Predictive Modeling Using Support Vector Regression

Muthukrishnan. R, Maryam Jamila. S

Abstract: Regression analysis is a statistical process for estimating the relationships among variables. The main focus of the analysis is on the relationship between a dependent variable and one or more independent variables. Usually classical regression is used to fit a model and predict using it for a given data. There is a need of a better procedure for estimating the model that will give more accurate results on prediction when the classical procedure turns out bad. Thus, we explore Support Vector Regression, a type of Support Vector Machine, which attempts to minimize the generalized error so as to achieve better performance. It is an advanced technique of regression that performs well with large number of observations. This study compares the classical Regression with Support Vector Regression using datasets containing one and more than one predictor. The accuracy of regression procedures is studied with RMSE and R^2 with the help of R software.

Keywords: Linear Regression, Robust Regression, Support Vector Regression, Kernel, Support Vectors, RMSE, R^2.

1 INTRODUCTION

The concept of regression analysis supports the need of modeling and prediction. In the era of big data, dealing with a large amount of data is a big deal. The classical regression methods used to perform well while the data set is large and also its tolerance towards contamination is very poor. That implied to the development robust regression procedures. The purpose of this paper is to reveal the efficiency of support vector regression over robust regression and linear regression. The method of support vector machine (SVM) has the foundation of the concept of hyperplane. The maximal margin hyperplane, which is the separating hyperplane that is far away from the training observations. A test observation then could be classified based on which side of the hyperplane it lies. This is called the maximal margin classifier. Its generalization to the non-separable case is called the support vector classifier. The SVM is an extension of the support vector classifier which results from enlarging the feature space using kernels. In this procedure the feature space is enlarged so that it could accommodate a non-linear boundary between the classes. The leftover content of this paper is constructed as follows. The regression procedures, linear regression and robust regression, used for comparing with support vector regression is explained briefly in section 2. An experiment is carried out to expose the efficiency of support vector.

2 REGRESSION METHODS

Regression analysis is a statistical method that deals with investigating relationship between the dependent and independent variables on a dataset. The methods used in this paper, for comparing with support vector regression are linear and robust regressions, are discussed in this section regression in section 3. Section 4 comprises of conclusion of the work done in this paper.

- *Muthukrishnan. R, Department of Statistics, Bharathiar University, India.*
- *Maryam Jamila. S, Department of Statistics, Bharathiar University, India. E-mail: smaja9555@gmail.com.*

2.1 LINEAR REGRESSION

Linear regression (Linear Model - LM) is a regression technique which produces the model having linear combinations of the regression parameter. This method was initially developed by Legendre [7] and Gauss [2] and the predominantly used linear regression procedure is least squares method, which was published by Gauss [3]. The general form of linear regression model with p independent variables is

$$y = \beta_0 + \beta_1 x + \varepsilon. \quad (1)$$

In the method of least squares the coefficients β are taken to minimize the residual sum of squares

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - x_i^T \beta)^2. \quad (2)$$

This classical method regression fails to succeed in the presence of a contaminated data point. Thus, there was a need to construct a better regression model that can tolerate contamination.

2.2 ROBUST REGRESSION

To overcome the intolerance of the classical regression procedure, robust regression techniques were developed. Robust regression (Robust Linear Model - RLM) is done by using several robust regression estimators, namely M-estimator, MM-estimator, least trimmed squares, least of median squares, etc. For comparing with support vector regression MM-estimator is used. MM-estimator, was proposed by Yohai [15], is calculated as a solution of minimization

$$\sum_{i=1}^{n} \rho\left(\frac{Y_i - x_i{}'\beta}{\widehat{\sigma_n}(\widehat{\beta_n})}\right) = min, \ \ \beta \epsilon \mathbb{R}^p \quad (3)$$

where $\widehat{\beta_n}$ is obtain from the computation of S-estimator [11] that involves the minimization of estimator of the scale.

2.3 SUPPORT VECTOR REGRESSION

The primary SVM algorithm was formulated by Vapnik and Chervonenkis [14] in 1964. Boser et. al. [1] recommended a way to develop non-linear classifiers to maximize the margin of the hyperplanes by applying the kernel trick. The current incarnation (soft margin) was proposed by Cortes and Vapnik in 1993 and published in 1995. Smola and Scholkopf [12] elucidates the mathematical formulation of the Support Vector Regression (SVR) model. Suppose the training data given are $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset \mathcal{X} \times \mathbb{R}$ where $\mathcal{X}$ denotes the space of the input pattern. The function f can be described as

$$f(x) = \langle w, x \rangle \text{ with } w \epsilon \mathcal{X}, b \epsilon \mathbb{R} \tag{4}$$

where $\langle \cdot,\cdot \rangle$ denotes the dot product in $\mathcal{X}$. The small w can be obtained by minimizing the norm (ie) $\|w\| = \langle w, w \rangle$. This problem can be written as a convex optimization problem

$$\left.\begin{array}{l} \text{minimize } \frac{1}{2}\|w\|^2 \\ \text{subject to} \begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon \\ \langle w, x_i \rangle + b - y_i \le \varepsilon \end{cases} \end{array}\right\} \tag{5}$$

