon(y_i) = \begin{cases} 0 & \text{if } |y_i - f(x, w)| \leq \epsilon \\ |y_i - f(x, w)| - \epsilon & \text{otherwise} \end{cases} \tag{2}$$

In order to find a flat function  $f(x, w)$ , minimizing the Euclidean norm  $\|w\|^2$  is needed (Smola and Schölkopf 2004). This can be done by introducing some positive slack variables  $(\xi_i, \xi_i^*)$ , to measure deviations of the training vectors outside the  $\epsilon$ -tube. Thus, a convex optimization problem can be formulated as:

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{subject to} && \begin{cases} y_i - f(x, w) - b \leq \epsilon + \xi_i \\ f(x, w) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{cases} \end{aligned} \tag{3}$$

The coefficient  $C$  is defined as the trade-off between complexity of the model and the number of deviations that larger than the coefficient  $\epsilon$  is tolerated (Smola and Schölkopf 2004). The coefficient  $\epsilon$  controls the width of the  $\epsilon$ -zone, which is used to fit the training data (Vapnik 2000). Thus, the estimated SVR function can be written as follows:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{4}$$

where  $\alpha_i$  and  $\alpha_i^*$  define as Lagrange multipliers, and  $k(x_i, x)$  is the function of kernel.

### 3. The Proposed Method Based on Fixed Parameter SVR

The main contribution of this study is to detect outliers and leverage points first, and then to minimize their effects. To achieve the first step, we have used the FP-SVR (Dhhan et al. 2015) to detect outliers and leverage points. The FP-SVR technique is insensitive to outliers because it succeeded to control the free parameters ( $\epsilon, C$ , and  $h$ ). According to FP-SVR model (5), the optimal parameters to detect outliers are:  $\epsilon = 0, C = 100000$ , and  $h = 1$ .

$$z = \hat{y}_{SVR} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b, \alpha_i, \alpha_i^* \in [0, C] \tag{5}$$

Based on FP-SVR, any point with a  $z$  value larger than the  $CP$  (6) is considered to be an outlier

$$\begin{aligned} CP &= 2 \text{Median}|z| + 2 \text{sd}(\text{Median}) \\ \text{sd}(\text{Median}) &= \sqrt{\pi \text{var}(z) / 2n} \end{aligned} \tag{6}$$

This method can be employed to prevent the effects of abnormal data the estimation. This can be done by using the following weight function.

$$\tau_i = \min[1, CP/z] \tag{7}$$

The Eq. (4) can be rewritten based on  $\tau$  function to achieve the final robust SVR function

$$f(x) = \sum_{i=1}^n \tau_i (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{8}$$

### 4. Artificial and Real Case Studies

In this part, we apply our proposed method (DSVR) to Belgian phone data, Hawkins-Bradu-Kass data and simulation example (Rousseeuw and Leroy, 1987), and compare the results with standard SVR. The mean square error (MSE) of the residuals is used to evaluate the proposed technique.

#### 4.1 Belgian Phone Data

In this example, the proposed technique is tested using real data which has vertical outliers. In the Belgian Statistical Survey, a data set was found containing the total number of international phone calls made between the years 1950 and 1973, which contains heavily contaminated data (Leroy and Rousseeuw, 1987).

Table 1. The MSE of SVR and the DSVR methods for Belgian phone data

Parameters	SVR			DSVR			
	C=1	C=50	C=100	C=1	C=50	C=100	
$\epsilon = 0$	$h=0.5$	0.068	0.247	0.249	<b>0.059</b>	<b>0.084</b>	<b>0.087</b>
	$h=1$	0.186	0.139	0.140	<b>0.079</b>	<b>0.094</b>	<b>0.092</b>
	$h=5$	0.129	0.150	0.150	<b>0.048</b>	<b>0.031</b>	<b>0.032</b>
$\epsilon = 0.1$	$h=0.5$	0.062	0.229	0.214	<b>0.054</b>	<b>0.060</b>	<b>0.059</b>
	$h=1$	0.156	0.188	0.184	<b>0.058</b>	<b>0.056</b>	<b>0.052</b>
	$h=5$	0.138	0.238	0.222	<b>0.038</b>	<b>0.031</b>	<b>0.030</b>
$\epsilon = 0.2$	$h=0.5$	0.092	0.217	0.257	<b>0.064</b>	<b>0.063</b>	<b>0.064</b>
	$h=1$	0.147	0.178	0.237	<b>0.059</b>	<b>0.062</b>	<b>0.062</b>
	$h=5$	0.127	0.317	0.274	<b>0.060</b>	<b>0.049</b>	<b>0.044</b>

