STM JOURNALS

Utilization of Statistical Learning Algorithms for Prediction of Elastic Modulus of Jointed Rock Mass

Prof. Pijush Samui
Centre for Disaster Mitigation and Management, VIT University, Vellore-632014, India

ABSTRACT

This study uses two statistical learning algorithms for the prediction of elastic modulus (E_j) of jointed rock mass. The first algorithm uses support vector machine (SVM) that is firmly based on the theory of statistical learning and uses regression technique by introducing ε-insensitive loss functionhas been adopted. The second algorithm uses relevance vector machine (RVM). It is based on a Bayesian formulation of a linear model with an appropriate prior that results in a sparse representation. The RVM model gives variance of predicted data. The inputs of models are joint frequency (J_n), joint inclination parameter (n), joint roughness parameter (r), confining pressure (σ_3) and elastic modulus (E_i) of intact rock. Equations have been developed for the determination of E_j of jointed rock mass based on the SVM and RVM models. The results of SVM and RVM models are compared with a widely used artificial neural network (ANN) model. This study shows that the developed SVM and RVM models can be used for the prediction of E_j of jointed rock mass.

Keywords: elastic modulus, jointed rock, support vector machine, relevance vector machine, artificial neural network

Author for Correspondence Email: pijush.phd@gmail.com, Phone:91-416-2202283, Fax:91-416-2243092

INTRODUCTION

The presence of joint is common in the rock mass. The behavior of rock mass is largely influenced by the joint. So, the prediction of elastic modulus (E_j) of jointed rock mass is an imperative task in rock engineering. Geotechnical engineers use different empirical relations for determination of E_j of jointed rock mass [1-2]. These empirical relations constrain the data along a particular geometry, which may not always be favorable to capture the non-linear relations existing between various parameters [3]. Maji and Sitharam (2008)[3] successfully used artificial neural network (ANN) to overcome the limitations of the empirical relations. But, ANN model has some limitations such as black box approach, arriving at local minima, slow convergence speed, low generalization capability, overtraining problem and absence of probabilistic output [4–5]. As a result, alternative methods are needed, which provide more accurate model for prediction of E_j of jointed rock mass.

This paper uses two statistical learning algorithms for the prediction of E_j of jointed rock mass. The first statistical learning algorithm adopts support vector machine (SVM), which is a novel type of learning machine based on statistical learning theory

[6]. It provides a new, efficient and novel approach to improve the generalization performance and can attain a global minimum. In general, SVM has been used for pattern recognition problems. But recently, it has been

used to solve non-linear regression estimation and time series prediction by introducing ε-insensitive loss function [6,7–9]. The details of SVM and its application to the geotechnical engineering problems can be found in literature [10–18].

The second statistical learning algorithm uses relevance vector machine (RVM). RVM introduced by Tipping (2000)[19] produces sparse solutions using an improper hierarchical prior and optimizing over hyperparameters. RVM is referred to as Bayesian kernel method that chooses sparse basis sets using an "Automatic Relevance Determination" [20] style prior that pushes non-essential weights to zero. The paper has the following aims:

- To examine the capability of SVM and RVM models for the prediction of E_j of jointed rock mass
- To determine the variance of predicted output based on the developed RVM model
- To develop equations for the determination of E_j of jointed rock mass based on the developed SVM and RVM models
- To make a comparative study between the SVM, RVM and ANN models developed by Maji and Sithatam (2008)[3]

DETAILS OF SVM

An interesting property of SVM approach is that it is an approximate implementation of the structural risk minimization (SRM) induction principle which tells that the generalization ability of learning machines depends more on capacity concept than merely the dimensionality of the space or the number of free parameters of the loss function. This study uses the SVM as a regression technique by introducing a ε-insensitive loss function. In this section, a brief introduction on how to construct SVM for regression problem is presented. More details can be found elsewhere [21–23,10]. The ε-insensitive loss function can be described in the following way (Figure 1):

$$L_\varepsilon(y) = 0 \quad \text{for} \quad |f(x) - y| < \varepsilon \quad \text{otherwise}$$

$$L_\varepsilon(y) = |f(x) - y| - \varepsilon \tag{1}$$

Consider the problem of approximating a set of data,

$$D = \{(x_1, y_1), \ldots, (x_l, y_l)\}, \quad x \in R^N, y \in r \tag{2}$$

where x is the input, y is the output, R^N is the N-dimensional vector space and r is the one dimensional vector space. In this study, input parameters are joint frequency (J_n), joint inclination parameter (n), joint roughness parameter (r), confining pressure (σ_3) and elastic modulus (E_i) of intact rock. The output of SVM is E_j of jointed rock mass. So, in this study, $x = [E_i, J_n, r, n, \sigma_3]$ and $y = E_j$. The main aim in SVM is to find a function that gives a deviation of ε from the actual output and at the same time is as flat as possible. Let us assume a linear function:

$$f(x) = (w.x) + b \; w \in R^N, \quad b \in r \tag{3}$$

where w = an adjustable weight vector and b = the scalar threshold. Flatness in the case of (3) means that one seeks a small w. One way of obtaining this is by minimizing the Euclidean norm $\|w\|^2$. This is equivalent to the following convex optimization problem:

