# *Application of Support Vector Machine for Rock Slope Stability Analysis*

***Pijush Samui***
*Center for Disaster Mitigation and Management*
*Vellore Institute of Technology,*
*Vellore-632014, India*
*Tel: 91-416-2202283, Fax: 91-416-2243092*
*Email: pijush.phd@gmail.com*

**ABSTRACT**

This paper examines the capability of Support Vector Machine (SVM) based classification approach for rock slope stability analysis. SVM achieves good generalization ability by adopting a structural risk minimization (SRM) induction principle that aims at minimizing a bound on the generalization error of a model rather than the minimizing the error on the training data only. This study uses SVM as a classification tool. The inputs of SVM model are unit weight ($\gamma$), cohesions ($c_A$) and ($c_B$), angles of internal friction ($\phi_A$) and $\phi_B$, angle of the line of intersection of the two joint-sets ($\psi_p$), slope angle ($\psi_f$) ,height (H), where A and B refer to the two joint sets. An equation has been developed for the prediction of stability of rock slope based on SVM model. This study shows that SVM is a powerful model for the determination of stability of rock slope.

***Keywords:*** Rock slope; Stability; Support vector machine; Prediction

## 1. INTRODUCTION

The stability of rock slope plays an important role when designing dams, roads, tunnels and other engineering structures. So, the prediction of stability of rock slope is an imperative task in rock engineering. The determination of stability of rock slope is a difficult task due to inhomogeneous and discontinuous (joints, fractures and bedding planes) nature of rock mass. Practicing engineers use different methods for the prediction of stability of rock slope (Jaeger, 1971; Hoek and Bray, 1981; Zanbak, 1983; Goodman and Kieffer, 2000; Siad, 2003; Hack et al., 2003; Yarahmadi and Verdel, 2005). However, most of the available methods simplify the problem by incorporating several assumptions associated with the factors that affect stability of rock slope. Therefore, the available methods are nor reliable. As a result, alternative methods are needed, which provide more accurate prediction of stability of slope.

This study examines the potential of Support Vector Machine (SVM) for prediction of stability of rock slope. SVM is based on statistical learning theory which has been developed by Vapnik (1995). SVM adopts structural risk minimization (SRM) principle. SRM seeks to minimize the upper bound of the generalization error rather than minimize the training error. The details of SVM and its application to geotechnical engineering problems can be found in literatures (Vapnik, 1998; Goh and Goh, 2007; Samui, 2008; Samui et al., 2008; Smola and Schilopf, 2004). This study uses the database collected by Sakellariou and Ferentinou (2005). The dataset contains the information about unit weight ($\gamma$), cohesions ($c_A$) and ($c_B$), angles of internal friction ($\phi_A$) and $\phi_B$, angle of the line of intersection of the two joint-sets ($\psi_p$), slope angle ($\psi_f$) ,height (H), where A and B refer to the two joint set and status of slope i.e. stable or failed. The paper has the following aims:

- To investigate the feasibility of SVM for prediction of stability of rock slope
- To determine an equation for prediction of status of rock slope based on SVM model

## 2. DETAILS OF SVM

SVM has recently emerged as an elegant pattern recognition tool and a better alternative to ANN methods. The method has been developed by Vapnik (1995) and is gaining popularity due to many attractive features. This section of the paper serves an introduction to this relatively new technique. Details of this method can be found in Boser et al. (1992), Cortes and Vapnik (1995), Gualtieri et al. (1999) and Vapnik (1998). A binary classification problem is considered having a set of training vectors (D) belonging to two separate classes.

$$D = \{(x^1, y^1)..........,(x^l, y^l)\} \qquad x \in R^n, y \in \{-1,+1\} \qquad (1)$$

