Support vector regression is an improvement for principle component analysis

Mohamed, A. M.[1] Abdel Latif, S. H.[2] Alwan, A. S.[3*]

[1]Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt
[2]Faculty of Graduate Studies for Statistical Research,Cairo University, Cairo, Egypt
[3]College of Administration & Economics, Sulaimani University, Sulaimani , Iraq
[1]dramanymousa04@gmail.com,[2]shereen_hamdy_@hotmail.com
*[3]*asraa.alwan@univsul.edu.iq*

Abstract:

The principle component analysis is used more frequently as variables reduction technique. And recently, an evolving group of studies make use of machine learning regression algorithms to improve the estimation of empirical models. One of the most frequently used machine learning regression model is support vector regression with various kernel functions. However, ensemble of support vector regression and principal component analysis is also possible. So, this paper aims to investigate the competence of support vector regression techniques after performing principal component analysis to explore the possibility of reducing data and having more accurate estimations. Some new proposals are introduced and the behavior of two different models ε-SVR and v-SVR are compared through an extensive simulation study under four different kernel functions; linear, radial, polynomial, and sigmoid kernel functions, with different sample sizes ranges from small, moderate to large. The models are compared with their counterparts in terms of coefficient of determination (R^2) and root mean squared error (RMSE).The comparative results show that applying SVR after PCA models improve the results in terms of SV numbers between 30% and 60% in average and it can be applied with real data. In addition, linear kernel function gave the best values rather than other kernel functions and the sigmoidkernel gave the worst values. Under ε-SVR the results improved which did not happen with v-SVR. It is also drawn that, RMSE valuesdecreased with increasing sample size.

Keywords: ε-Support Vector Regression, Kernel Functions, Principal Component Analysis,v -Support Vector Regression.

1. Introduction

The recent trends in collecting huge and diverse datasets,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[3].

PCAhelps 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[11],[6],[7],[12],[13].

PCA can be widely applied in all forms of analysis from neuroscience tocomputer graphics and in a variety of real-world applications including image segmentation [15], climate research [8], genome-wide expression studies [14], and deep learning [2] due to its superior properties, such as linear un-correlation, low-dimensionality and visualization in multivariate data, over other linear dimension reduction (LDR) methods[10],[13].

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[10]. SVMis a supervisedlearning model, with associated learning algorithm that analyze data usedfor classification, known as SV classifier, and regression (function approximation), known as support vector regression (SVR) [16].

During past few decades, an extension to the SVM classification algorithmhas been received a considerable attention, see [1], which is mainly due to VladimirVapnik and co-workers Harris Drucker, Christopher J. C. Burges, Linda Kaufman and Alexander J. SmolaIn 1996 for introducingSVM for regression, known as ε-SVR model,that handles regression problems [6]. SVR has additional advantages compared to other regression methods, see[16].

PCAis a widely applied feature extraction method in the framework of SVR. In the literature,[4]proposed an integration of PCA and SVR, or PCA-SVR,toenhance 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, [16]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 $\varepsilon -$SVR. The other is also applying the PCA but within $v -$SVR.

The rest of the paper is organized as follows: Section 2 presents themethodology thatused in this paper PCA and SVR. Section 3discussed the results and evaluation. Finally, conclusion and future work are given in Section 4.

2. Methodology

The data used in the PCA were recorded from a Monte Carlo simulationtechnique. The simulation was designed in the R software using R package. 5000 simulations with 1000 observations were carried out. The data followingnormal distribution. Eight sample sizes (50, 100, 250, 500, 1000, 1750, and 2000) were used in this study. The dataset is 9-dimensional with 9 variables giving 9 principal components.Several analyses with two, three, and four features were performed to obtain the effective parameters on the fragmentation; these parameters were selectedas input parameters for the predictive models.

In this study, two main procedures were conducted on the same dataset. The first procedure was preceded in two steps: First, PCA was performed on a set of parameters– input (features) and output to estimate how far the effect of PCA on several sample sizes by calculatingrootmean square error (RMSE). Second, ε −SVR was performed to show its effect, using a set of mathematical functions known as positive definite kernels– linear, polynomial, radial, and sigmoid. The second procedure followed the same steps of the first procedure but applying ν −SVR instead of ε −SVR.The next sections will cover all of these methods.

2.1. Principal Component Analysis (PCA)

Many raw data sets have a high-dimension space, and are accordingly difficult to interpret. In addressing this issue, PCA finds smaller number of uncorrelated components from high dimensional original inputs by calculating the eigenvectors of the covariance matrix. Given a set of m dimensional input vectors $x_i = (x_i(1),\ x_i(2), \dots , x_i(m))^T$where $i = 1,2, \dots , n$. PCA is a transformation of x_i into a new vector y_i by:

$$y_i = U^T x_i \quad (1)$$

where$U$is the m × morthogonal matrix whose$j$th column u_j is the jth eigenvector of the sample covariance matrix $C = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$. In other words, PCA solves the following eigenvalue problem:

$$\lambda_j u_j = C u_j\ , j = 1,2, \dots , m \quad (2)$$

whereλ_j is one of the eigenvalues of Candu_j is the corresponding eigenvector.

Based on the estimated u_j, the components of y_i are calculated as the orthogonal transformation of x_i.That is,

$$y_i(j) = u_j^T x_i, j = 1,2, \dots , m \ (3)$$

The new components are called principal components. By using only the first several eigenvectors sorted in adescending order of eigenvalues, the number of principal components in y_i can be reduced [11]. Thus, PCA can be used to reduce dimensions where the principal components are uncorrelated and have sequentially maximum variances.

2.2. Support Vector Regression (SVR)

The SVR extends the basic principles of SVM for classification [5] 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.

2.2.1. ε-SVR Model

ε-SVR maps the input vectors $x_i \in R^m$ into a high dimensional feature space. Given a training set (x_i, y_i), $i = 1,2,\ldots,n$, 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

$$y = f(x) = w,\ x + b = w^T x + b \ (4)$$

where$w$ is the weight vector corresponding to x and b isthe bias. The Vapnik's linearε -Insensitivity loss (error) function is also given as

$$L(y, f(x)) = \begin{cases} 0 & if|y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & otherwise \end{cases} (5)$$

Based on the above, the linear regression $f(x)$ is estimated by simultaneously minimizing $||w||^2$ 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 w is a design parameter chosen by the user.

$$R = \frac{1}{2}||w||^2 + c\left(\sum_{i=1}^{n} |y - f(x)|_\varepsilon\right) \ (6)$$

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

$$\text{Minimize}R = \frac{1}{2}||w||^2 + c\sum_{i=1}^{n}(\xi_i + \xi_i^*) \ (7)$$

$$subject\ to \begin{cases} (w^T x_i + b) - y_i \leq \varepsilon + \xi_i \\ y_i - (w^T x_i + b) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0\ , i = 1,2,\ldots,m \end{cases} (8)$$

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 functionto obtain the desired weight vector of the regression function[17].

