

Original Research Article

<https://doi.org/10.20546/ijcmas.2019.802.137>

Estimation and Comparison of Support Vector Regression with Least Square Method

S. Vishnu Shankar^{1*}, G. Padmalakshmi² and M. Radha³

¹Agricultural Statistics, ²Agricultural Economics, ³Faculty of Agricultural Statistics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

*Corresponding author

ABSTRACT

Keywords

Least square, Support vector regression, Root mean square error

Article Info

Accepted:

10 January 2019

Available Online:

10 February 2019

Regression is one among the most used vital machine learning and statistical tool. Regression is a method of modeling a target value based on independent predictors. It allows making predictions from data by understanding the relationship between features of data and observed continuous-valued response. Support Vector Regression (SVR) is one of the useful and flexible techniques, helping the user to deal with the limitations pertaining to distributional properties of underlying variables, the geometry of the data and the common problem of model overfitting. In this paper an attempt has been made to establish the significance of SVR through the numerical study. A 34 years of Metrological data is used here to compare Support Vector Regression with Least Square Regression. Based on the numerical study SVR model is identified as best fit by using Relative Mean Square Error (RMSE).

Introduction

Regression analysis is a statistical method that allows studying the relationship between two or more variables of interest. This analysis helps to determine the factors of interest and its influence on other variables. There are different types of regression analysis which are used for a specific purpose. All these regression methods examine the influence of one or more independent variables on a dependent variable. In general regression analysis is used with experimentally manipulated variables and naturally-occurring variables. The main disadvantage of

regression is that it cannot determine the causal relationships among the variables.

The least square technique is a widely used regression technique. It is a Conventional method that assume a linear relationship between the input variables and the single output variable i.e. x and y.

But all regression problems cannot be described using a linear model. So Support Vector Regression is used which avoid the difficulties of using linear functions in the high dimensional feature space and optimization problem.

Support Vector Machines (SVMs) is one of the Machine learning techniques. It is mainly used for classification purpose in Data sciences. Support vector Regression can be also used as Regression technique i.e. Support Vector Regression (SVR) based on the nature of data. SVR is considered as a nonparametric technique because it mostly counts on kernel functions. These include the selection of few model parameters, avoidance of overfitting to the data and unique, optimal and global solution. The Support Vector Regression (SVR) uses the principles as same as SVM for classification, with only a few difference. Support Vector Regression is of two types i.e. linear and Non-linear. Non-linear SVR is performed using kernel function.

In this article, least square method is compared with the support vector regression. For the comparison of these two methodologies, 34 years of Metrological data of Coimbatore district i.e. Rainfall as predictor variable and Evapotranspiration as response variable is used here. The superiority one technique over another is shown here by Root Mean Square Error (RMSE). “Root mean square error or Root mean square deviation is the measure of the differences between values (sample or population values) predicted and the values observed. RMSE is always non-negative, and a value of 0 would indicate a perfect fit to the data” by Vladimir N. Miaorov and Gordon M. Crippen (1994).

Materials and Methods

Theory and basic principles

(i) Least square method

Least squares are a form of mathematical regression analysis that hooks the line of best fit for a dataset. It demonstrates the relationship between the data points visually through graphs and charts. Each individual

point of data is characteristic of the relationship between x and y i.e. known independent variable and an unknown dependent variable. The least squares method is popular for finding a curve fitting a given data. Say $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ be n observations from the data.

$$y=f(x)=ax+b$$

Now at $x=x_1$ while the observed value of y is y_1 , the expected value of y from the curve above is $f(x_1)$. Let us define the residual for n by

$$e_n=y_n-f(x_n)$$

Some of the residuals may be positive and some may be negative. While finding the curve fitting the given data the residual at any x_i should be as small as possible. Now since some of the residuals are positive and others are negative and equal importance is given to all the residuals as it is desirable to consider the sum of the squares of these residuals, say E and thereby find the curve that minimizes E . Thus, we consider

$$E=\sum_{i=1}^n e_i^2$$

The best representative of curve $y=f(x)$ is by minimizing $E=\sum_{i=1}^n e_i^2$. By solving the normal equations we can find the a and b values of the equation.