Where $x \in R^n$ is an n-dimensional data vector with each sample belonging to either of two classes labelled as $y \in \{-1,+1\}$, and l is the number of training data. The main aim is to find a generalized classifier that can distinguish the two classes (-1, +1) from the set of the training vectors mentioned above (D) and also can classify equally well the unseen data. In the current context of classifying slope failure, the two classes labelled as (-1, +1) may mean failed rock slope and stable rock slope. In this study, $\gamma$, $c_A$,$c_B$ $\phi_A$, $\phi_B$, $\Psi_p$,$\Psi_f$ and H are used as input parameters. So, $x = [\gamma, H, c_A, c_B, \phi_A, \phi_B, \psi_p, \psi_f]$. For a set of data, this would mean a linear hyper plane defined by Eq. 2 which can distinguish the two classes.

$$f(x) = w.x + b = 0 \qquad (2)$$

Where $w \in R^n$ determines the orientation of a discriminating hyperplane, $b \in R$ is a bias. An example of hyperplane is shown in Fig. 1. For the linearly separable case, a separating hyperplane can be defined for the two classes as

$$w.x_i + b \geq 1 \text{ (for } y_i = 1) \longrightarrow \text{stable slope}$$

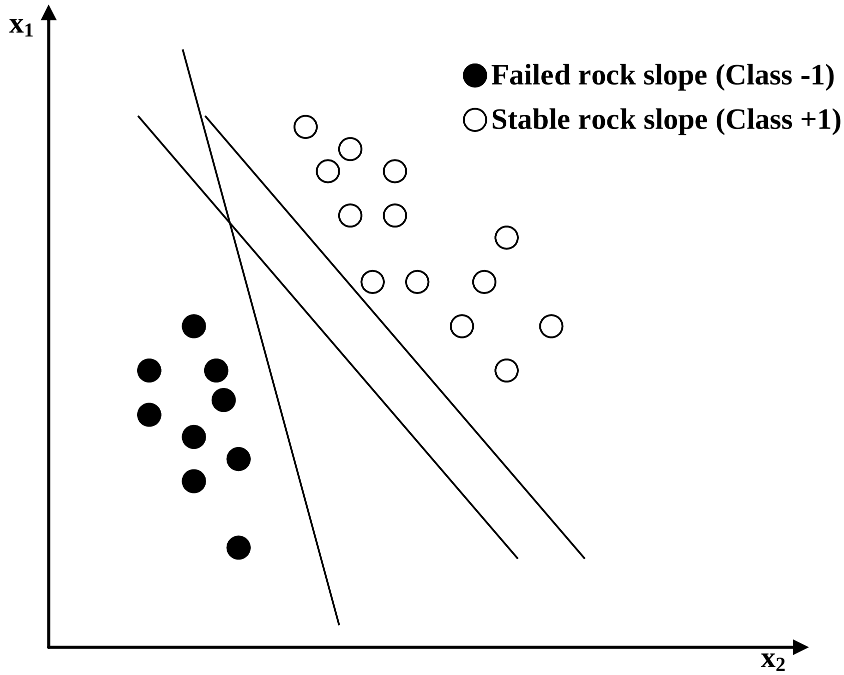


Fig. 1 - An example of hyperplane

$$w.x_i + b \leq -1 \text{ ( for } y_i = -1) \longrightarrow \text{failed slope} \quad (3)$$

The above two equations can be combined as

$$y_i \left(w.x_i + b\right) \geq 1 \quad (4)$$

Sometimes, due to the noise or mixture of classes introduced during the selection of training data, variables $\xi_i > 0$, called slack variables, are used due to the effects of misclassification. So the Eq. 4 can be written as

$$y_i \left(w.x_i + b\right) \geq 1 - \xi_i \quad (5)$$

The perpendicular distance from the origin to the plane $w.x_i + b = -1$ is $\frac{|1+b|}{\|w\|}$. Similarly, the perpendicular distance from the origin to the plane $w.x_i + b = 1$ is $\frac{|b-1|}{\|w\|}$. The margin ($\rho(w,b)$) between the planes is simply

$$\rho(w,b) = \frac{2}{\|w\|} \quad (6)$$

The optimal hyperplane is located where the margin between two classes of interest is maximized (Fig. 2) and the error is minimized. The maximization of this margin leads to the following constrained optimization problem