After introducing the slack variables to (5) a Lagrange function is derived from the objective function, which in turn solves to find w as $w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) x_i$. And thus, the function f becomes

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

For computing b, Karush-Kuhn-Tucker condition ([5],[6]) was exploited. After the computational process of b and the construction of the regression model, the examples that come with the non-vanishing coefficients are called the support vectors. More number of support vectors explains the relationship more accurately. The support vector method uses the concept of kernel to covert the given data into higher dimension. There are four main types of kernels used, namely linear, polynomial, sigmoid and radial basis function kernel. Kernel used in this paper for the numerical study is radial basis function kernel,

$$K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2) \tag{6}$$

3 NUMERICAL STUDY

The performance of the Support Vector Regression against linear and robust regression by providing results using real datasets containing one and more than one predictor is discussed in this section.

3.1 CASE 1

Data for the Hertzsprung-Russell Diagram of the Star Cluster CYG OB1, which contains 47 stars in the direction of Cygnus is used for this experiment. It has one predictor, logarithm of the effective temperature at the surface of the star (log.Te) and a dependent variable, logarithm of its light intensity (log.light). For the dataset, simple linear regression (LM), robust linear regression (RLM) and support vector regression (SVR) analysis are performed. There are 9 outlying observations in the dataset when detected using cook's distance. The root mean squared error and r-square values are computed for data with and without outliers and summarized in table 1.

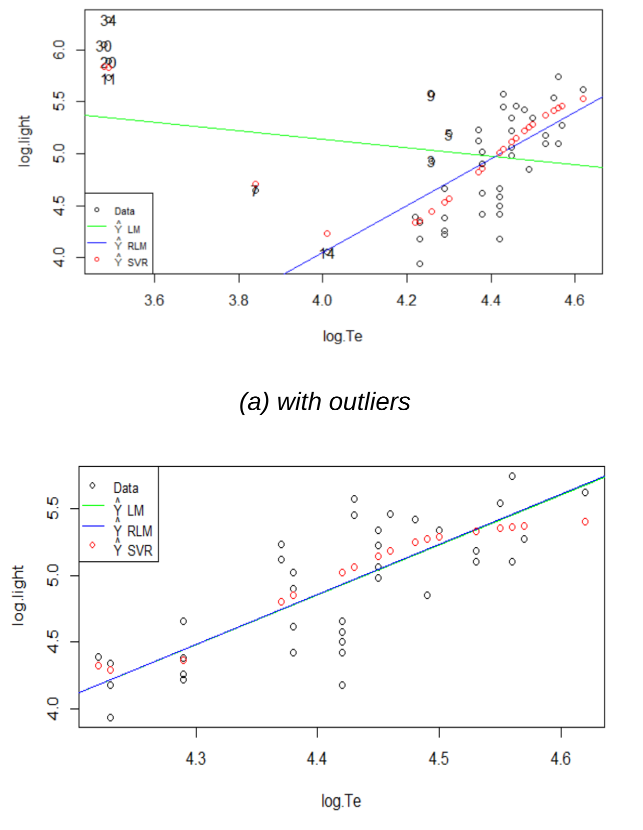
TABLE 1

RMSE AND R^2 (STARSCYG DATA)

Regression Methods	RMSE	R^2
LM	0.552 (0.309)	0.044 (0.612)
RLM	0.553 (0.309)	0.044 (0.612)
SVR	0.359 (0.315)	0.597 (0.595)

(.) without outliers

For the support vector regression model, the given dataset has 45 support vectors and when the outliers are removed the number of support vectors are found to be 37.



(a) with outliers

(b) without outliers

Fig. 1 - Scatter plot with regression fits

From the fig. 1(a), the performance of simple linear regression is not appreciable, whereas robust linear regression attempts to cover majority of the data points. And it is also clear that the prediction based on support vector regression method is more accurate than robust linear regression. Using Cooks distance procedure, the outliers are removed and the three regression methods are performed for the dataset without outliers. Fig. 1(b) clearly shows that the support vector regression method performs well no matter whether the dataset contains outliers or not. For the original data, robust linear regression gives a better fit but it is not exposed neither by RMSE nor by R2. This is evident on working with several other datasets, thus concluding that the error measures and goodness of fit measures do not completely reveal the efficient performance of the regression methods.