In Equation (4), the generalization performance of such linearfunction,$f(x)$, is fairly limited and unable to reflect thetrue regression procedure. In order to overcome suchweakness, a standard mathematical solution is the introductionof kernel function,φ(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 φ(X). The most popular kernel functions used in this study areshownin Table (2.1).

Table (2.1) 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

2.2.2. v -SVR Model

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 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 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

$$min_w \frac{1}{2}||w||^2 + C(v\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi + \xi^*))$$

$$subjectto \begin{cases} y_i - w^T\varphi(x_i) - b \leq \varepsilon + \xi \\ w^T\varphi(x_i) + b - y_i \leq \varepsilon + \xi^* \text{ (9)} \\ \xi, \xi^*, \varepsilon \geq 0 \end{cases}$$

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 $\alpha, \alpha^*, \eta, \eta^*andb$ gives

$$L(w,\xi,\xi^*,\alpha,\alpha^*,\eta,\eta^*,\beta) = \frac{1}{2}||w||^2 + Cv\varepsilon + \frac{C}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$+\sum_{i=1}^{l}\alpha_i^*(y_i - w^Tx_i - b - \varepsilon - \xi_i) + \sum_{i=1}^{l}\alpha_i(w^Tx_i + b - y_i - \varepsilon - \xi_i^*)$$

$$-\sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*)\beta\varepsilon(10)$$

Following the KKT conditions that partial derivatives with respect tothe variables $w, b, x, x^*, and\varepsilon$ 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

$$max_{\alpha,\alpha^*} \sum_{l}^{l_{sv}} y_i(\alpha_i - \alpha_i^*) - \frac{1}{2}\sum_{i}^{l_{sv}}\sum_{j}^{l_{sv}}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j),$$

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

$$\sum_{i}^{l_{sv}}(\alpha_i - \alpha_i^*) = 0,\ \alpha_i, \alpha_i^* \in \left[0, \frac{C}{l}\right], \sum_{i}^{l_{sv}}(\alpha_i - \alpha_i^*) \leq Cv(11)$$

Then, the regression estimate takes the form

$$f(x) = \sum_{i}^{l_{sv}}(\alpha_i - \alpha_i^*)\,k(x_i, x) + b(12)$$

Compared to the optimization problem in ε-SVR [10], we can see thatthe parameter ε vanishes but instead there is the new parameter v in v-SVR [9]. 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. Simulation Study: Results and Discussion

This simulation was conducted for the purpose of comparison between ε-SVR and$v$-SVR models after and before applying PCA using four different kernel functions to detect the PCA effect on data reduction. In this study, all trained models designed are evaluated using measured data based on two different measures, root mean square error (RMSE), and coefficient of determination (R^2). 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. The sample sizes are arbitrarily determined to represent small, moderate, and large sample sizes such as; $n = 50$,100, 250, 500, 1000,1500,1750, and 2000. The simulation results were based on10000 replications. All computations are using the R program (R x64 3.2.5) version.

Two types of procedures have been adopted within the practical aspect.The first procedure lays in studying and applying the data reduction method, PCA, within ε-SVR model with four different kernel functions; linear, polynomial, redial, sigmoid, which was referred to previously through which we can use this to simulate a large number of hypothetical cases that may arise within the practical aspect. Thus, generalizing the results becomes more comprehensive. According to the importance of the components, table (a) [SeeAppendix (a)], show that components one and two together capture from 86% to 95%.

3.1. Effect on Support Vector Regression Number:

Tables from (b.1) to (b.8)[See Appendix (b)], show that before using PCA, the ε-SVR application reduced the number of SVR by 30% to 40%, with the linear kernel being superior to its counterparts at most sample sizes, and it reached 70% for sample size 100, and the polynomial kernel showed good performance with large sample sizes 1750 and 2000.But the v-SVR application reduced the number of SVR by a rate ranging from 53% to 60% with the sigmoid kernel function being superior to its counterparts at all most all sample sizes,the performance of all functions converged.And after applying PCA with ε-SVR, the percentage was about 30% with the radial kernel function superior to its counterparts with all most all sample sizes, and only in one case the decrease in the number of SVR reached 56% at $n = 1750$.But with$v$-SVR, the results of sigmoid and linear kernel functions were equal for 50% of the cases, and the percentage ranged from 54% to 60%, and the differences between them and polynomial were very narrowfollowed by the radial kernel function.

3.2. Effect of Sample Size

The following two figures illustrated the effect of sample size on the RMSE values.

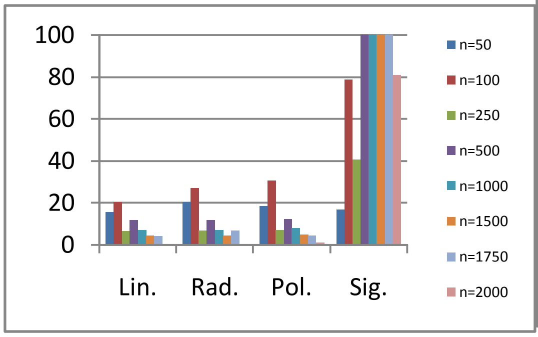


Figure (3.1) the effects of n under ε-SVR before using PCA

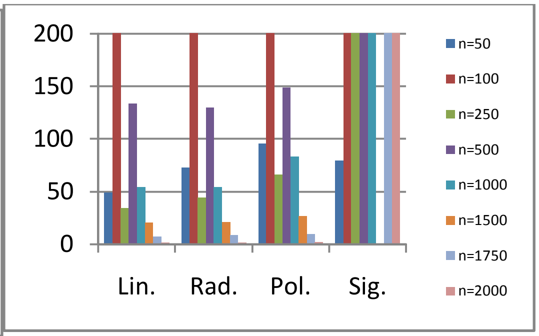


Figure (3.2) the effects of n underν-SVR before using PCA

Under ε-SVR and ν-SVR and with all sample sizes from $n = 50$ to $n = 2000$, after using PCA, and from figures (3.1) and (3.2), it is clear that the root mean square error with almost different kernel functionsdecreased with increasing sample size except for sample size $n = 100$ and $n = 500$. But the worst values of RMSE were at $n = 100$and $n = 500$.A sample size $n = 250$ gave good results for linear, radial, and polynomial kernel functions between sample sizes of 50 to 1000. [See Appendix (c), figures (c.1) and (c.2) for before using PCA case]

3.3. Effect of PCA

The following figures show the effects PCA with small, moderate, and large sample size underε-SVRbefore and after using PCA.