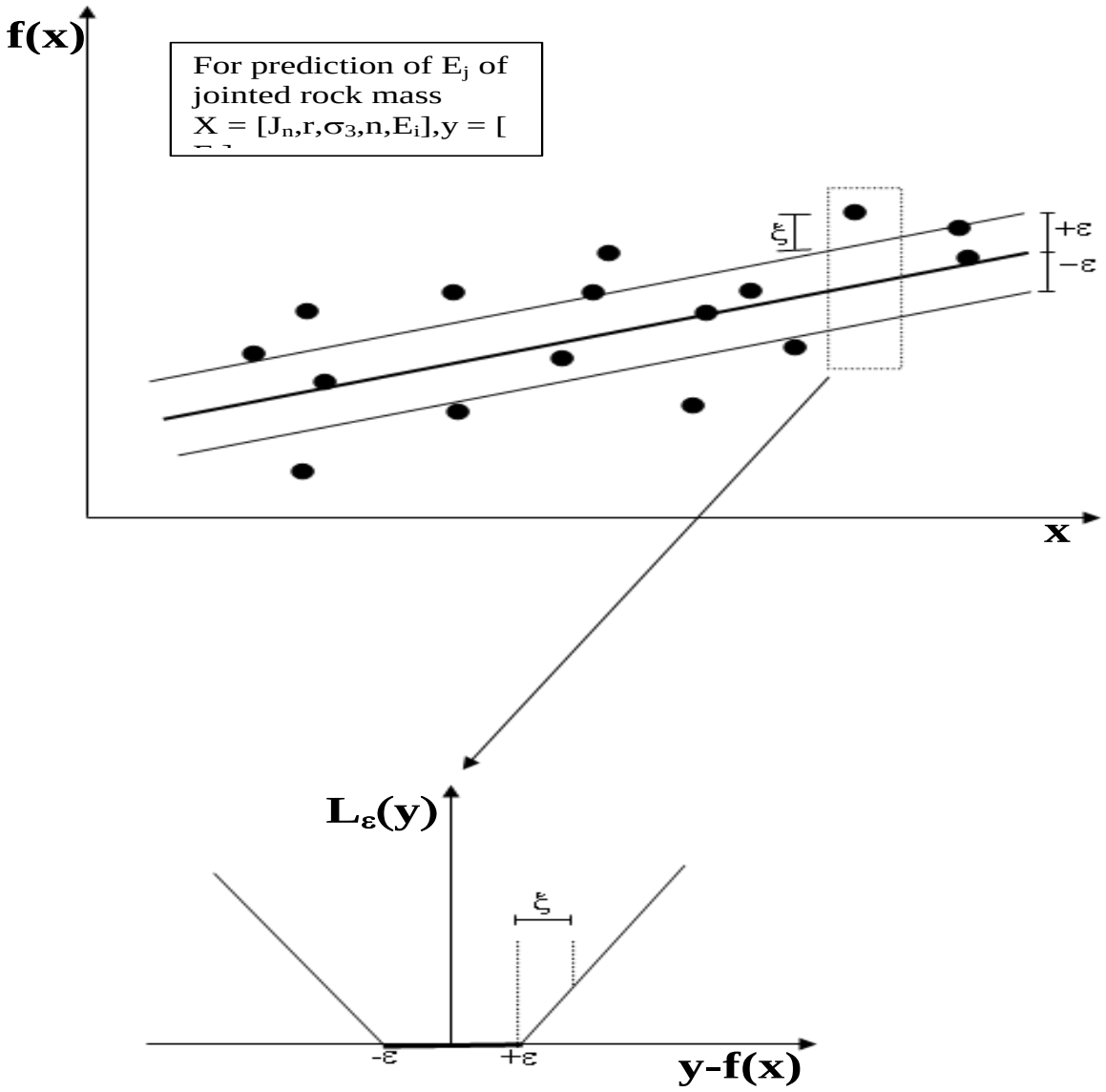


Fig. 1 Prespecified Accuracy ε and Slack Variable ξ_i and ξ_i^* in Support Vector Regression [24].

Minimize: $\frac{1}{2}\|w\|^2$

Subjected to: $y_i - \left(\left\langle w.x_i \right\rangle + b\right) \le \varepsilon$, i = 1, 2,...,l

$$\left(\left\langle w.x_i \right\rangle + b\right) - y_i \le \varepsilon, i = 1,2,...,l \quad (4)$$

In order to allow for some errors, the slack variables ξ_i and ξ_i^* (see Figure 1) are introduced in (4). The formulation can then be restated as:

Minimize: $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)$

Subjected to: $y_i - \left(\left\langle w.x_i \right\rangle + b\right) \le \varepsilon + \xi_i$, i = 1, 2,...,l

$\left(\left\langle w.x_i \right\rangle + b\right) - y_i \le \varepsilon + \xi_i^*$, I = 1, 2,...,l

$\xi_i \ge 0$ and $\xi_i^* \ge 0$, i = 1, 2,...,l (5)

The constant $0 < C < \infty$ determines the trade-off between the flatness of f and the amount up to which deviations larger than ε are tolerated [25]. In practice, the C value is selected by trial and error approach. This optimization problem (4) is solved by Lagrangian multipliers [10], and its solution is given by

$$f(x) = \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(x_i.x\right) + b \quad (6)$$

Where $b = -\left(\frac{1}{2}\right)w.\left[x_r + x_s\right]$ and α_i, α_i^* are the Lagrangian multipliers. An important aspect is that some Lagrangian multipliers (α_i, α_i^*) will be zero, implying that these training objects are considered to be irrelevant for the final solution (sparseness). The training objects with nonzero Lagrangian multipliers are called support vectors.

When linear regression is not appropriate, then input data has to be mapped into a high dimensional feature space through some nonlinear mapping [21] (see Figure 2). The two steps that are involved are first to make a fixed nonlinear mapping of the data onto the feature space and then carry out a linear regression in the high dimensional space. The input data is mapped onto the feature space by a map Φ(see Figure 2). The dot product given by $\Phi(x_i).\Phi(x_j)$ is computed as a linear combination of the training points. The concept of kernel function $[K(x_i, x_j) = \Phi(x_i).\Phi(x_j)]$ has been introduced to reduce the computational demand [22, 26]. So, equation (6) becomes written as

$$f(x) = \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)K\left(x_i..x_j\right) + b \quad (7)$$

Some common kernels have been used such as polynomial (homogeneous), polynomial (nonhomogeneous), radial basis function, Gaussian function, sigmoid,

This study uses the above SVM based model for predicting E_j of jointed rock mass. The data has been collected from the work of Maji and

Sitharam (2008) [3]. The complete database comprised of 896 datasets. Out of that, 515 datasets are with confining case and rest with unconfined case. The dataset contains information about J_n,n,r, σ_3,E_i, and E_j.