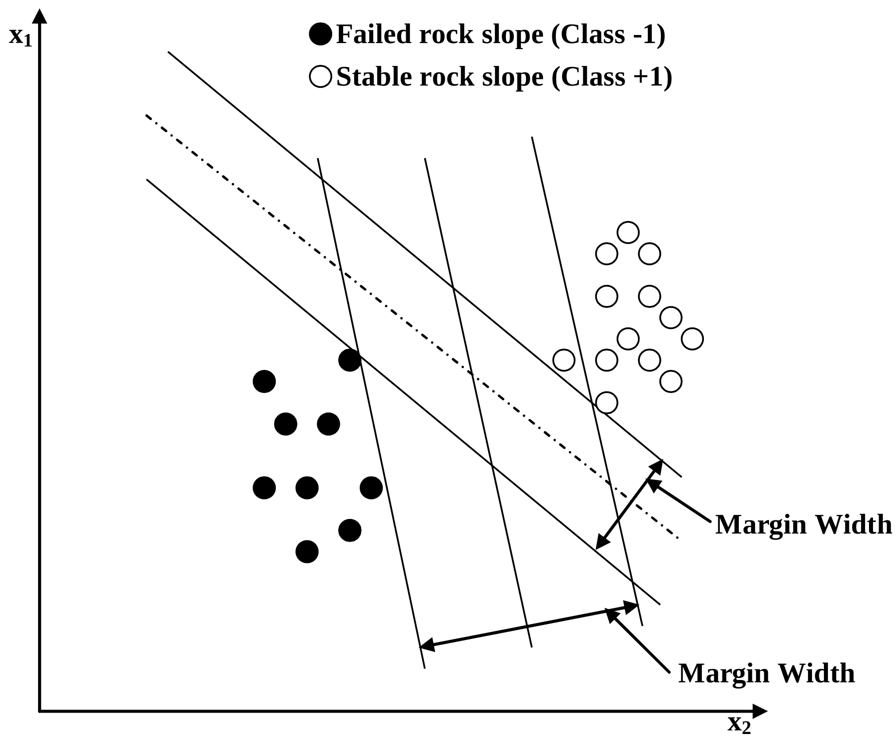


Fig. 2 - Margin width of different hyperplanes

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$

$$\text{Subjected to: } \quad y_i\left(w.x_i + b\right) \geq 1 - \xi_i \qquad (7)$$

The constant 0<C<? , a parameter defines the trade-off between the number of misclassification in the training data and the maximization of margin. A large C assigns higher penalties to errors so that the SVM is trained to minimize error with lower generalization while a small C assigns fewer penalties to errors; this allows the minimization of margin with errors, thus higher generalization ability. If C goes to infinitely large, SVM would not allow the occurrence of any error and result in a complex model, whereas when C goes to zero, the result would tolerate a large amount of errors and the model would be less complex.

In order to solve the above optimization problem (Eq. 7), the Lagrangian is constructed as follows:

$$L(w,b,\alpha,\beta,\xi) = \frac{\|w\|^2}{2} + C\left(\sum_{i=1}^{l}\xi_i\right) - \sum_{i=1}^{l}\alpha_i\left\{\left[\left(w.x_i + b\right)\right]y_i - 1 + \xi_i\right\} - \sum_{i=1}^{l}\beta_i\xi_i \qquad (8)$$

Where α, β are the Lagrange multipliers. The solution to the constrained optimization problem is determined by the saddle point of the Lagrangian function L(w,b,α,β,ξ), which has to be minimized with respect to w, b and ξ. Thus, differentiating L(w,b,α,β,ξ) with respect to w, b and ξ and setting the results equal to zero, the following three conditions have been obtained:

Condition 1: $$\frac{\partial L(w,b,\alpha,\beta,\xi)}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{l} \alpha_i y_i x_i$$

Condition 2: $$\frac{\partial L(w,b,\alpha,\beta,\xi)}{\partial b} = 0 \Rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0$$

Condition 3: $$\frac{\partial L(w,b,\alpha,\beta,\xi)}{\partial \xi} = 0 \Rightarrow \alpha_i + \beta_i = C \qquad (9)$$