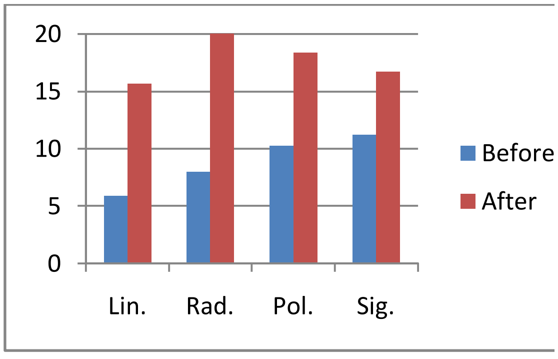


Fig. (3.3) effect of PCA for ε-SVRat n=50

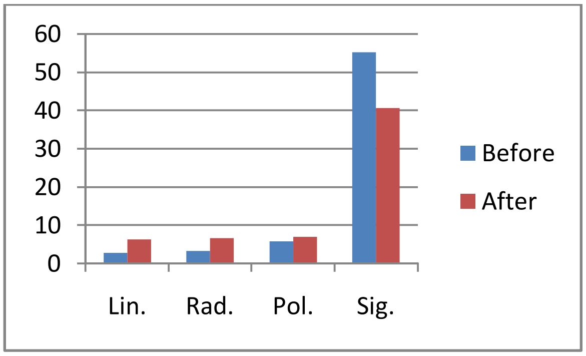
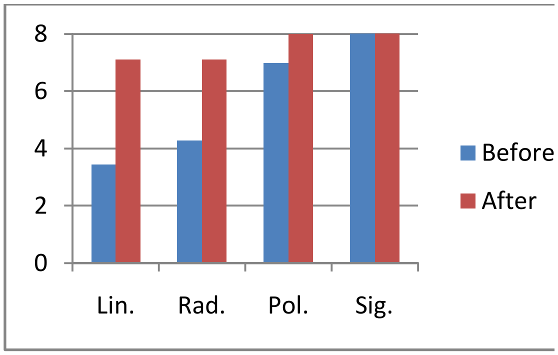


Fig. (3.4) effect of PCA for ε-SVR at n=250



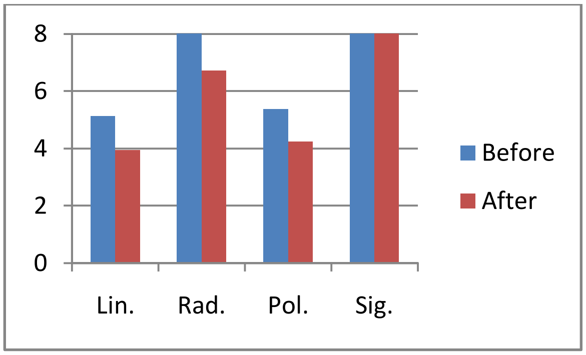


Fig. (3.5) effect ofPCA for ε-SVRat n=1000 Fig. (3.6) effect of PCA for ε-SVRat n=1750

Under ε-SVR, and with sample size $n = 50, 250, 1000,1750$, before and after using PCA, and from figure (3.3), figure (3.4), figure (3.5) and figure (3.6) respectively, it can be concluded that for sample size greater than or equal 1500, results improved after applying PCA. And, it is clear that the root mean square error with linear kernel function gave the best values other than other kernel functions. In addition, the sigmoidkernel gave the worst values for RMSE amongst the other kernel functions.[See Appendix(c), figures from (c.3) to (c.6) for $n \ = 100, 500, 1500, 2000$]. The following figures show the effect of PCA under ν -SVR.