The data has been divided into two sub-sets; a training dataset, to construct the model, and a testing dataset to estimate the model performance. So, for our study a set of 726 data is considered as the training dataset and remaining set of 170 data is considered as the testing dataset. The data is scaled between 0 and 1. This study uses radial basis function ($K(x_i, x) = \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2\sigma^2}\right\}$) as a kernel function. When applying SVM, the optimum values of the C and width (σ) of radial basis function will be determined during the modeling experiment. etc., for non-linear cases. Figure 3 shows a typical architecture of the SVM for E_j of jointed rock mass.

DETAILS OF RVM

The RVM, introduced by Tipping (2000)[19], is a sparse linear model. Let $D = \{(x_i, t_i), i = 1,..., N\}$ be a dataset of observed values. Where x_i = input, t_i = output, $x_i \in R^d$ and $t_i \in R$. In this study, the input parameters are n, J_n, r, σ_3 and E_i. So, $x = [n, J_n, r, \sigma_3, E_i]$. The output of the RVM model is E_j of jointed rock mass. Therefore, $t = [E_j]$. One can express the output as the sum of an approximation vector $y = (y(x_1),..., y(x_N))^T$, and zero mean random error (noise) vector $\varepsilon = (\varepsilon_1,..., \varepsilon_N)^T$ where $\varepsilon_n \sim \mathbf{N}(0, \sigma^2)$ and $\mathbf{N}(0, \sigma^2)$ is the normal distribution with mean 0 and variance σ^2. So, the output can be written as:

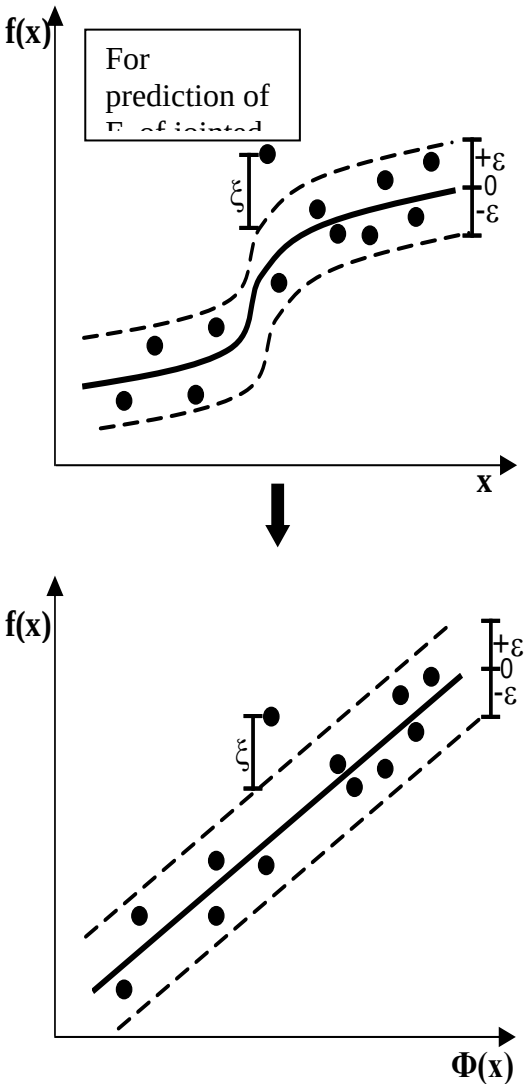


Fig. 2 Concept of Nonlinear Regression.

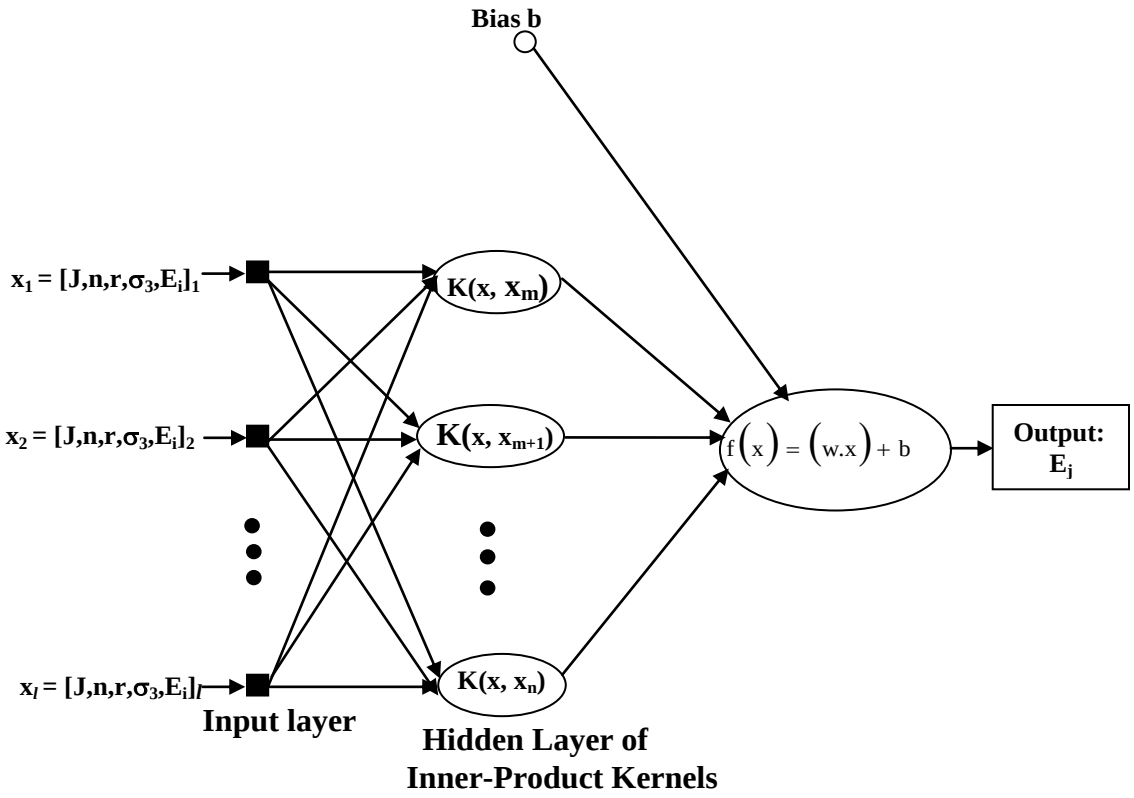


Fig. 3 SVM Architecture for Prediction of E_j of Jointed Rock Mass.