Hence from Eqs. 8 and 9 the equivalent optimization problem becomes (Osuna et al., 1997),

Maximize: $$\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \left( x_i . x_j \right)$$

Subjected to: $$\sum_{i=1}^{l} \alpha_i y_i = 0 \text{ and } 0?\alpha_i?C, \quad \text{for } i=1, 2,\ldots, l \qquad (10)$$

Solving Eq. 10 with constraints determines the Lagrange multipliers. According to the Karush-Kuhn-Tucker (KKT) optimality condition (Fletcher, 1987), some of the multipliers will be zero. The nonzero multipliers are called support vectors (Fig. 3). In conceptual terms, the support vectors are those data points that lie closest to the optimal hyperplane and are therefore the most difficult to classify. The value of w and b are calculated from $w = \sum_{i=1}^{l} y_i \alpha_i x_i$ and $b = -\frac{1}{2} w \left[ x_{+1} + x_{-1} \right]$ where $x_{+1}$ and $x_{-1}$ are the support vectors of class labels +1(stable slope) and -1(failed slope) respectively. The classifier can then be constructed as:

$$f(x) = \text{sign}(w.x + b) \qquad (11)$$

Where sign (•) is the signum function. It gives +1(stable slope) if the element is greater than or equal to zero and -1(failed slope) if it is less than zero.

In case where linear supporting hyper plane is inappropriate, SVM maps input data into a high dimensional feature space through some nonlinear mapping (Boser et al., 1992) (Fig. 4). This method easily converts a linear classification learning algorithm into a non-linear one, by mapping the original observations into a higher-dimensional non-linear space so that linear classification in the new space is equivalent to non-linear classification in the original space. After replacing x by its mapping in the feature space $(\Phi(x))$, the optimization problem of Eq. 11 becomes

