Pakistan Journal of Statistics and Operation Research

The Support Vector Regression Model: A New Improvement for some Data Reduction Methods with Application

Moustafa Salem[1,*] and Mohamed G. Khalil [2]

* Corresponding Author

[1]Department of Applied Statistics, Damanhour University, Egypt. Email: moustafasalemstat@com.dmu.edu.eg
[2] Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt.
Email: hndaoy@gmail.com

Abstract
Support Vector Regression (SVR) formulates is an optimization problem to learn a regression function that maps from input predictor variables to output observed response values. The SVR is useful because it balances model complexity and prediction error, and it has good performance for handling high-dimensional data. In this paper, we use the SVR model to improve the principal component analysis and the factor analysis methods. Simulation experiments are performed to assessment the new method. Some useful applications to real data sets are presented for comparing the competitive SVR models. It is noted that with increasing sample size, the ε-SVR type under the principal component analysis is the best model. However, under the small sample sizes the v −SVR type under the factor analysis provided adequate results.

Key Words: Support Vector Regression; ε-Support Vector Regression; Factor Analysis; Kernel Functions; Principal Component Analysis; v -Support Vector Regression.

1.Introduction

The recent trends in collecting huge and diverse data sets, such as documents, videos and digital images, financial time series, and gene expressions and DNA copy numbers, have posed a great challenge that is brought by the high dimensionality and aggravated by the presence of irrelevant dimensions in tasks such as predictive modeling (Glaser et al. (2019)).

The principal component analysis (PCA) helps in building a predictive model that is simple as it contains the smallest number of variables and efficient that accounts for as much of the information "explained variation" as possible, see Mechelli and Vieira (2019), Rosipal et al. (2001), Shi et al. (2008), Astuti (2018) and Drucker et al. (1997) for more details.

The PCA can be widely applied in all forms of analysis from neuroscience to computer graphics and in a variety of real-world applications including image segmentation (Shokri et al. (2015)), climate research (Jolliffe (1986)), genome-wide expression studies (Jiang et al. (2011)), and deep learning (Yu et al. (2014)) due to its superior properties, such as linear un-correlation, low-dimensionality and visualization in multivariate data, over other linear dimension reduction (LDR) methods (Chao et al. (2009) and Drucker et al. (1997)).

Support Vector Machine (SVM) is one of the most robust prediction methods, based on the statistical learning framework or VC theory proposed by Vapnik and Chervonenkis (1974) and Vapnik (1982, 1995). SVM seeks to maximize the predictive accuracy from computation of a confidence interval for the importance of a variable in order to describe the relationship between inputs and outputs (see Chao et al. (2009)). SVM is a supervised learning model, with associated learning algorithm that analyze data used for classification, known as SV classifier, and regression (function approximation), known as SVR (Chowdhury et al. (2017)).

During past few decades, an extension to the SVM classification algorithm has been received a considerable attention, see Chowdhury et al. (2018) and Rosipal et al. (2001)). The SVR has additional advantages compared to other regression methods (see Chowdhury et al. (2017)).

PCA is a widely applied feature extraction method in the framework of SVR. In the literature, Lee and Verleysen (2009) proposed an integration of PCA and SVR models, which can be also noted as PCA-SVR, to enhance the performance of prediction (forecasting) model for financial time series. PCA-SVR produced less mean average precision MAP (%), mean absolute error (MAE), root mean square error (RMSE) and mean square error (MSE) than single SVR, Chowdhury et al. (2017) proposed PCA-SVM stock selection model which achieves the entire accuracy of 75.44% in training set and of 61.79% in testing set.

Two types of procedures have been adopted within the practical aspect. The first procedure is applying the PCA within ε-SVR type. The other is also applying the PCA but within v-SVR type.

The rest of the paper is organized as follows: Section 2 presents the methodology that used in this paper PCA, factor analysis, SVR. Section 3 discussed the results and evaluation. Finally, conclusion Section 4.