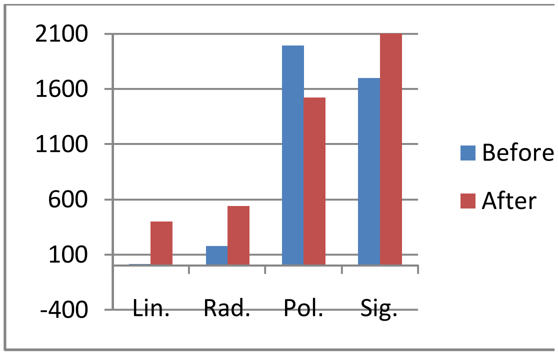


Fig. (3.7) effect of PCA for ν-SVRat $n = 100$

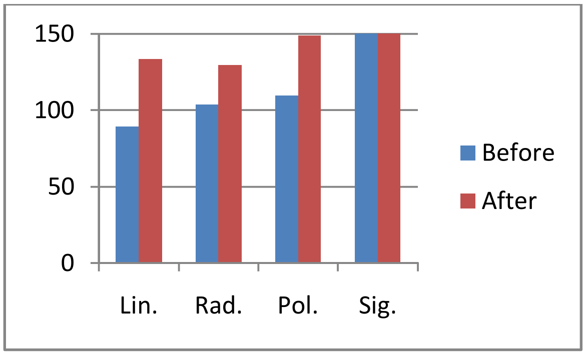


Fig. (3.8) effect of PCA for ν-SVR at $n = 500$

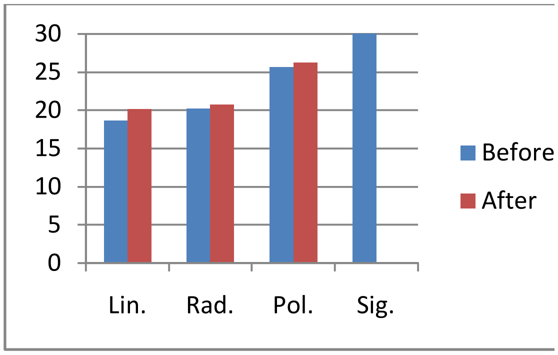


Fig. (3.9) Effect of PCA forν-SVR at $n =$ 1500

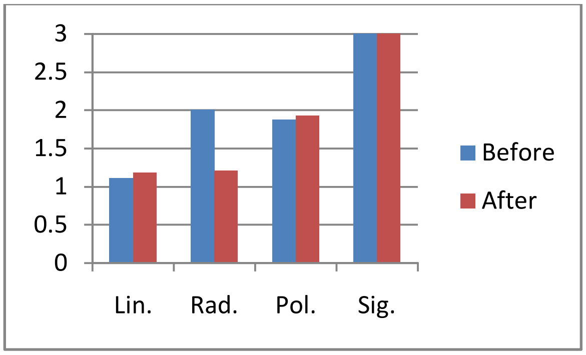


Fig. (3.10) Effect of PCA forν-SVRat $n = 2000$

As it can be seen, under ν-SVR, and with sample size $n = 100, 500, 1500,2000$, from figure (3.7) to figure (3.10), it can be concluded that the ν-SVR results did not improve after applying PCA with all sample sizes.Kernel function gave the best values rather than other kernel functions. In addition, the sigmoidkernel gave the worst values for RMSE amongst the other kernel functions. [See Appendix (c), figures from (c.7) to (c.10) for$n = 50, 250, 1000, 1750$]

3.4. Effect of Support Vector Regression

The following figures showed thatwith all types of kernel functions and at all sample sizesε-SVR is better than ν-SVR after using PCA, except for only one case for$n = 1500$ and sigmoid. These results were similar to results obtained before using PCA [See Appendix (c), figures from (3.11) to(3.14)].

Figure (3.7) the effects ofε-SVR andν-SVR with all kernels functions, and for all sample sizes,after using PCA

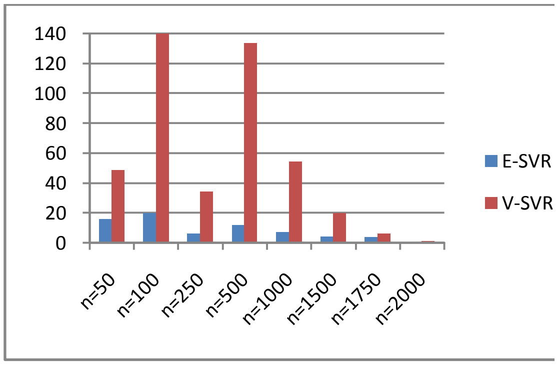


Fig. (3.11) Effect of SVR for Linear kernel

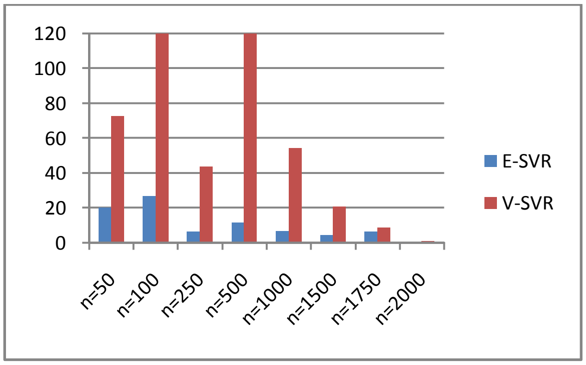


Fig. (3.12) Effect of SVR for Radial kernel

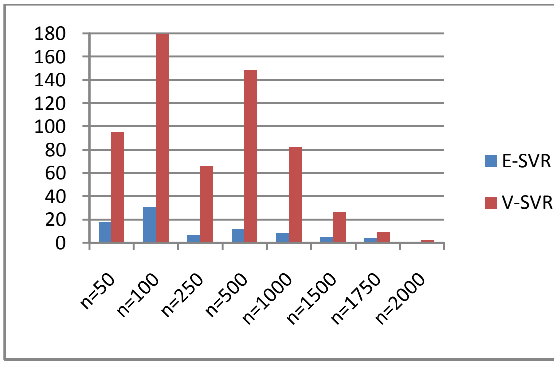


Fig. (3.13) Effect of SVR for Polynomial kernel

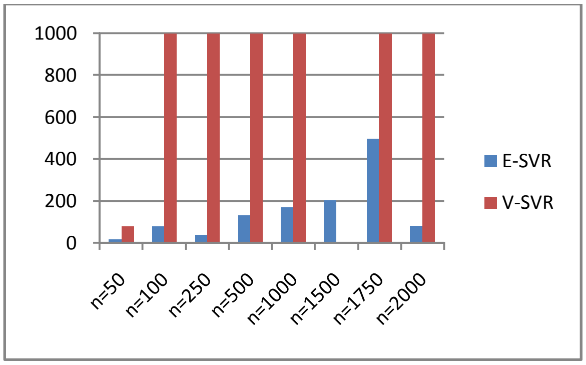


Fig. (3.14) Effect of SVR for Sigmoid kernel

3.5. Effect of kernel function

The following to figures (3.15) and (3.16) illustrates the effects of kernel function on RMSE with ε-SVR andν-SVR, at all sample sizes, after using PCA.

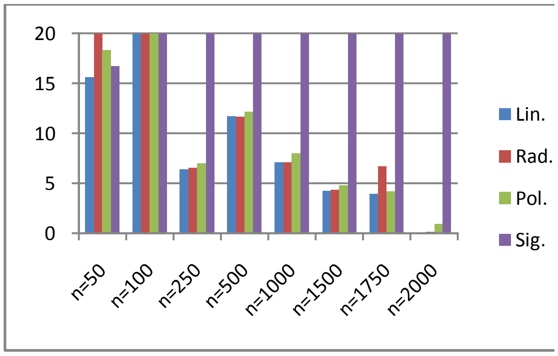


Fig. (3.15)ε-SVR after PCA

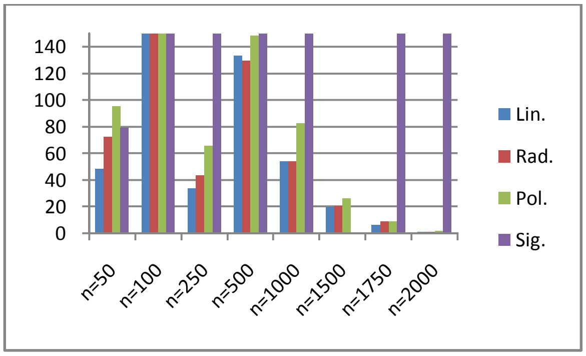


Fig. (3.16) ν-SVR after PCA

Under ε-SVR, and for all sample sizes from $n = 50$ to $n = 2000$, after using PCA, and from figure (3.15), it is clear that the linear kernel function gave the best of root mean square error then the radial kernel function for all sample sizes except at $n = 50$, it gave the worst value. But the sigmoid kernel function gave the worst value of RMSE with all sample sizes (except at $n = 50$). It is worth noting that with the increase in the sample size, the results of linear, radial and polynomial are close to each other, except for radial when the sample size $n = 1750$. The results of v-SVR are very similar to the results of ε-SVR except for sample size $n =$ 100 the polynomial kernel function gave the worst error value, and for sample sizes $n =$ 500 the radial kernel function gave the best error value, figure (3.16).These results are similar to results before using PCA. [See Appendix (c), figures (3.15) and (3.16)].