(ii) Support vector regression

Support Vector Regression performs linear regression in the high-dimension using insensitive loss function and tries to reduce model complexity. This is described by introducing (non-negative) slack variables (ξ_i, ξ_i^*) which measure the deviation of training samples outside ε -insensitive zone. In ε -SV regression (Vapnik, 1995), the objective is to find a function $f(x)$ that has at

most ε deviation from the actually obtained targets y_i for all the training data. In other words, the errors are acceptable as long as they are less than ε , but will not accept any deviation larger than this. For linear functions f ,

$$f(x) = \langle w, x \rangle + b \text{ with } w \in X, b \in R (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product in X . Flatness in the above equation means that the w should be small. One way to ensure this is to minimize the norm, i.e. $\|w\|^2 = \langle w, w \rangle$. This can be written as a convex optimization problem:

$$\begin{aligned} &\text{Minimize} && \frac{1}{2}(\|w\|^2) \\ &\text{Subject to} && \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \end{aligned}$$

The tacit assumption in subject to was that such a function factually exists that estimates all pairs (x_i, y_i) with ε precision, or in other words, that the convex optimization problem is *feasible*. Sometimes, this may not be the case, or some errors are allowed. The “soft margin” loss function which was used in support vector machines by Cortes and Vapnik (1995), shows that slack variables ξ_i, ξ_i^* is introduced to cope with infeasible constraints of the optimization problem (subject to). Hence the formulation stated in Vapnik (1995),

$$\begin{aligned} &\text{Minimize} && \frac{1}{2}(\|w\|^2) + C \sum (\xi_i + \xi_i^*) \\ &\text{Subject to} && \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned}$$

The constant $C > 0$ concludes the trade-off between the flatness of f and the amount up to which deviations larger than ε are accepted. This corresponds to dealing with a so called ε -insensitive loss function $|\zeta|_\varepsilon$ described by

$$|\zeta|_\varepsilon = \begin{cases} 0 & \text{if } |\zeta| \leq \varepsilon \\ |\zeta| - \varepsilon & \text{otherwise} \end{cases}$$

Simply, Linear SVR is

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b$$

In Non-Linear SVR, the kernel functions transform the data into a dimension of higher feature space which makes it possible to perform the linear separation

$$\begin{aligned} y &= \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \\ y &= \sum_{i=1}^N (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x) \rangle + b \end{aligned}$$

It is well known that SVM performance (estimation accuracy) depends on a good setting of meta-parameters parameters like C, ε and the kernel parameters. Selection of kernel type and kernel function parameters is generally based on domain knowledge. Parameter C governs the tradeoff between the model complexity (flatness) and the degree to which deviations larger than ε are tolerated in optimization formulation. The parameter ε controls the width of the ε -insensitive zone, used to fit the data. The kernel function used here is Radial Basis Kernel function i.e.

$$y(x) = \sum_{i=1}^N \omega_i \phi(\|x - x_i\|)$$

Results and Discussion

Least square method is widely used statistical technique for fitting the best line. In simple regression, the error rate should be as low as possible. But now a day due to complexity of data, LS could not attain it correctly. So many

alternate methods for fitting the best line was found. One of that is Support Vector Regression. SVR is an advantageous and flexible technique, with the limitations concerning to distributional properties of variables, the geometry of the data and

overfitting problem in the model. The selection of kernel function in the model is critical. While LS cannot capture the nonlinearity in a dataset, SVR becomes convenient in such situations (Fig. 1).

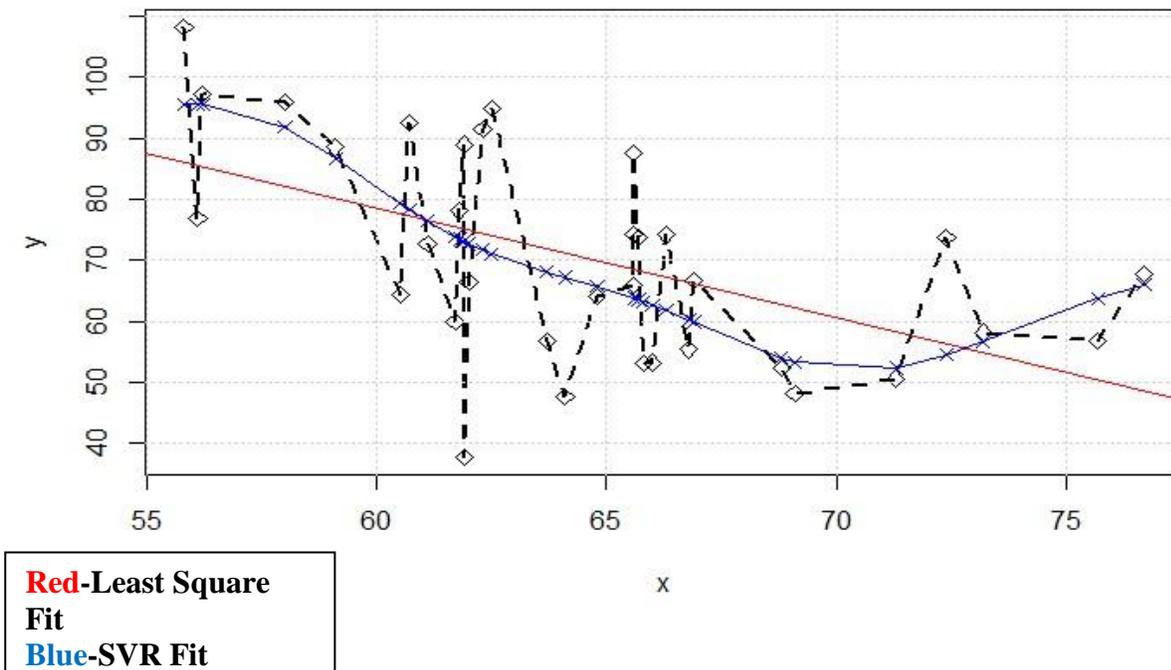
Table.1 Estimated parameter values of LS method