As shown in Table and Figure 1, the proposed approach is effective to minimize the effects of outliers much better than the standard SVR model.

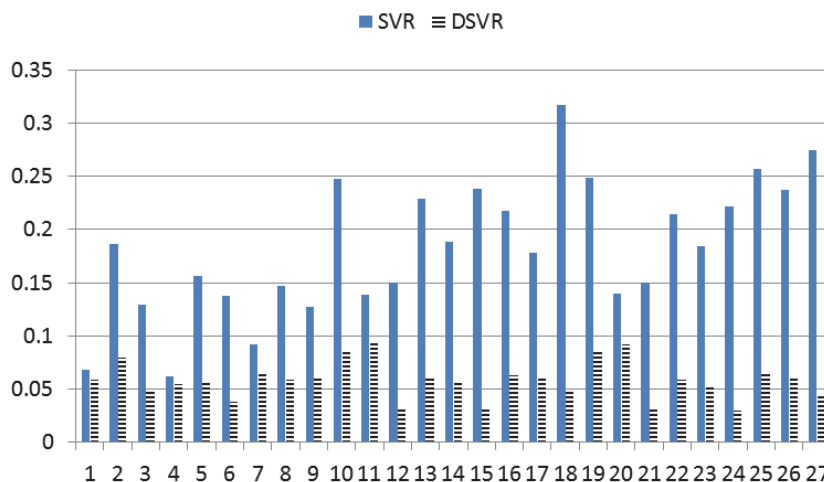


Figure 1. The MSE of SVR and the DSVR methods for Belgian phone data

4.2 Hawkins-Bradru-Kass Data

To illustrate the superiority of the proposed method, we have applied it to Hawkins-Bradru-Kass data. The data set consists of three predictors in addition to the dependent variables, with 75 observations. The first 10 observations are classified as bad leverage points (have bad effect on the regression estimator), and next four observations are considered as good leverage points (have no or small effect on the regression estimator).

Table 2. The MSE of SVR and the DSVR methods for HBK data

Parameters	SVR			DSVR			
	C=1	C=50	C=100	C=1	C=50	C=100	
$\epsilon = 0$	$h=0.5$	0.222	0.247	0.950	<b>0.165</b>	<b>0.019</b>	<b>0.004</b>
	$h=1$	0.123	0.480	0.481	<b>0.092</b>	<b>0.0001</b>	<b>0.0001</b>
	$h=5$	0.024	0.320	0.320	<b>0.021</b>	<b>0.0001</b>	<b>0.0001</b>
$\epsilon = 0.1$	$h=0.5$	0.208	0.349	0.497	<b>0.159</b>	<b>0.023</b>	<b>0.014</b>
	$h=1$	0.115	0.385	0.385	<b>0.095</b>	<b>0.010</b>	<b>0.010</b>
	$h=5$	0.032	0.356	0.356	<b>0.029</b>	<b>0.009</b>	<b>0.009</b>
$\epsilon = 0.2$	$h=0.5$	0.208	0.303	0.293	<b>0.165</b>	<b>0.048</b>	<b>0.043</b>
	$h=1$	0.118	0.335	0.335	<b>0.104</b>	<b>0.040</b>	<b>0.040</b>
	$h=5$	0.056	0.413	0.413	<b>0.049</b>	<b>0.034</b>	<b>0.034</b>

The comparison results are recorded in Table and Figure 2. These results clearly show the superiority of our proposed method DSVR over the standard SVR method in terms of achieve lower values of the MSE. Based on these results we can conclude that the use of the proposed method is recommended for this data set.