$$t_n = y(x_n, \omega) + \varepsilon_n \tag{8}$$

where, ωis the parameter vector. Letus assume

$$p(t_n \mid x) \sim \mathbf{N}(y(x_n), \sigma^2) \tag{9}$$

where $\mathbf{N}(y(x_n), \sigma^2)$ is the normal distribution with mean $y(x_n)$ and variance $\sigma^2 y(x)$ can be expressed as a linearly weighted sum of M nonlinear fixed basis function,

$\{\Phi_j(x) \mid j = 1,..., M\}$:

$$y(x; \omega) = \sum_{i=1}^{M} \omega_i \Phi_i(x) = \mathbf{\Phi}\omega \tag{10}$$

The likelihood of the complete data set can be written as

$$p(t \mid w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\|t - \Phi w\|^2\right\} \tag{11}$$

where $t = (t_1..., t_N)^T$, $\omega = (\omega_0,..., \omega_N)$ and

$$\Phi^T = \begin{bmatrix} 1 & K(x_1, x_1) & K(x_1, x_2) & \cdots & K(x_1, x_n) \\ 1 & K(x_1, x_2) & K(x_2, x_2) & \cdots & K(x_2, x_n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & K(x_n, x_1) & K(x_n, x_2) & \cdots & K(x_n, x_n) \end{bmatrix}$$

where $k(x_i, x_n)$ is a kernel function.

To prevent overfitting, automatic relevance detection (ARD) prior is set over the weights.

$$p(w \mid \alpha) = \prod_{i=0}^{N} N(\omega_i \mid 0, \alpha_i^{-1}) \tag{12}$$

Where α is a hyperparametervector that controls how far from zero each weight is

allowed to deviate [25]. Consequently, using Baye's rule, the posterior over all unknowns could be computed given the defined noninformative prior distribution:

$$p\left(w,\alpha,\sigma^2/t\right)=\frac{p\left(y/w,\alpha,\sigma^2\right)p(w,\alpha,\sigma)}{\int p\left(t/w,\alpha,\sigma^2\right)p\left(w,\alpha,\sigma^2\right)dw\,d\alpha\,d\sigma^2} \quad (13)$$

Full analytical solution of this integral [19] is obdurate. Thus decomposition of the posterior according to $p\left(w,\alpha,\sigma^2/t\right)=p\left(w/t,\alpha,\sigma^2\right)p\left(\alpha,\sigma^2/t\right)$ is used to facilitate the solution [27]. The posterior distribution over the weights is thus given by:

$$p\left(w/t,\alpha,\sigma^2\right)=\frac{p\left(t/w,\sigma^2\right)p(w/\alpha)}{p\left(t/\alpha,\sigma^2\right)} \quad (14)$$

The resulting posterior distribution over the weights is the multi-variate Gaussian distribution

$$p\left(w/t,\alpha,\sigma^2\right)=\mathbf{N}(\mu,\Sigma) \quad (15)$$

where the mean and the covariance are respectively given by:

$$\Sigma=\left(\sigma^{-2}\Phi^{T}\Phi+A\right)^{-1} \quad (16)$$

$$\mu=\sigma^{-2}\Sigma\Phi^{T}t \quad (17)$$

With diagonal $A=\mathrm{diag}(\alpha_0,...,\alpha_N)$.

For uniform hyperpriors over and one needs only maximize the term $p\left(t/\alpha,\sigma^2\right)$:

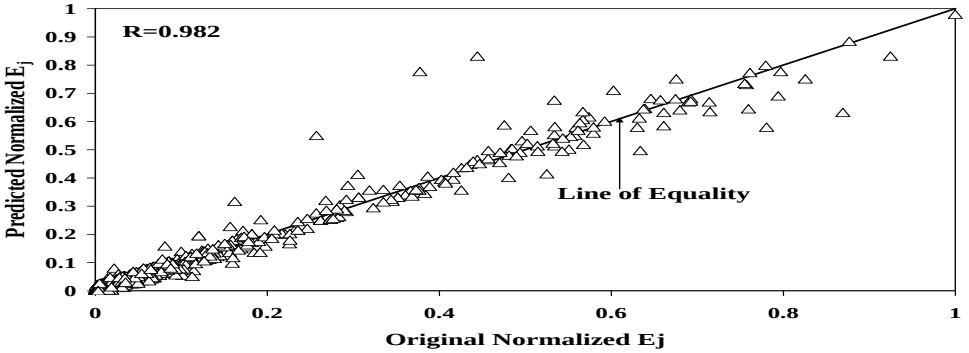
$$p\left(t/\alpha,\sigma^2\right)=\int p\left(t/w,\sigma^2\right)p(w/\alpha)dw$$

$$=\left(\frac{(2\pi)^{\frac{-N}{2}}}{\sqrt{\left|\sigma^2+\Phi A^{-1}\Phi T\right|}}\right)\times\exp\left\{-\frac{1}{2}y^{T}\left(\sigma^2+\Phi A^{-1}\Phi T\right)^{-1}y\right\} \quad (18)$$

Maximization of this quantity is known as the type II maximum likelihood method [27,28] or the "evidence for hyper parameter"(MacKay 1992). Hyper parameter estimation is carried out in iterative formulae, e.g., gradient descent on the objective function [22]. The outcome of this optimization is that many elements of α go to infinity such that w will have only a few nonzero weights that will be considered as relevant vectors.

The main scope of this study is the use of above RVM-based model for prediction of E_j of jointed rock mass. In RVM, the same training, testing, normalization technique and kernel function have been used as used in the SVM model. The SVM and RVM models have been constructed by using MATLAB.