4. General Conclusions

When reducing the data dimensions, it's important not to lose more information than is necessary. Principal Component Analysis (PCA) is a well-established mathematical technique for reducing the dimensionality of data, while keeping as much variation as possible as we notes 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 simulation study under four different kernel functions; linear, radial, polynomial, and sigmoid kernel functions, with different sample sizes ranges from small, moderate to large. Generally,it can be concluded that according to the reduction of SVR, after applying PCA and with all sample sizes, underε-SVR, the percentage of reduction was about 30% with the radial kernel function. But under$v$-SVR, the result of sigmoid and linear kernel functions were the best between other counterparts, and the percentage ranged from 54% to 60%. But with regard to the value of RMSE, under ε-SVR, for sample size greater than or equal 1500, results improved. And, it is clear that the RMSE with linear kernel function gave the best values rather than other kernel functions. In addition, the sigmoidkernel gave the worst values amongst the other kernel functions. But under v-SVR, results did not improve after applying PCA with all sample sizes. It is also drawn that, with ε-SVR and v-SVR, the RMSE with almost different kernel functionsdecreased with increasing sample size which is considered as an indicator to the consistency. In addition, from sample sizes ranges from $n = 50$to $n = 1000$, the sample size $n = 250$ gave good results for linear, radial, and polynomial kernel functions.

References

1. Boser, Bernhard E.; Guyon, Isabelle M.; Vapnik, Vladimir N. (1992). "A training algorithm for optimal margin classifiers". Proceedings of the fifth annual workshop on Computational learning theory – COLT '92.p. 144. CiteSeerX 10.1.1.21.3818. doi:10.1145/130385.130401. ISBN 978897914970. S2 CID 207165665
2. Chan, T. H., Jia, K., Gao, S., Lu, J., Zeng, Z., & Ma, Y. (2015). PCANet: A simple deep learning baseline for image classification?. IEEE transactions on image processing, 24(12), 5017-5032.

3. Chao, D., Zhou, W., Ye, C., Zhang, Q., Chen, Y., Gu, L., ...&Qiao, S. Z. (2019). An Electrolytic Zn–MnO2 Battery for High-Voltage and Scalable Energy Storage. AngewandteChemie International Edition, 58(23), 7823-7828.
4. Chowdhury, U. N., Chakravarty, S. K., &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.
5. Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. Machine Learning, 20,273-297. https://doi.org/10.1007/BF00994018
6. Engelbrecht, A. P. (2007). Computational intelligence: an introduction. John Wiley & Sons.
7. Glaser, J. I., Benjamin, A. S., Farhoodi, R., &Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. Progress in neurobiology, 175, 126-137.
8. Hannachi, A., Jolliffe, I. T., & Stephenson, D. B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: A review. International Journal of Climatology: A Journal of the Royal Meteorological Society, 27(9), 1119-1152.
9. Higuchi, I., &Eguchi, S. (2004). Robust principal component analysis with adaptive selection for tuning parameters. Journal of Machine Learning Research, 5(May), 453-471.
10. Hua, X. G., Ni, Y. Q., Ko, J. M., & Wong, K. Y. (2007). Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. Journal of Computing in Civil Engineering, 21(2), 122-135.
11. Jolliffe, I. T. (1986). Principal components in regression analysis. In Principal component analysis (pp. 129-155). Springer, New York, NY.
12. Jolliffe, I. T., &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.
13. Lee, J. A., &Verleysen, M. (2009). Quality assessment of dimensionality reduction: Rank-based criteria. Neurocomputing, 72(7-9), 1431-1443.
14. Ringnér, M. (2008). What is principal component analysis?. Nature biotechnology, 26(3), 303-304.
15. Shi, S., Li, G., Chen, H., Hu, Y., Wang, X., Guo, Y., & 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.
16. Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In Machine Learning (pp. 123-140). Academic Press.
17. Zhang, Y., d’Aspremont, A., & El Ghaoui, L. (2012). Sparse PCA: Convex relaxations, algorithms and applications. In Handbook on Semidefinite, Conic and Polynomial Optimization (pp. 915-940). Springer, Boston, MA.

Appendix (a)

Table (a): Evaluation of componentsfor all sample size

	n	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	50	2.399582	1.493121	0.850262	0.466969	0.263411
Proportion of variance		0.639777	0.247712	0.080327	0.024229	0.00771
Cumulative Proportion		0.639777	0.887489	0.967817	0.992046	0.999755
Standard deviation	100	2.405992	1.406846	0.941638	0.540262	0.194417
Proportion of variance		0.6432	0.219913	0.09852	0.032431	0.0042
Cumulative Proportion		0.6432	0.863113	0.961633	0.994064	0.998264
Standard deviation	250	2.4141047	1.6512802	0.49868025	0.43419256	0.08290096
Proportion of variance		0.6475446	0.3029696	0.02763133	0.02094702	0.00076361
Cumulative Proportion		0.6475446	0.9505142	0.97814554	0.99909256	0.99985618
Standard deviation	500	2.251667	1.696724	0.999785	0.221098	0.051637
Proportion of variance		0.563334	0.319875	0.111063	0.005432	0.000296
Cumulative Proportion		0.563334	0.883208	0.994272	0.999703	0.999999
Standard deviation	1000	2.124317	1.818052	0.977601	0.410422	0.209954
Proportion of variance		0.501414	0.367257	0.106189	0.018716	0.004898
Cumulative Proportion		0.501414	0.868671	0.97486	0.993576	0.998474
Standard deviation	1500	2.637062	1.097019	0.91294	0.083799	0.036359
Proportion of variance		0.772678	0.133717	0.092607	0.00078	0.000147
Cumulative Proportion		0.772678	0.906394	0.999001	0.999781	0.999928
Standard deviation	1750	2.581409	1.22201	0.699998	0.447875	0.374477
Proportion of variance		0.740408	0.165923	0.054444	0.022288	0.015581
Cumulative Proportion		0.740408	0.906331	0.960776	0.983063	0.998645
Standard deviation	2000	2.601504	1.028339	0.904634	0.485987	0.333533
Proportion of variance		0.751981	0.117498	0.090929	0.026243	0.012361
Cumulative Proportion		0.751981	0.869479	0.960408	0.98665	0.999011