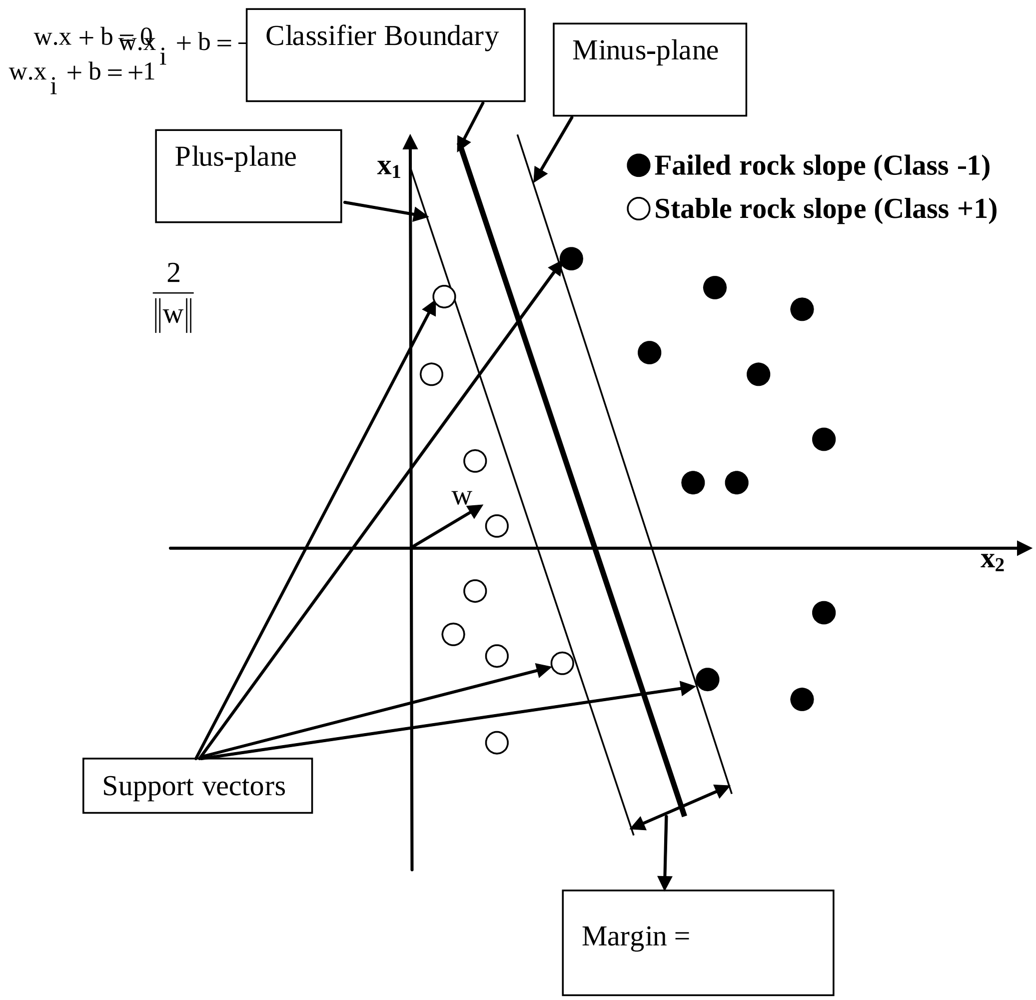


Fig. 3 - Support vectors with maximum margin

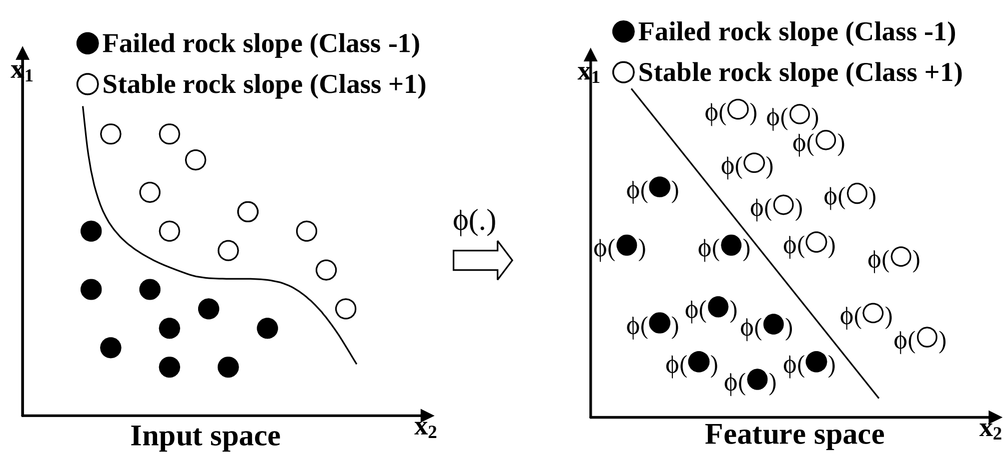


Fig. 4 – Concept of nonlinear SVM for classification problem

Maximize: $\sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j \left(\Phi(x_i).\Phi(x_j)\right)$

Subjected to: $\sum_{i=1}^{l} \alpha_i y_i = 0$ and 0?αi?C, for i=1, 2,…, l (12)

Kernel function $K(x_i.x_j) = \Phi(x_i).\Phi(x_j)$ has been introduced instead of feature space $(\Phi(x))$ to reduce computational demand (Cortes and Vapnik, 1995; Cristianini and Shawe-Taylor, 2000). Polynomial, radial basis functions and certain sigmoid functions has been used as a kernel functions. To get the Eq. 11, same procedures have been applied as in linear case. So, the final classifier takes the following form:

$$f(x) = sign\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i.x) + b\right) \qquad (13)$$

This study uses the above methodology for prediction of stability of rock slope. The data is scaled in between 0 and 1. In carrying out the formulation, the data has been divided into two sub-sets: such as

(a) A training dataset: This is required to construct the model. In this study, 16 out of the 22 data are considered for training dataset.

(b) A testing dataset: This is required to estimate the model performance. In this study, the remaining 6 data is considered as testing dataset. To train the RVM model, radial basis function has been used as kernel function. The program is developed using MATLAB.