2. Methodology

2.1 The Principal Component Analysis

The PCA is fundamentally a dimensionality reduction algorithm, but it can also be useful as a tool for visualization, for noise filtering, for feature extraction and engineering, and much more.

2.2 The Factor Analysis

The factor analysis (FA) is a powerful data reduction technique that enables researchers to investigate concepts that cannot easily be measured directly. When applied to a large amount of data, it compresses the set into a smaller set that is far more manageable, and easier to understand.

2.3 The Support Vector Regression model

The SVR extends the basic principles of SVM for classification of Jolliffe and Cadima (2016) by measuring the error of approximation instead of the margin used in classification. SVR estimates a continuous-valued function that encodes the fundamental interrelation between a given input and its corresponding output in the training data. This function then can be used to predict outputs for given inputs that were not included in the training set. This is similar to a neural network. However, a neural network's solution is based on empirical risk minimization. In contrast, SVR introduces structural risk minimization into the regression and thereby achieves a global optimization, while a neural network achieves only a local minimum. Brief descriptions of two types of SVR which have been considered in the paper are given.

2.3.1 The ε-SVR Model

The ε-SVR maps the input vectors $x_i \in R^m$into a high dimensional feature space. Given a training set

$$\left(x_i, y_i\right),\ i = 1,2, \dots, n,$$

where $x_i \in R^m$ is the m-dimensional input vector and $y_i \in R$is the response variable. SVR generates the linear regression function in the form of generic cost estimation model that can be written as

$$y = f(x) = \mathcal{W},\ x + b = \mathcal{W}^T x + b \tag{1}$$

where $\mathcal{W}$ is the weight vector corresponding to x and b is the bias. The Vapnik's linear ε -Insensitivity loss (error) function is also given as

$$L(y, f(x)) = \begin{cases} 0 & \text{if} \quad |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & \text{otherwise} \end{cases} \tag{2}$$

Based on the above, the linear regression $f(x)$ is estimated by simultaneously minimizing $||\mathcal{W}||^2$ and the sum of the linear ε -Insensitivity losses as shown in Equation (7). The constant $c$controls a trade-off between an approximation error and the weight vector norm $\mathcal{W}$ is a design parameter chosen by the user.

$$R = \frac{1}{2}||\mathcal{W}||^2 + c(\sum_{i=1}^{n}|y - f(x)|_\varepsilon) \tag{3}$$

Minimizing the risk R is equivalent to minimizing the following risk under the following constraints mentioned in Equations

$$\text{Minimize } R = \frac{1}{2}||\mathcal{W}||^2 + c\sum_{i=1}^{n}(\xi_i + \xi_i^*), \tag{4}$$

subject to

$$\begin{cases} (\mathcal{W}^T x_i + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (\mathcal{W}^T x_i + b) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1,2,\dots,m \end{cases}, \tag{5}$$

Here, ξ_i and $\xi_i^*$are slack variables, one for exceeding the target value by more than ε and other for being more than ε below the target. As used in SVM, the above constrained optimization problem is solved using Lagrangian theory and the Karush-Kuhn-Tucker (KKT) conditions for the optimum of a constrained function to obtain the desired weight vector of the regression function [17].

In Equation (4), the generalization performance of such linear function, $f(x)$, is fairly limited and unable to reflect thetrue regression procedure. In order to overcome suchweakness, a standard mathematical solution is the introductionof kernel function, $\varphi(X)$, which is a nonlinearmapping function from the input space to a higher dimensionalfeature space. We can reachinfinite dimensions for a more expressive$f$ by using $\varphi(X)$. The most popular kernel functions used in this study are shown in Table 1 below.