Appendix (b)

Table (b.1):Application results ofε-SVR andν-SVR with n=50

Kernel	Using PCA	ε-SVR=50			ν-SVR=50		
		No. SVR	RMSE		No. SVR	RMSE	
Linear	Before	35	5.905735	0.611466	26	4.487617	0.648434
	After	34	15.63424	0.920805	22	6.973119	0.922534
Polynomial	Before	41	10.24338	0.518209	31	10.64776	0.525611
	After	36	18.34726	0.824364	22	9.766403	0.820838
Radial	Before	37	7.983526	0.317025	30	10.44179	0.367461
	After	36	20.07923	0.109865	23	8.525746	0.150681
Sigmoid	Before	37	11.18416	0.783092	24	10.15318	0.752485
	After	40	16.73461	0.692276	20	8.911357	0.737407

Table (b.2): Application ofε-SVR andν-SVR with$n = 100$

Kernel	Using PCA	ε-SVR=100			ν-SVR=100		
		No. SVR	RMSE	R^2	No. SVR	RMSE	R^2
Linear	Before	30	3.812108	0.989790	44	3.898366	0.989469
	After	70	20.12219	0.734026	42	19.94299	0.730193
Polynomial	Before	70	45.75615	0.618775	48	44.65717	0.584493
	After	72	30.64740	0.380796	43	38.99390	0.114258
Radial	Before	42	22.98349	0.645508	51	13.33926	0.644507
	After	70	26.87716	0.475720	46	23.15135	0.449809
Sigmoid	Before	80	74.41939	0.044944	41	41.14962	0.000311
	After	72	78.73017	0.105730	42	75.49266	0.107149

Table (b.3): Application results of ε-SVR andν-SVR with n=250

Kernel	Using PCA	ε-SVR=250			ν-SVR=250		
		No. SVR	RMSE	R^2	No. SVR	RMSE	R^2
Linear	Before	149	2.7591288	0.896497	103	3.0581533	0.8846127
	After	179	6.4037621	0.44426	102	5.8314337	0.5741451
Polynomial	Before	184	5.87614584	0.613779	105	7.9868842	0.3640643
	After	182	7.0393593	0.35889	104	8.12225091	0.2861072
Radial	Before	156	3.3402499	0.858116	111	4.5304271	0.7494853
	After	176	6.6047535	0.40594	108	6.5998772	0.4548942
Sigmoid	Before	196	55.216609	0.001994	102	40.6345665	0.008469012
	After	199	40.581904	0.00173	102	38.9927044	0.0165334

Table (b.4):Application results ofε-SVR andν-SVR with$n = 500$

Kernel	Using PCA	ε-SVR=500			ν-SVR=500		
		No. SVR	RMSE	R^2	No. SVR	RMSE	R^2
Linear	Before	361	9.975281	0.343531	203	9.466509	0.329342
	After	366	11.73772	0.098096	202	11.55629	0.329342
Polynomial	Before	360	11.99407	0.146126	205	10.463957	0.126680
	After	357	12.17787	0.02367	203	12.19348	0.1266801
Radial	Before	360	11.44221	0.152811	218	10.176497	0.146262
	After	368	11.65505	0.125827	214	11.388279	0.146262

Sigmoid	Before	399	122.0327	0.035544	202	102.1085	0.0025699
	After	394	130.80145	0.010471	201	90.1976	0.00256991

Table (b.5): Application results ofε-SVR andν-SVR with $n = 1000$

Kernel	Using PCA	ε-SVR=1000			ν-SVR=1000		
		No. SVR	RMSE	R^2	No. SVR	RMSE	R^2
Linear	Before	618	3.42460	0.8826332	405	3.100595	0.882584
	After	720	7.08211	0.5049097	401	7.3756314	0.504739
Polynomial	Before	687	6.975845	0.560541	413	8.443815	0.547097
	After	721	7.98206	0.3652298	402	9.083121	0.350718
Radial	Before	636	4.2676199	0.825030	438	4.0220206	0.818858
	After	717	7.1008020	0.504075	413	7.3552450	0.501977
Sigmoid	Before	794	174.4608	0.0968525	402	161.9323	0.053553
	After	795	171.090	0.0099719	401	191.52642	0.009081

Table (b.6): Application results ofε-SVR andν-SVR with $n = 1500$

Kernel	Using PCA	ε-SVR=1500			ν-SVR=1500		
		No. SVR	RMSE	R^2	No. SVR	RMSE	R^2
Linear	Before	1069	4.29137	0.4949637	603	4.32192	0.497399
	After	1080	4.30776	0.4872366	602	4.48947	0.4874544
Polynomial	Before	1076	4.979287	0.333197	605	5.064822	0.344293
	After	1084	4.82179	0.351424	603	5.12483	0.3504718
Radial	Before	1067	4.3989396	0.466456	623	4.501377	0.464170
	After	1079	4.3563	0.474433	618	4.5555559	0.463043
Sigmoid	Before	1199	314.6709	0.1760055	601	216.9770	0.0941660
	After	1190	205.508	0.0081285	602	80.25321	0.03242593

Table (b.7):Application results of ε-SVR and v-SVR with n=1750

Kernel	Using PCA	ε-SVR=1750			ν-SVR=1750		
		No. SVR	RMSE	R^2	No. SVR	RMSE	R^2
Linear	Before	1080	5.12654	0.9296202	708	3.034645	0.9312248
	After	1027	3.94297	0.6098332	702	2.5561089	0.672139
Polynomial	Before	821	5.38031	0.9234148	711	3.0423366	0.9240209
	After	1083	4.234208	0.5439352	704	3.0099337	0.5328131
Radial	Before	1333	8.289448	0.9816035	724	3.13559	0.9827222
	After	769	6.7023958	0.8830206	709	2.974966	0.8840259
Sigmoid	Before	1381	577.73834	0.0121804	702	69.768746	0.00003041
	After	1394	496.99376	0.1055873	702	48.677263	0.09880482