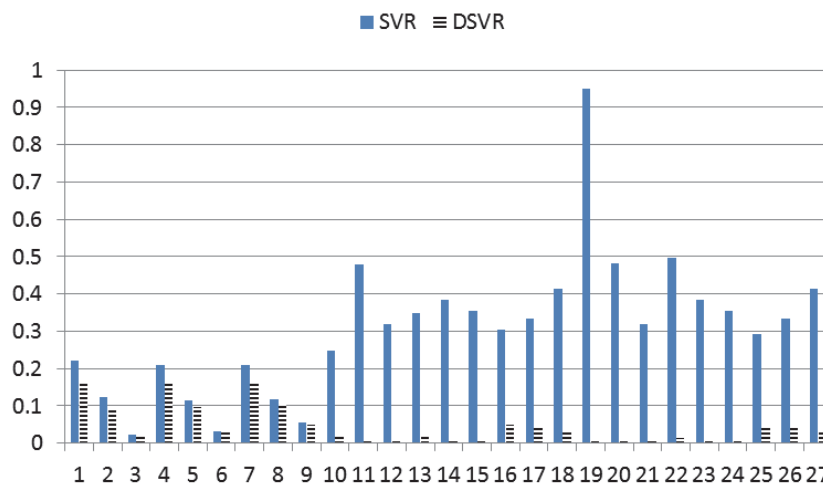


Figure 2. The MSE of SVR and the DSVR methods for HBK data

### 5. Simulation Study

In order to generalize the DSVR technique, we simulate a nonlinear regression model with two predictors.

$$y = \frac{\sin \sqrt{x_1^2 + x_2^2}}{\sqrt{x_1^2 + x_2^2}} + r_i \text{ with } x \in [-10, 10] \tag{9}$$

The values of  $x_1$  and  $x_2$  are sampled from uniform distribution, while the residuals  $r_i, i = 1, 2, \dots, n$  are simulated based on standard normal distribution. We contaminate the data by different outliers' percentages (10%, 15% and 20%). This contamination is implemented by replacing some points of  $x_1$  and  $y$  by extreme value equal to 100.

The results of the comparison of the proposed DSVR and the standard SVR are shown in Table 3. This explains the estimates of DSVR and SVR for different samples and contamination levels. These results are explained graphically in Figures 3, 4 and 5. These results reveal that the DSVR has smaller MSE than the standard SVR.

Table 3. The MSE of the SVR and the DSVR methods for different sample sizes and percentages of contamination

Cont.	n	Parameters	SVR				DSVR			
			$\epsilon=0$		$\epsilon=0.2$		$\epsilon=0$		$\epsilon=0.2$	
			C=1	C=100	C=1	C=100	C=1	C=100	C=1	C=100
10%	50	$h=1$	1.18	1.43	1.16	1.34	<b>1.17</b>	<b>1.32</b>	<b>1.14</b>	<b>1.24</b>
		$h=5$	1.11	1.61	1.10	1.43	<b>1.11</b>	<b>1.40</b>	<b>1.10</b>	<b>1.26</b>
	100	$h=1$	1.16	1.45	1.13	1.33	<b>1.14</b>	<b>1.26</b>	<b>1.11</b>	<b>1.17</b>
		$h=5$	1.08	1.14	1.06	1.11	<b>1.07</b>	<b>1.10</b>	<b>1.07</b>	<b>1.07</b>
	250	$h=1$	1.22	2.69	1.19	2.19	<b>1.19</b>	<b>1.84</b>	<b>1.15</b>	<b>1.50</b>
		$h=5$	1.12	1.34	1.10	1.26	<b>1.10</b>	<b>1.19</b>	<b>1.08</b>	<b>1.13</b>
15%	50	$h=1$	1.17	1.43	1.15	1.34	<b>1.16</b>	<b>1.32</b>	<b>1.14</b>	<b>1.24</b>
		$h=5$	1.11	1.61	1.10	1.43	<b>1.11</b>	<b>1.40</b>	<b>1.10</b>	<b>1.26</b>
	100	$h=1$	1.16	1.42	1.13	1.31	<b>1.14</b>	<b>1.24</b>	<b>1.11</b>	<b>1.16</b>
		$h=5$	1.07	1.14	1.06	1.11	<b>1.07</b>	<b>1.10</b>	<b>1.06</b>	<b>1.07</b>
	250	$h=1$	1.22	2.58	1.18	2.11	<b>1.18</b>	<b>1.77</b>	<b>1.14</b>	<b>1.46</b>