RESULTS AND DISCUSSION



The performance of the SVM and RVM models has been examined by the value of coefficient of correlation(R). The design values of C, ε and σ is 100, 0.02 and 0.7 respectively. The number of support vector is 201. Figure 4 shows the performance of SVM model for training dataset. Figure 4 also shows the value R is close to one. In order to evaluate the prediction capabilities

Fig. 4 Performance of SVM for Training Dataset.

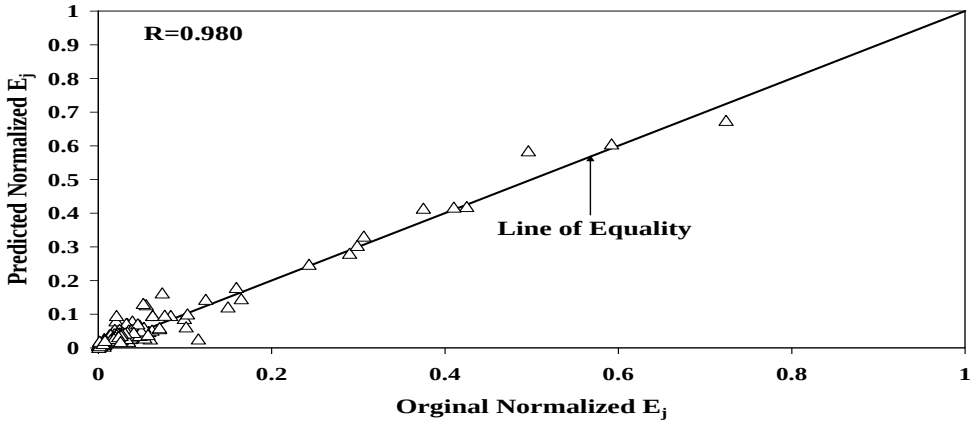


Fig. 5 Performance of SVM for Testing Dataset.

of the SVM model, the model is validated with the test set data. The performance of SVM for testing dataset has been shown in Figure 5. Figure 5 confirms that the value of R is close to one. So, the developed SVM model has the ability to predict E_j of jointed rock mass. The following equation (by putting

$$K(x_i, x) = \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{0.98}\right\},$$

b=0, l=726 and σ=0.7 in Eqn (7)) can be developed for the prediction of Ej of jointed rock mass based on the developed SVM model.

$$E_j = \sum_{i=1}^{726}\left(\alpha_i - \alpha_i^*\right)\exp\left\{-\frac{(x_i - x)(x_i - x)^T}{0.98}\right\} \quad (19)$$

Figure 6 shows the value of $(\alpha_i - \alpha_i^*)$.

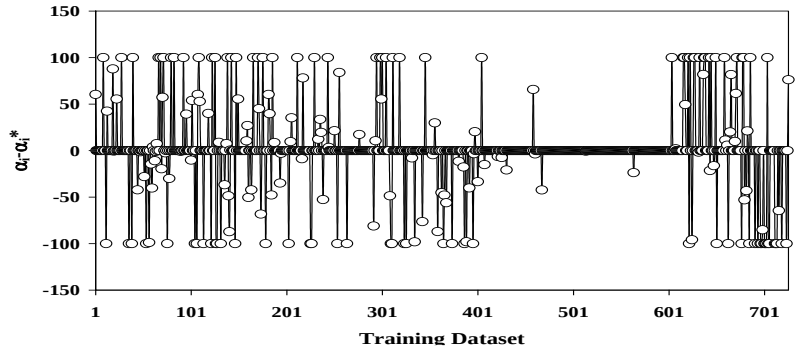


Fig. 6 Values of $(\alpha_i - \alpha_i^*)$ for SVM Model.

For RVM model, the design value of σ has been determined by trial and error approach. The design value of σ is 0.02 and the number of relevance vector is 428. Figure 7 depicts the performance of RVM model for testing dataset.

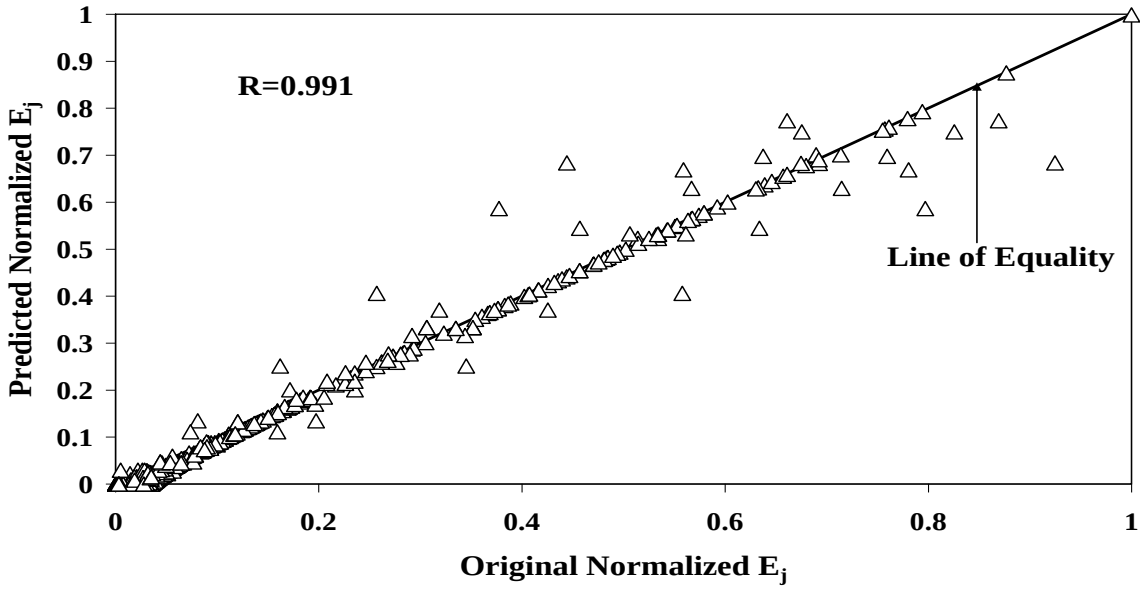


Fig. 7 Performance of RVM Model for Training Dataset.