Table 1: Admissible kernel functions

Name	Definition	Parameter
Linear	$k(x_1, x_2) = x_1.x_2$	-
Polynomial	$k(x_i, x_j) = (x_i.x_j + 1)^d$	d
Radial basis function	$k(x_i, x_j) = exp\,(-\gamma\|x_i - x_j\|^2)$	γ
Sigmoid	$k(x, y) = tanh\,(\alpha x^T y + c)$	c

2.3.2 v-SVR Model

The v-SVR is one of the most popular modifications proposed by Scho¨lkopf, Bartlett, Smola, and Williamson (1999). The benefit of$v$-SVR is that it provides a way to automatically minimize ε.

In the ε-SVR, selection of a proper ε value is essential for an accurate regression approximation. However, it is difficult to specify ε beforehand, other than an empirical choice. In the v-SVR a new parameter of a prior $v \in (0,1)$ is introduced to automatically adjust a flexible tube by controlling the number of support vector and tolerated training errors. Then, the parameter ε becomes a variable in the optimization process and is controlled by the new parameter$v$. Inv-SVR, the optimization problem can be written, given a function $\varphi(x)$to the kernel space for a nonlinear case, as follows

$$min_{\mathcal{W}}\frac{1}{2}||\mathcal{W}||^2 + C\left(v\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi + \xi^*)\right),$$

subject to

$$\begin{cases} y_i - \mathcal{W}^T\varphi(x_i) - b \leq \varepsilon + \xi \\ \mathcal{W}^T\varphi(x_i) + b - y_i \leq \varepsilon + \xi^* \\ \xi, \xi^*, \varepsilon \geq 0 \end{cases} \tag{6}$$

Here, the newly introduced constant variable $v \in (0,1)$is used as atrade-off against model complexity and slack variables. Forming aLagrangian formulation from ($\varphi(.) = R^d \to F$) by introducing positive multipliers $\alpha, \alpha^*, \eta, \eta^*$and b gives

$$L(\mathcal{W}, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*, \beta) = \frac{1}{2}||\mathcal{W}||^2 + Cv\varepsilon + \frac{C}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) + \sum_{i=1}^{l}\alpha_i^*(y_i - \mathcal{W}^T x_i - b - \varepsilon - \xi_i) \tag{7}$$

$$+\sum_{i=1}^{l}\alpha_i(\mathcal{W}^T x_i + b - y_i - \varepsilon - \xi_i^*) - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*)\beta\varepsilon$$

Following the KKT conditions that partial derivatives with respect to the variables $\mathcal{W}, b, x, x^*, and\varepsilon$ are equal to be zero and the products of theLagrange multipliers and the constraint are equal to zero, we have thefollowing dual optimization problem of v-SVR

$$max_{\alpha,\alpha^*}\sum_l^{l_{sv}} y_i(\alpha_i - \alpha_i^*) - \frac{1}{2}\sum_i^{l_{sv}}\sum_j^{l_{sv}}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j),$$

where $k(x_i, x_j) = \varphi(x_i)\varphi(x_j)$*subject to*

$$\sum_i^{l_{sv}}(\alpha_i - \alpha_i^*) = 0,\ \alpha_i, \alpha_i^* \in \left[0, \frac{C}{l}\right], \sum_i^{l_{sv}}(\alpha_i - \alpha_i^*) \le Cv \tag{8}$$

Then, the regression estimate takes the form

$$f(x) = \sum_i^{l_{sv}}(\alpha_i - \alpha_i^*)\, k(x_i, x) + b \tag{9}$$

Compared to the optimization problem in ε-SVR (Chao et al. (2009)), we can see that the parameter ε vanishes but instead there is the new parameter v in v-SVR (Chao et al. (2019)). Scho¨lkopf et al. had proved that$v \in (0, 1)$ is an upper bound on the fraction of errors (i.e., data points outside of the tube divided by the total number of data points l) and a lower bound on the fraction of support vectors (i.e., the numbers of support vectors divided by the total number of data points l).