Regression Type	Simple Linear
Intercept	186.2134723
Coefficient of x	-1.792803232
P-Value of x	0.000972
RMSE	14.32

Table.1 Estimated parameter values of SVR model

SVM Type	eps-Regression
C	1
E	0.799
Kernel type	Radial Basis Function
Loss function	ϵ -insensitive
The quantity of support vectors	33
RMSE	13.12

Figure.1 Original +Least Square +SVR



Based on the regression fit it can be identified that the relationship between response and predictor variables are non linear. So the comparison of the LS method with SVR is made here for the data and the results are given in the table 1 and 2. At first, the Least Square Method resulted with the RMSE value of 14.32. Next SVR method was performed with the Radial Basis Kernel Function and resulted with the RMSE value of about 13.12. SVM package e1071 was used to perform SVR.

The performance of SVR Model has been assessed through regression model. Estimated parameter results are obtained by using R software. It is interesting to note that the method of least square method yields the high RMSE value compared with SVR for 34 years of Metrological data of Coimbatore district. Based on the statistical evaluation, the SVR method founds superior to LS method. The study establishes the fact that the performance of performance of Least Square and Support Vector Regression is almost identical with SVR having a slight edge over least square. Hence it is concluded that the Support Vector Regression Model can be considered as modification of the Least Square procedure and such procedures may not fail when there is non-linearity in the dataset.

Acknowledgement

In preparation of my Research paper, I had to take the help and guidance of some respected persons, who deserve my deepest gratitude. As the completion of this paper gave me much pleasure, I would like to show my gratitude to Dr.M.Duraisamy Prof.& Head TNAU, Dr.Patil Santosh, Assistant Prof. TNAU and all other Staffs for giving me a good guidelines for Research paper throughout numerous consultations. Many people, especially Vijay Kumar Selvaraj - Data Science Lead on I Nurture ltd, Naffees

Gowsar S R, Nandhini C, Gomathi T, Nivedha R, Mano Chitra K, Muthu Prabakaran K, Arulpandiyan K, Arulprabhu K, Vinoth S.K, Aravind. K and Naveena. R, have made valuable comment suggestions on my paper which gave me an inspiration to improve the quality of the Research Paper.

References

- Chu, H., Wei, J., Li, T., and Jia, K. (2016). Application of Support Vector Regression for Mid-and Long-term Runoff Forecasting in “Yellow River Headwater” Region. *Procedia Engineering*, 154, 1251-1257.
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Kavitha, S., Varuna, S., and Ramya, R. (2016, November). A comparative analysis on linear regression and support vector regression. In *Green Engineering and Technologies (IC-GET), 2016 Online International Conference on* (pp. 1-5). IEEE.
- Liu, Z., and Xu, H. (2014). Kernel parameter selection for support vector machine classification. *Journal of Algorithms & Computational Technology*, 8(2), 163-177.
- Meyer, D., and Wien, F. T. (2001). Support vector machines. *R News*, 1(3), 23-26.
- Parveen, N., Zaidi, S., and Danish, M. (2016). Support vector regression model for predicting the sorption capacity of lead (II). *Perspectives in Science*, 8, 629-631.
- Smola, A.J., and Scholkopf, B. (2004), A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222.
- Vladimir N. Miorov, Gordon M. Crippen (1994), Significance of Root-Mean Square deviation in comparing three dimensional structures of globular

proteins. *J.Mol.Biol*, 625-634.

Ye, Z., and Li, H. (2012, October). Based on radial basis Kernel function of support vector machines for speaker

recognition. In *Image and Signal Processing (CISP), 2012 5th International Congress on* (pp. 1584-1587). IEEE.

How to cite this article:

Vishnu Shankar, S., G. Padmalakshmi and Radha, M. 2019. Estimation and Comparison of Support Vector Regression with Least Square Method. *Int.J.Curr.Microbiol.App.Sci*. 8(02): 1186-1191. doi: <https://doi.org/10.20546/ijcmas.2019.802.137>