Table (b.8): Application results ofε-SVR andν-SVR with$n = 2000$

Kernel	Using PCA	ε-SVR=2000			ν-SVR=2000		
		No. SVR	RMSE	R^2	No. SVR	RMSE	R^2
Linear	Before	1080	1.0057917	0.9994493	805	1.0542278	0.99333
	After	1394	0.0539351	0.6137618	802	1.0873407	0.9999
Polynomial	Before	1260	1.449058	0.6674626	813	1.3706637	0.6756977
	After	1447	0.9712849	0.3253675	804	1.3895499	0.6756977
Radial	Before	1389	1.0335056	0.9883338	949	1.4172537	0.9933313
	After	1408	0.1752729	0.5916619	815	1.1007692	0.9933313
Sigmoid	Before	1598	109.76821	0.1382282	802	69.33145	0.06405554
	After	1591	80.915245	0.0192846	803	65.374727	0.06405554

Appendix (c)

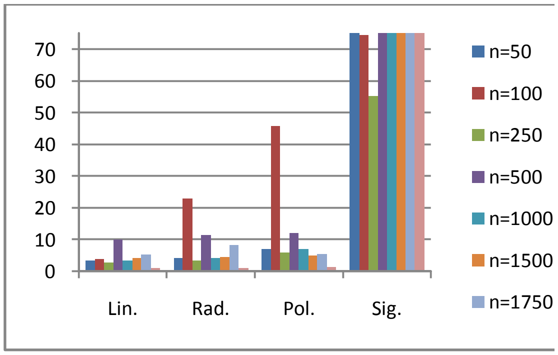


Fig. (c.1) Effect of PCA for ε-SVRat n=1000

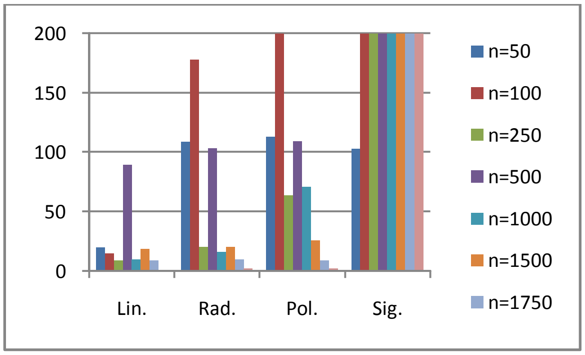
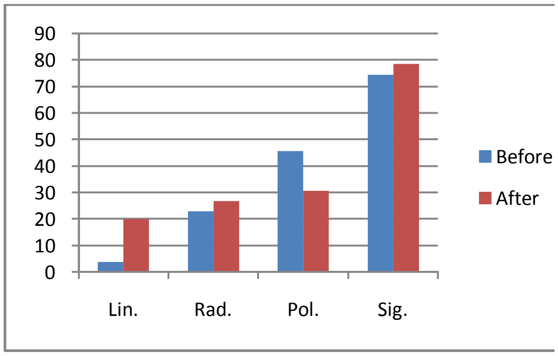


Fig. (c.2) Effect of PCA for ε-SVR at n=1000



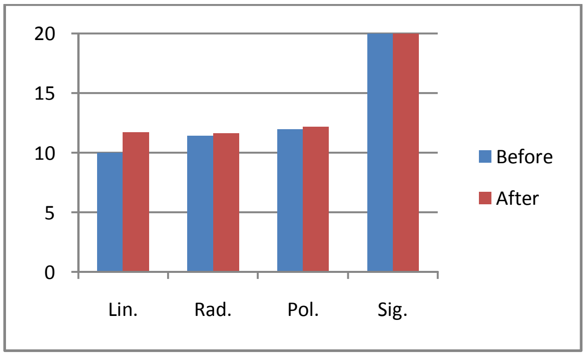


Fig. (c.3) Effect of PCA for ε-SVRat n=100Fig.(c.4) Effect of PCA for ε-SVR at

n=500

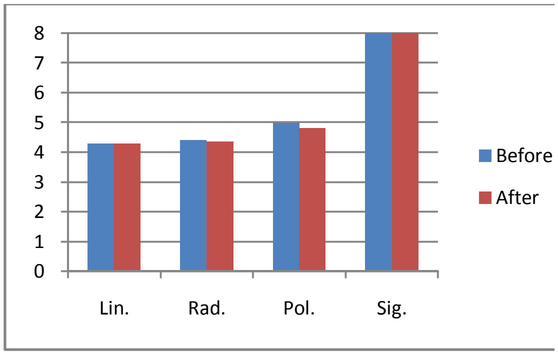


Fig. (c.5) Effect of PCA for ε-SVR at $n =$ 1500

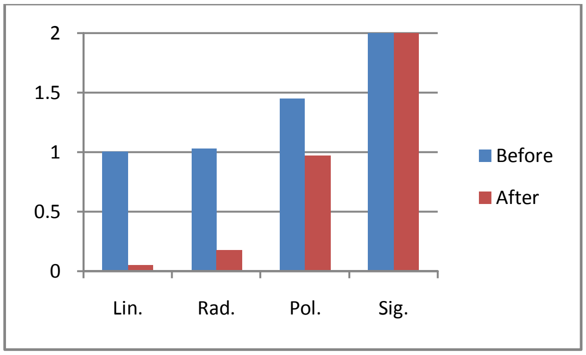
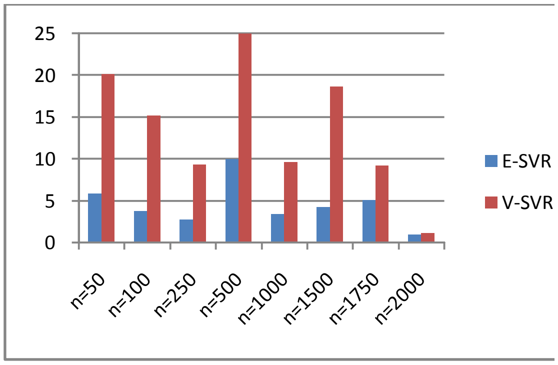