3.2 CASE 2

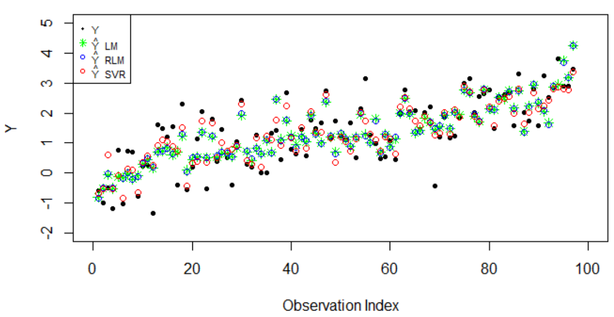
This experiment deals with Prostate dataset that came from a study that examined the correlation between the level of prostate specific antigen and a number of clinical measures in men who were about to receive a radial prostatectomy. The dataset has 97 observations each having 8 independent variables namely lweight (log of prostate weight), age, lbph (log of benign prostatic hyperplasia amount), svi (seminal vesicle invasion), lcp (log of capsular penetration), gleason (Gleason score), pgg45 (percentage Gleason scores 4 or 5), lpsa (log of prostate specific antigen) and one dependent variable lcavol (log of cancer volume). This dataset contains 9 outliers when it is checked using cook's distance.

TABLE 2
RMSE AND R^2 (PROSTATE DATA)

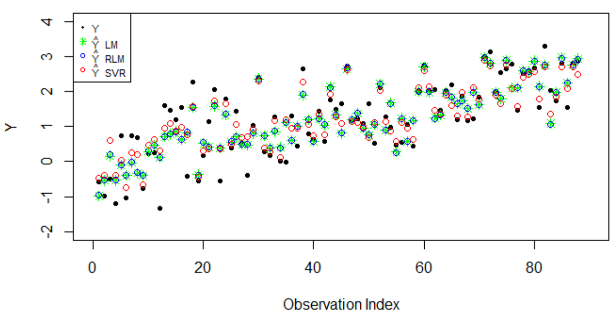
Regression Methods	RMSE	R^2
LM	0.667 (0.557)	0. 676 (0.747)
RLM	0.667 (0.557)	0.677 (0.746)
SVR	0.548 (0.490)	0.781 (0.803)

(.) without outliers

From table 2 it is clear that support vector regression shows minimum root mean square error and greater r-square when compared with other two methods. Actual values are plotted against the predicted ones for the data and are presented in fig. 2 (a).



(a) with outliers



(b) without outliers

Fig. 2 *- Actual vs predicted values (Prostate data)*

Predicted values using the three regression techniques are plotted against the actual values of the dependent variable for the dataset without outliers in fig. 2 (b). Support vector regression performs efficiently and similarly despites the contamination in the dataset.

4 CONCLUSION

It is clear that support vector regression is more efficient in minimizing the error of the regression model. More the number of support vectors states that the model of support vector regression to be more efficient. The performance of SVR may not be clearly displayed in the scatter plot for dataset containing more than one predictor, whereas for dataset with one predictor is distinctly visible that the prediction is more accurate than other regression procedures. By the analysis of many other experiments it is clear that support vector regression works the best for dataset with greater number of observations.

REFERENCES

[1] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers", Proceedings of the Annual Conference on Computational Learning Theory, ACM Press, Pittsburgh, Pennsylvania, pp. 144-152, 1992.
[2] C. F. Gauss, "Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum", Werke, vol.7, pp. 1-261, 1809.
[3] C. F. Gauss, "Theoria combinationis observationum erroribus minimis obnoxia: Pars posterior", Werke, vol.4, pp. 95-100, 1821.
[4] B. Hamner, M. Frasco, and E. LeDell, "Metrics: Evaluation Metrics for Machine Learning", R package version 0.1.4, https://github.com/mfrasco/Metrics, 2018.
[5] W. Karush, "Minima of functions of several variables with inequalities as side constraints", Master's thesis, Department of Mathematics, University of Chicago, 1939.
[6] H. W. Kuhn, and A. W. Tucker, "Nonlinear programming", Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probabilistics, Berkeley, University of California Press, pp. 481-492, 1951.
[7] A. M. Legendre, "Nouvelles methodes pour la determination des orbites des cometes (New methods for determining the orbits of comets)", Firmin Didot, Paris, 1805.
[8] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingesse, F. Leich, C. C. Chang, and C. C. Lin, "e1071: Misc Functions of Department of Statistics, Probability Theory Group (Formerly: E1071)", TU Wein, R package version 1.7-0, 2018.
[9] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2019.
[10] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, and D. Firth, "MASS: Support Functions and Datasets for Venables and Ripley's MASS", R package version 7.3-51.4, http://www.stats.ox.ac.uk/pub/MASS4/, 2019.
[11] P. J. Rousseeuw, and V. J. Yohai, "Robust Regression by Means of S Estimators in Robust and Nonlinear Time Series Analysis", Lecture Notes in Statistics, vol.26, pp. 256-274, 1984.
[12] A. J. Smola, and B. Scholkopf, "A tutorial on support vector regression", Statistics and Computing, vol.14, pp. 199-222, 2004.
[13] V. N. Vapnik, "The Nature of Statistical Learning Theory", Springer, New York, 1995.
[14] V. N. Vapnik, and A. Y. Chervonenkis, "A note on one class of perceptrons", Automation and Remote Control, vol.25, pp. 821-837, 1964.
[15] V. J. Yohai, "High breakdown point and high efficiency robust estimates for regression", Annals of Statistics, vol.15, pp. 642-656, 1987.