	$h=5$	1.11	1.34	1.09	1.25	<b>1.09</b>	<b>1.19</b>	<b>1.08</b>	<b>1.13</b>	
20%	50	$h=1$	1.17	1.41	1.14	1.33	<b>1.16</b>	<b>1.31</b>	<b>1.13</b>	<b>1.24</b>
		$h=5$	1.11	1.60	1.10	1.42	<b>1.11</b>	<b>1.40</b>	<b>1.10</b>	<b>1.26</b>
	100	$h=1$	1.17	1.37	1.14	1.27	<b>1.14</b>	<b>1.20</b>	<b>1.11</b>	<b>1.14</b>
		$h=5$	1.08	1.12	1.07	1.09	<b>1.07</b>	<b>1.08</b>	<b>1.06</b>	<b>1.07</b>
	250	$h=1$	1.21	2.53	1.18	2.07	<b>1.18</b>	<b>1.75</b>	<b>1.14</b>	<b>1.45</b>
		$h=5$	1.10	1.32	1.09	1.24	<b>1.09</b>	<b>1.18</b>	<b>1.07</b>	<b>1.12</b>

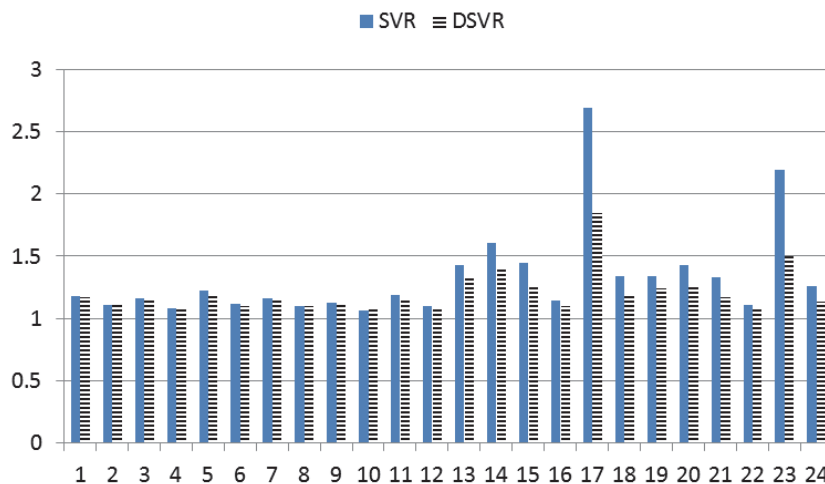


Figure 3. The MSE of the SVR and the DSVR methods for different sample sizes and 10% percentage of contamination