Fig.(c.6) Effect of PCA for ε-SVR at $n =$ 2000



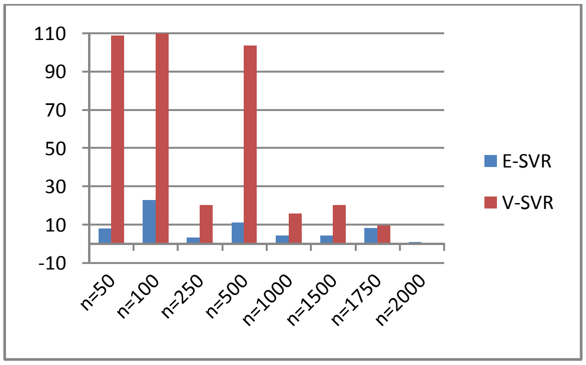
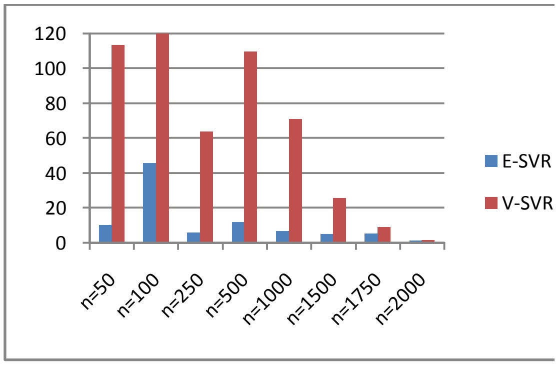


Fig.(c.9) effect of PCA for$v$-SVRat $n = 50$Fig. (c.8) effect of PCA for$v$-SVRat

$n = 250$



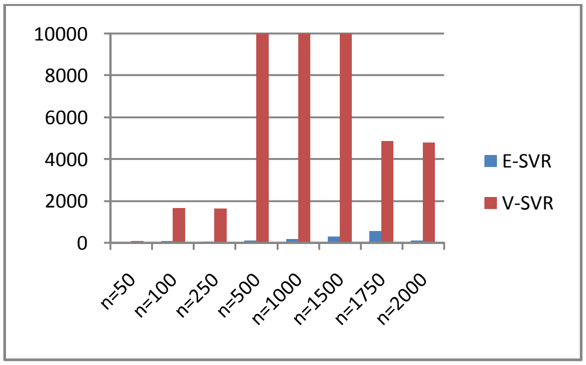


Fig. (c.9) effect of PCA for$v$-SVRat n=1000 Fig. (c.10) effect of PCA for v-SVRat n=1750

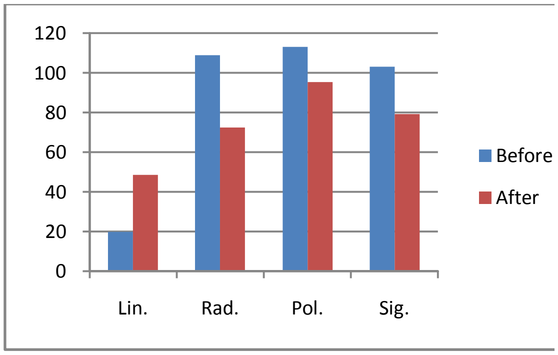


Fig.(c.11) Effect of SVR for Linear kernel

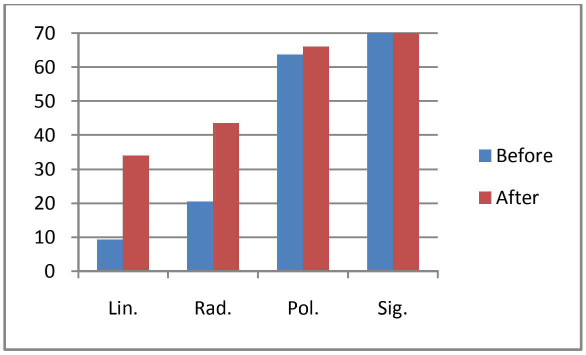


Fig. (c.12) Effect of SVR for Radial kernel

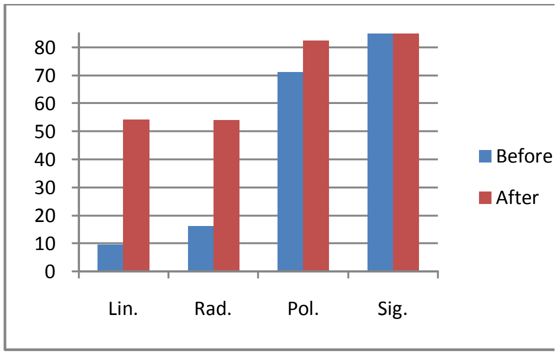


Fig. (c.13) Effect of SVR for Polynomial kernel

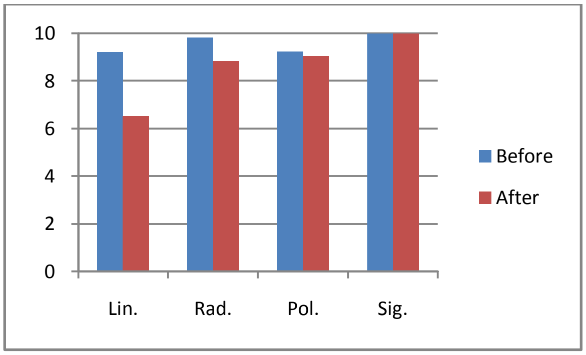


Fig. (c.14) Effect of SVR for Sigmoid kernel

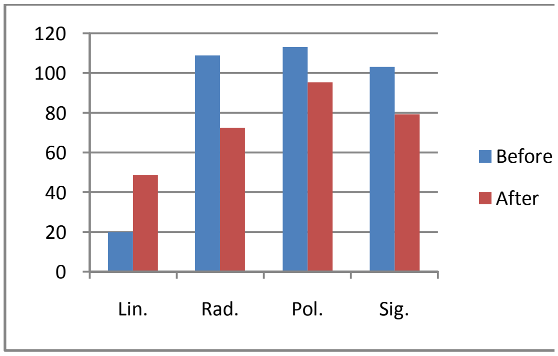


Fig. (c.15) ε-SVR before PCA

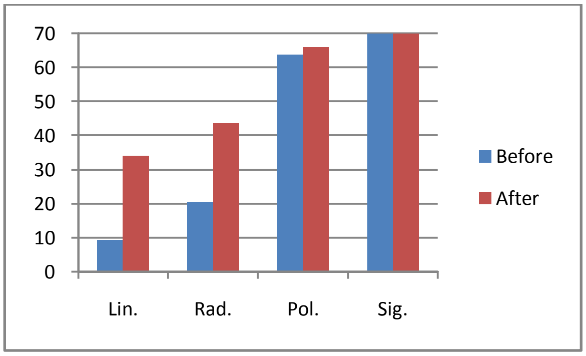


Fig. (c.16) ν-SVR before PCA