According to the results of RVM training, RVM has successfully captured the relationship between the input parameters and output. Now, the performance of RVM model has been determined for testing dataset. The performance of testing dataset has been shown in Figure 8. Figure 8 also confirms that the developed RVM model can be used for determination of E_j of jointed rock mass. The following equation is developed for the determination of E_j of jointed rock mass

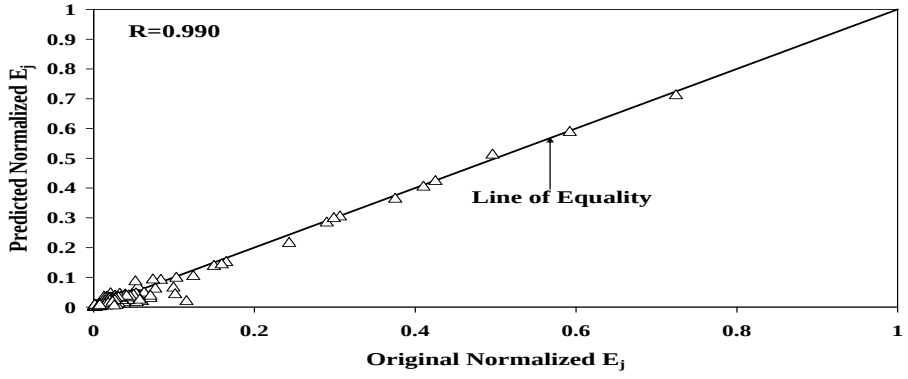


Fig. 8 Performance of RVM for Testing Dataset.

$$E_j = \sum_{i=1}^{726} w_i \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{0.0008}\right\} \quad (20)$$

Figure 9 shows the value of w.

The developed RVM model also gives variance of predicted data. Figures 10 and 11 depict the variance of training and testing dataset respectively. The predicted variance can be used to determine the confidence interval.

A comparative study has been done between developed SVM, RVM, and ANN model developed by Maji and Sitharam(2008)[3]. Maji and Sitharam (2008) [3] successfully used backpropagation (BP) and radial basis function (RBF)-based ANN model for the prediction of E_j of jointed rock mass. Comparison has been done for testing dataset. Figure 12 represents the bar chart of average absolute error (%) for the different models.

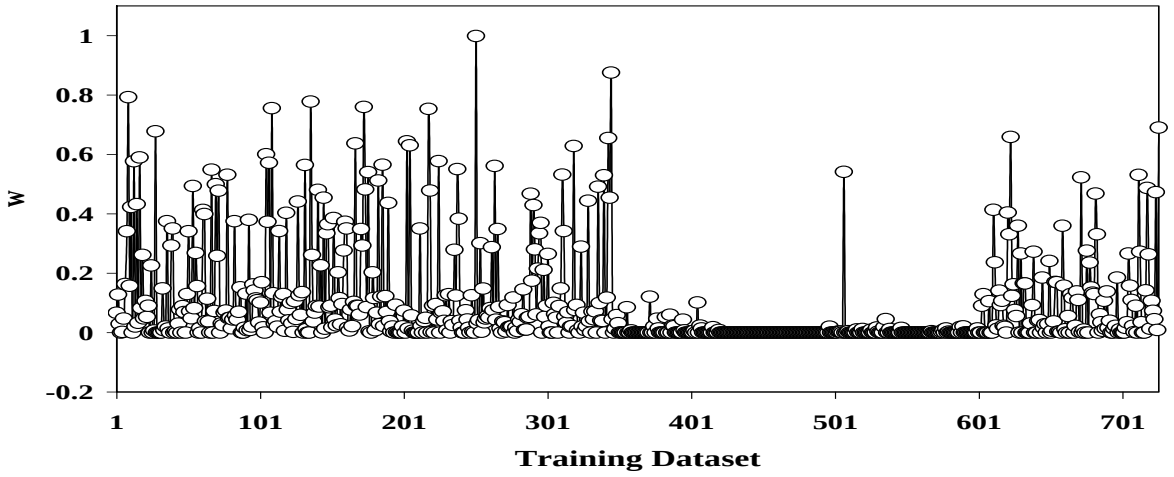


Fig. 9 Values of w for RVM Model.

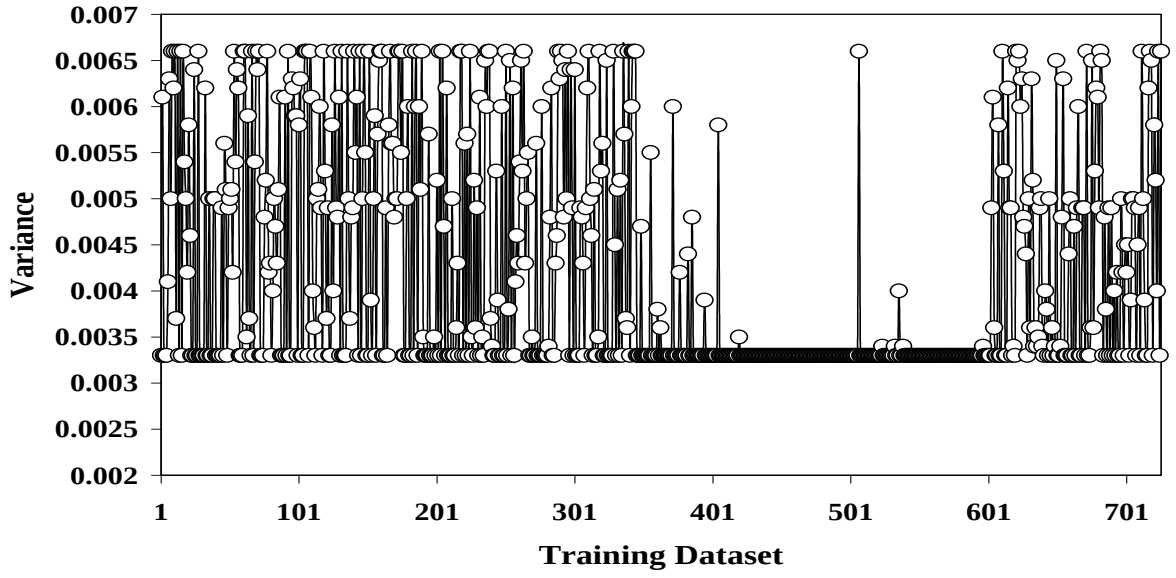


Fig. 10 Variance of Training Dataset for RVM Model.