3. Experimental procedures

In this part, real data and simulations were used conducted for the purpose of comparison between ε-SVR and v-SVR models after and before applying PCA and FA using four different kernel functions to detect the PCA and FA effects on data reduction, In this study, all trained models designed are evaluated using measured data based on root mean square error (RMSE), coefficient of determination (R^2), Relative Efficiency (RE) and Reduction No. of SVR.

RMSE is a commonly used measure of the difference between predicted values of model and the actual values from the system that is being modeled, RE is (RMSE / the biggest RMSE) *100%. The sample sizes are n=50, 100, and 150. The simulation results were based on 10000 replications. All computation is using the R program. The table below represents the results of this article.

Table 2 gives the simulation results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 30. Table 3 gives the simulation results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 100. Table 4 gives the simulation results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 150.

Table 5 gives the real data results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 30. Table 6 gives the real data results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 100. Table 7 gives the real data results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 150.

Table 2: The simulation results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 30.

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	18	1.9594	0.6247	0.9998	9	2.7173	0.809	0.593924
	Polynomial	16	3.1365	1	0.8682	16	3.3587	1	0.999823
	Radial	18	1.5879	0.5063	0.9781	13	2.5251	0.7518	0.569619

	Sigmoid	24	2.297	0.7324	0.488	24	2.8239	0.8408	0.868214
FA	Linear	16	0.1659	0.192	0.1458	12	0.1444	0.2441	0.1613
	Polynomial	16	0.6737	0.7797	0.0088	18	0.5632	0.9523	0.004
	Radial	14	0.0277	0.032	0.2471	13	0.0053	0.0089	0.3309
	Sigmoid	24	0.864	1	0.0696	22	0.5914	1	0.0154

Table 3: The simulation results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 100

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	73	1.9594	0.6247	0.9998	69	2.8525	0.8573	1
	Polynomial	73	3.1365	1	0.8682	45	3.3273	1	0.6897
	Radial	73	1.5879	0.5063	0.9781	56	2.648	0.7958	0.9063
	Sigmoid	76	2.297	0.7324	0.488	76	2.9052	0.8731	0.5265
FA	Linear	73	1.151632	0.50289	0.276829	43	1.137073	0.34177	0.279444
	Polynomial	73	1.587911	0.69341	0.978067	49	2.852466	0.857293	1
	Radial	76	2.138123	0.93367	0.019628	47	1.226413	0.36862	0.184143
	Sigmoid	76	2.297006	1	0.487993	45	3.327294	1	0.6897229

Table 4: The simulation results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 150

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
	Linear	98	0.0923984	0.09209	0.1485413	66	5.1164	0.2269	0.12259
	Polynomial	111	0.99995	1	54.41528	75	1.7035	0.0755	0.54595
	Radial	100	0.69855	0.69924	12.45879	69	10.964	0.4864	0.71345
	Sigmoid	120	0.85871	0.85956	2.80651	100	22.541	1	0.03979
FA	Linear	99	3.0601	0.6644	0.655535	56	5.7075	0.2329	0.08775
	Polynomial	112	4.6058	1	0.42215	89	1.841	0.07515	0.605545
	Radial	102	2.3699	0.51454	0.9101	68	4.763	0.1944	0.9101
	Sigmoid	119	3.4027	0.73878	0.50777	99	24.496	1	0.47863

Table 5: Real data results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 30

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	11	3.101	0.0251	0.7955	12	1.7169	0.063	0.7990
	Polynomial	12	81.33	0.6596	0.586	13	30.7709	1.1293	0.5851
	Radial	20	3.108	0.0252	0.774	13	5.677	0.2083	0.7789

	Sigmoid	24	123.3	1	0.2903	11	27.2462	1	0.3354
FA	Linear	10	0.6422	0.537	0.3725	13	0.6567	0.46363	0.3818
	Polynomial	24	1.1955	1	0.1586	15	1.4165	1	0.1603
	Radial	11	0.4221	0.353	0.4325	16	0.6824	0.4817	0.4208
	Sigmoid	23	1.1633	0.973	0.0430	21	1.2746	0.8998	0.0608