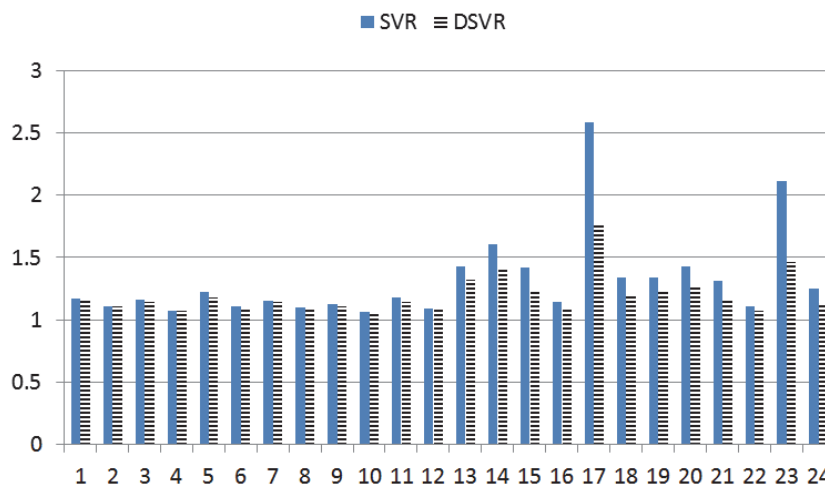


Figure 4. The MSE of the SVR and the DSVR methods for different sample sizes and 15% percentage of contamination

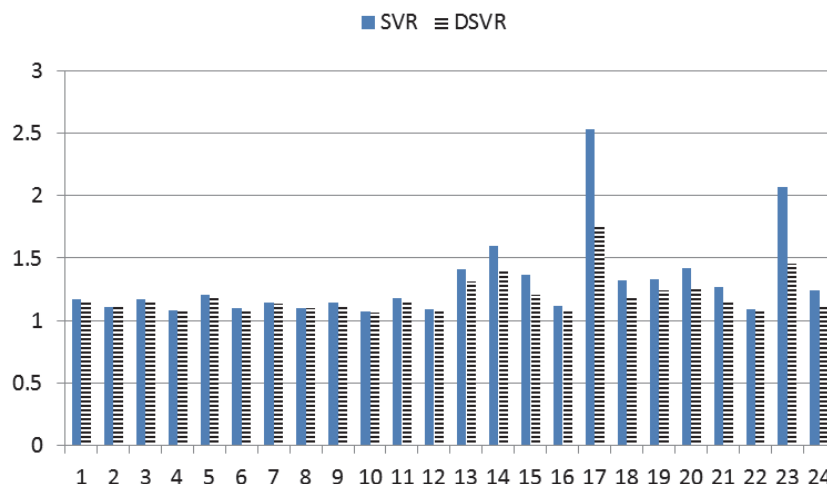


Figure 5. The MSE of the SVR and the DSVR methods for different sample sizes and 20% percentage of contamination

## 6. Conclusion

In this article, we have proposed the robust support vector regression technique for nonlinear function. Although, SVR is nonparametric method but it is still affected by outliers. In order to compare our proposed method, DSVR, with standard SVR method, real and simulation examples have been used. The results are calculated based on some of SVR parameters and sample sizes. Finally, the comparison results are demonstrated, the superiority of proposed method, DSVR over standard SVR method for nonlinear target functions

## References

- Calafiore, G. C. (2000). Outliers robustness in multivariate orthogonal regression. *Systems, Man and Cybernetics, Part A: Systems and Humans*, *30*(6), 674-679.
- Ceperic, V., Gielen, G., & Baric, A. (2014). Sparse  $\epsilon$ -tube support vector regression by active learning. *Soft Computing*, *18*(6), 1113-1126.
- Chuang, C., Su, S., Jeng, J., & Hsiao, C. (2002). Robust support vector regression networks for function approximation with outliers, *Neural Networks*. *IEEE Transactions*, *13*(6), 1322-1330.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297.
- Dhhan, W., Rana, S., & Midi, H. (2015). Non-sparse  $\epsilon$ -insensitive support vector regression for outlier detection. *Journal of Applied Statistics*, *42*(8), 1723-1739.
- Hawkins, D. M. (1980). *Identification of Outliers*, Springer, New York.
- Jordaan, M. E., & Smits, F. G. (2004). Robust outlier detection using SVM regression. *Neural networks, IEEE international joint conference proceedings, Budapest*. *Hungary*, 2017-2022.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. J. Wiley & Sons, New York.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199-222.
- Vapnik, V. (2000). *The nature of statistical learning theory*, Springer, New York.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).