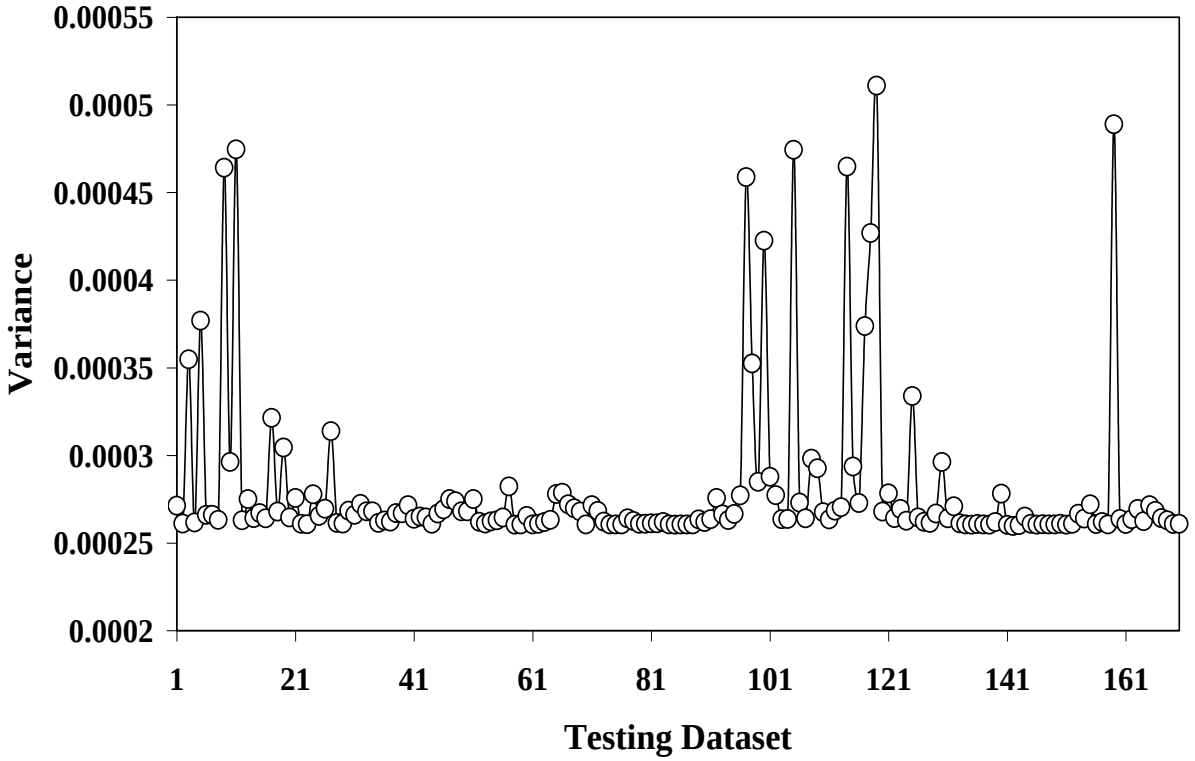


Fig. 11 Variance of Testing Dataset for RVM Model.

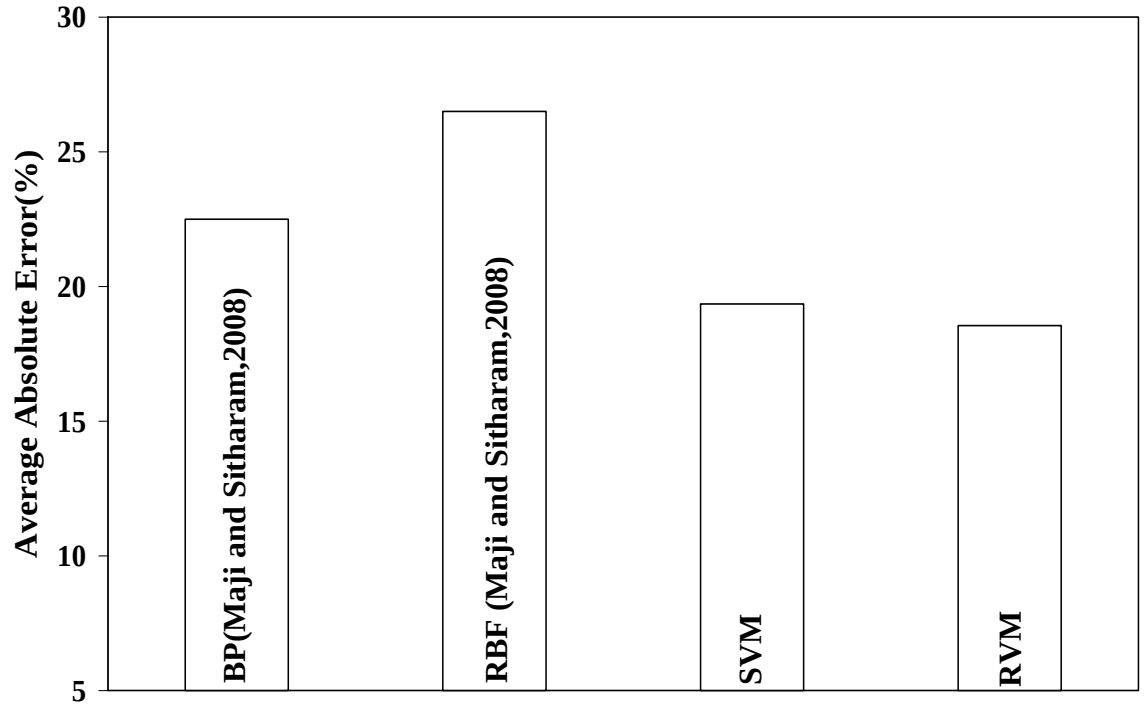


Fig. 12 Comparison between ANN, SVM and RVM Models.