Table 6: Real data results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 100

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	70	8.192	0.16322	0.70107	43	3.04894	0.02965	0.72162
	Polynomial	68	29.82	0.594142	0.80555	43	9.2942	0.09039	0.70836
	Radial	70	19.07	0.379956	0.74132	48	2.67359	0.02600	0.7727
	Sigmoid	75	50.19	1	2.43×10^{-5}	41	102.82	1	0.0003
FA	Linear	60	0.4690	0.06693	0.54805	43	0.78843	0.31488	0.550267
	Polynomial	69	1.4202	0.20266	0.10935	45	1.32879	0.53068	0.281505
	Radial	59	0.3608	0.05149	0.62744	52	0.82480	0.32940	0.639909
	Sigmoid	76	7.0078	1	0.00838	42	2.50392	1	0.0381035

Table 7: the results for the RMSE to all kinds of kernels to both types of SVR and after applying data reduction with sample size 150

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	88	0.3436	0.06392	0.88681	63	3.15563	0.029858	0.898509
	Polynomial	112	0.4064	0.07561	0.88069	64	55.5271	0.525387	0.892637
	Radial	89	0.363	0.06757	0.85871	68	2.80651	0.026555	0.749613
	Sigmoid	119	5.374	1	0.02122	62	105.688	1	0.02733
FA	Linear	95	2.203	0.006363	0.4779	62	0.69385	0.13384	0.472918
	Polynomial	98	21	0.060659	0.4184	66	1.87524	0.36172	0.377412
	Radial	96	2.978	0.008602	0.4349	71	0.56026	0.10807	0.444192
	Sigmoid	113	346.2	1	0.0260	61	5.18416	1	0.205365

It can be concluded that for sample size n=30 the results improved after applying the FA more than the PCA, it is clear that the RMSE with linear and radial kernel functions gave good results, also an approximate value for both SVR. With real data, sample size (50,100), it's obvious clear that after applying FA and the v –SVR gave better results and for sample size 150 after applying PCA and the ε –SVR we get better the RMSE. For more information about before applying data reduction checks the Appendix.

4. Conclusions

It's important not to lose more information than is necessary, when reducing the data dimensions. Principal Component Analysis is a well-established mathematical technique for reducing the dimensionality of data, while keeping as much variation as possible as we note in practical section. It is also known that using of SVR with various

kernel functions improves the estimation of models. The behavior of two different models ε-SVR and v-SVR are compared through an extensive real data and simulation study under four different kernel functions: linear, radial, polynomial, and sigmoid kernel functions, with different sample sizes ranges. Generally, it can be concluded that according to the reduction of SVR, after applying PCA and with increase sample sizes, under ε-SVR. But under v-SVR, the results of sigmoid and polynomial kernel functions were the best between other counterparts .But with regard to the value of RMSE, under ε-SVR, for sample size greater than or equal 150, results improved. And, it is clear that the RMSE with linear kernel function gave the best values rather than other kernel functions.

References

1. Chowdhury, U. N., Chakravarty, S. K., and Hossain, M. T. (2018). Short-term financial time series forecasting integrating principal component analysis and independent component analysis with support vector regression. Journal of Computer and Communications, 6(03), 51.

2. Yu, H., Chen, R., and Zhang, G. (2014). A SVM stock selection model within PCA. Procedia computer science, 31, 406-412

3. Glaser, J. I., Benjamin, A. S., Farhoodi, R., and Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. Progress in neurobiology, 175, 126-137.

4. Lee, J. A., and Verleysen, M. (2009). Quality assessment of dimensionality reduction: Rank-based criteria. Neurocomputing, 72(7-9), 1431-1443.

5. Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202.

6. Rosipal, R., Girolami, M., Trejo, L. J., and Cichocki, A. (2001). Kernel PCA for feature extraction and de-noising in nonlinear regression. Neural Computing and Applications, 10(3), 231-243.

7. Shi, S., Li, G., Chen, H., Hu, Y., Wang, X., Guo, Y., and Sun, S. (2018). An efficient VRF system fault diagnosis strategy for refrigerant charge amount based on PCA and dual neural network model. Applied Thermal Engineering, 129, 1252-1262.

8. Jolliffe, I. T. (1986). Principal components in regression analysis. In Principal component analysis (pp. 129-155). Springer, New York, NY.

9. Chao, D., Zhou, W., Ye, C., Zhang, Q., Chen, Y., Gu, L., and Qiao, S. Z. (2019). An electrolytic Zn–MnO2 battery for high-voltage and scalable energy storage. Angewandte Chemie International Edition, 58(23), 7823-7828.

10. Vapnik, V. (2013). The nature of statistical learning theory. Springer science and business media.

11. Mechelli, A., and Vieira, S. (Eds.). (2019). Machine learning: methods and applications to brain disorders. Academic Press.

12. Astuti, W. (2018, March). Support vector machine and principal component analysis for microarray data classification. In Journal of Physics: Conference Series (Vol. 971, No. 1, p. 012003). IOP Publishing.

13. Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1997). Support vector regression machines. Advances in neural information processing systems, 9, 155-161.

14. Jiang, M., Zhu, L., Wang, Y., Xia, L., Shou, G., Liu, F., and Crozier, S. (2011). Application of kernel principal component analysis and support vector regression for reconstruction of cardiac transmembrane potentials. Physics in Medicine and Biology, 56(6), 1727.

15. Shokri, S., Sadeghi, M. T., Marvast, M. A., and Narasimhan, S. (2015). Integrating principal component analysis and vector quantization with support vector regression for sulfur content prediction in HDS process. Chemical Industry and Chemical Engineering Quarterly, 21(3), 379-390.

16. Chowdhury, U. N., Rayhan, M. A., Chakravarty, S. K., and Hossain, M. T. (2017). Integration of principal component analysis and support vector regression for financial time series forecasting. International Journal of Computer Science and Information Security (IJCSIS), 15(8).

17. Vapnik, V. N. (1998). Adaptive and learning systems for signal processing communications, and control. Statistical learning theory.

Appendix (A)

Table (A.1) Simulation results for the RMSE to all kinds of kernels to both types of SVR and before applying data reduction with sample size 30.

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	18	1.526	0.8947	1	12	0.90834	0.91741	0.999999
	Polynomial	23	1.4893	0.8731	0.6897	20	0.990107	1	0.585139
	Radial	18	1.6413	0.9623	0.9063	18	0.90788	0.91695	0.799017
	Sigmoid	24	1.7057	1	0.5265	24	0.980555	0.99035	0.575994
FA	Linear	15	0.45217	0.58374	0.9101	13	0.00566	0.00678	0.98615
	Polynomial	22	0.60955	0.78692	0.47863	20	0.83469	1	0.79553
	Radial	18	0.54872	0.70839	0.1665	18	0.74132	0.88813	0.61521
	Sigmoid	24	0.7746	1	0.59872	24	0.80555	0.96508	0.58601

Table (A.2) Simulation results for the RMSE to all kinds of kernels to both types of SVR and before applying data reduction with sample size 100.

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	73	0.9158624	0.27289	0.02634888	46	0.568143	0.67388	0.02796826
	Polynomial	73	2.261642	0.67388	1	45	0.453264	0.53762	0.99982250
	Radial	77	1.0545485	0.31421	0.01638800	48	0.843081	1	0.03507764
	Sigmoid	77	3.3561050	1	0.68972290	56	0.465766	0.55245	0.86821400
FA	Linear	72	0.0848327	0.04599	0.25716614	45	0.8442265	0.43649	0.25716614
	Polynomial	72	1.8442265	1	0.99992310	46	1.848327	0.95564	1.00000000
	Radial	76	0.9341080	0.506503	0.01196654	46	1.1613841	0.60047	0.01196654
	Sigmoid	76	1.1613841	0.62974	0.8164677	64	1.934108	1	0.69609480

Table (A.3) Simulation results for the RMSE to all kinds of kernels to both types of SVR and before applying data reduction with sample size 150.