Figure 12 also shows that the developed SVM and RVM models outperform the ANN models. The developed SVM and RVM use three(C, ε, and σ) and one kernel parameter (σ), respectively. In ANN, there are a larger number of controlling parameters, including the number of hidden layers, number of hidden nodes, learning rate, momentum term, number of training epochs, transfer functions, and weight initialization methods.

The performance of RVM model is slightly better than the SVM model. The developed RVM model uses 58.95% of training data as relevance vector. These relevance vectors are only used for final prediction. Therefore, the developed RVM produces sparse solution. Sparseness means that a significant number of the weights are zero (or effectively zero), which has the consequence of producing compact, computationally efficient models, which in addition are simple and therefore produce smooth functions.

The developed SVM model uses 27.68% of training data as support vector. These support vectors have been used for final prediction. The developed SVM model produces more sparse solution than RVM model. The performance of SVM and RVM is almost same for the training and testing dataset. So, the developed SVM and RVM do not exhibit any overtraining. Therefore, the developed SVM and RVM have good generalization capability.

CONCLUSIONS

This paper describes the twostatistical learning algorithms (SVM and RVM) for prediction of E_j of jointed rock mass. Both algorithms give promising results. The developed SVM and RVM model outperforms the ANN models. In terms of prediction accuracy, the performance of RVM model is slightly better than SVM model. The developed RVM model has the added advantage of probabilistic interpretation that yields prediction uncertainty. In terms of sparseness, the performance of SVM is better than the RVM model. User can use the developed equations for prediction of E_j of jointed rock mass. The developed SVM and RVM models can be used as quick tools for prediction of E_j of jointed rock mass. In summary, this study has given two robust models based on the SVM and RVM for the prediction of E_j of jointed rock mass.

ACKNOWLEDGEMENT

The author thanks T.G. Sitharam and Vidya Bhushan Maji for providing the data.

REFERENCES

1.Ramamurthy, T. and Arora, V.K. *International Journal of Rock Mechanics and Mining Science &Geomechanics Abstracts*, (1994).13(1). 9–22p.

2.Sitharam, T.G., Sridevi, J. and Shimizu, N. *International Journal of Rock Mechanics*

and Mining Science &Geomechanics Abstracts. 2001.38,.437–448p.

3.Maji, V.B. and Sitharam, T.G. *Geotechnical and Geoenvironmental Engineering*, 2008.26. 443–452p.

4.Park, D. and Rilett, L.R. *Computer Aided Civil and Infrastructure Engineering*. 1999.14. 358–367p.

5.Kecman, V. *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. The MIT press, Cambridge, Massachusetts, London, England.2001.

6.Vapnik, V.N.(1995), *The nature of statistical learning theory*, Springer, New York.

7.Vapnik, V.N., Golowich, S. and Smola, A.*Advance in Neural Information Processing System* 9. Mozer M. and Petsch T. (Eds.). MIT Press. Cambridge, MA.1997.

8.Mukherjee, S., Osuna, E. and Girosi, F.*Proceedings of IEEEWorkshop on Neural Networks for Signal Processing* 7, Institute of Electrical and Electronics Engineers, New York. .1999. 511–519p.

9.Muller, K., Smola, A., Ratsch, G., et al. *Proceedings of International Conference on Artificial Neural Networks*, Springer-Verlag, Berlin 999.1999.

10.Vapnik, V.N. *Statistical Learning Theory*. Wiley, New York.1998.

11.Smola, A.J. and Scholkopf, B. *Statistics and Computing*. 2004.14. 199–222p.

12.Pal, M.*International Journal for Numerical and Analytical Methods in Geomechanics*. 2006. 30(10), 983–996p.

13. Goh, A.T.C. and Goh, S.H. *Computers and Geotechnics*. 2007.34(5). 410–421p.

14.Samui, P. *Computers and Goetechnics*. 2008.35(3), 419–427p.

15.Samui, P., Kurup, P. and Sitharam, T.G. *Journal of Geotechnical and GeoEnvironmental engineering*. 2008.134(6). 894–898.

16.Zhao, H.B. and Yin, S. *Marine Georesources & Geotechnology*. 2010. 28(1). 64–75p.

17.Zhao, H.B. *Computers and Geotechnics*. 2008.35(3). 459–467p.

18.Pal, M. and Deswal, S. *Computer and Geotechnics*(in press).2010).

19.Tipping, M.E. *Advances in Neural Information Processing Systems*.2000. 12. 625–658p.

20.Bishop, C. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.1995.

21.Boser, B.E., Guyon, I.M. and Vapnik, V.N. In D. Haussler (Ed.).*5th Annual ACM Workshop on COLT*. Pittsburgh, PA, ACM Press. 1992. 144–152p.

22.Cortes, C. and Vapnik, V.N.*Machine Learning*..1995. 20. 273–297.

23.Gualtieri, J.A., Chettri, S.R., Cromp, R.F. et al.*In the Summaries of the Eighth JPL Airbrone Earth Science Workshop*.1999.

24.Scholkopf, B. (1997), "Support vector learning," R. Oldenbourg, Munich.

25.Scholkopf, B. and Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA.2002.

26.Cristianini, N. and Shawe-Taylor, J. *An introduction to Support vector machine*, University press, London, Cambridge.2000.

27.Tipping, M.E. *Journal of Machine Learning Research*. 2001.1.211–244.

28.Berger, J.O. *Statistical Decision Theory and Bayesian Analysis.* 2nd ed. Springer, New York. 1985.

29.Wahba, G. *Annals ofStatistics* 1985.4. 1378–1402p.

30.MacKay, D.J. *Bayesian Methods for Adaptive Models*. Ph.D. thesis.Department of Computer Science and Computation &Neural Systems, California Institute of Technology, Pasadena, CA. 1992.