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	93	1.3707	0.55497805	0.4596	68	1.92882	1	0.4499
	Polynomial	96	1.26785	0.51333537	0.1544	94	1.80689	0.93678917	0.1556
	Radial	95	1.36821	0.55396907	0.7244	89	1.61874	0.83924223	0.724
	Sigmoid	100	2.46982	1	0.3244	63	1.84037	0.95414283	0.1615
FA	Linear	95	1.34361	0.24997763	0.4779	61	5.18417	1	0.2054
	Polynomial	98	1.40643	0.26166489	0.4185	62	0.69385	0.13384	0.4729
	Radial	96	1.36318	0.25361917	0.4349	66	1.87524	0.361725	0.3774
	Sigmoid	113	5.37491	1	0.0261	71	0.56027	0.108073	0.4442

Table (A.4) Real data results for the RMSE to all kinds of kernels to both types of SVR and before applying data reduction with sample size 30.

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	22	0.6599	0.0537	0.4423	18	0.6847	0.9357	0.453
	Polynomial	24	12.292	1	0.0193	24	0.4604	0.6291	0.0191
	Radial	21	0.6991	0.0569	0.5357	22	0.7317	1	0.5332
	Sigmoid	24	0.4948	0.0403	0.4748	14	0.4963	0.6783	0.4453
FA	Linear	20	0.6422	0.5372	0.3726	13	0.6568	0.4636	0.3819
	Polynomial	24	1.1955	1	0.1587	15	1.4165	1	0.1604
	Radial	21	0.4221	0.3531	0.4325	16	0.6824	0.4818	0.4208
	Sigmoid	23	1.1634	0.9731	0.0431	11	1.2746	0.8998	0.0608

Table (A.5) Real data results for the RMSE to all kinds of kernels to both types of SVR and before applying data reduction with sample size 100.

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	73	1.1343	0.5305	0.2853	73	1.9594	0.6247	0.9998
	Polynomial	73	1.1957	0.5592	0.1727	73	3.1365	1	0.8682
	Radial	73	1.1516	0.5386	0.2768	73	1.5879	0.5063	0.9781
	Sigmoid	76	2.1381	1	0.0196	76	2.297	0.7324	0.488
FA	Linear	43	1.1371	0.5566	0.2794	49	2.8525	0.8573	1
	Polynomial	47	1.2264	0.6003	0.1841	45	3.3273	1	0.6897
	Radial	49	1.1025	0.5396	0.2813	56	2.648	0.7958	0.9063
	Sigmoid	76	2.0431	1	0.0104	76	2.9052	0.8731	0.5265

Table (A.6) Real data results for the RMSE to all kinds of kernels to both types of SVR and before applying data reduction with sample size 150.

Data reduction	Kernel	ε –SVR				v-SVR			
		No. SVR	RMSE	RE of RMSE	R^2	No. SVR	RMSE	RE of RMSE	R^2
PCA	Linear	105	1.0559	0.2098	0.1124	93	1.0530	0.2545	0.117
	Polynomial	104	1.1954	0.2375	0.0122	98	1.1514	0.2783	0.0275
	Radial	101	1.0702	0.2126	0.1023	97	1.0464	0.2529	0.1288
	Sigmoid	112	5.0328	1	0.1062	109	4.1376	1	0.0819
FA	Linear	104	1.9594	0.6247	0.9998	105	2.8525	0.8573	1
	Polynomial	105	3.1365	1	0.8682	109	3.3273	1	0.8053
	Radial	100	1.5879	0.5063	0.9781	108	2.648	0.7958	0.9865
	Sigmoid	112	2.2970	0.7324	0.4880	109	2.9052	0.8731	0.0026