## 3. RESULTS AND DISCUSSION

The training or testing performance has been calculated from the following formula:

$$\text{Training or Testing performance}(\%) = \left(\frac{\text{No of data predicted accurately by SVM}}{\text{Total data}}\right) \times 100 \qquad (14)$$

Figure 5 shows that the effect of C on testing performance (%) and number of support vector.

From Fig. 5, it is clear that C value does not affect the testing performance (%) of radial basis function. In case of radial basis function, the number of support vector is decreasing up to C=30 and after that number of support vector remain constant with increasing C value (Fig. 5). The design value of C, width of radial basis function ($\sigma$) and number of support vector is 30, 1 and 10 respectively. The performance of SVM for training dataset is 100% using design value of C and $\sigma$. In order to determine the capability of SVM model, the performance of testing data has been determined by using design C and s values. It has been shown that only one data (out of six data) has been classified. Therefore, the developed SVM model has capability to predict status of rock slope. Tables 1 and 2 show the performance of SVM model for training and testing data respectively. The performance of training and testing is almost same for the developed SVM. So, the developed SVM model has ability to avoid overtraining. Therefore, it has good generalization capability. SVM uses only two parameters (σ and C). 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. Obtaining an optimal combination of these parameters is a difficult task as well. Another major advantage of the SVM is its optimization algorithm, which includes solving a linearly constrained quadratic programming function leading to a unique, optimal, and global solution compared to the ANN. In SVM, the number of support vectors has determined by algorithm rather than by trial-and-error which has been used by ANN for determining the number of hidden nodes.

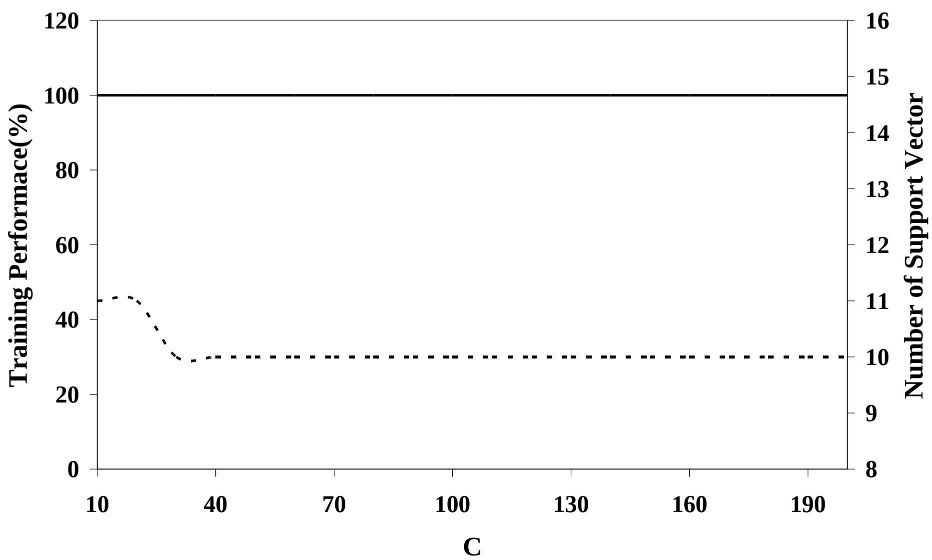


Fig. 5 - Variation of testing performance (%) and number of support vectors with capacity factor (C) values for radial basis function

Table 1 - Performance of training dataset

| γ (kN/m³) | $c_A$ (kPa) | $c_B$ (kPa) | $\phi_A(^0)$ | $\phi_B(^0)$ | $\psi_p(^0)$ | $\psi_f(^0)$ | H(m) | Actual class | Predicted class |
|---|---|---|---|---|---|---|---|---|---|
| 25.14 | 23.94 | 47.88 | 20 | 30 | 31.2 | 65 | 30.5 | 1 | 1 |
| 25 | 14.36 | 16.76 | 28 | 18 | 30 | 45 | 37 | -1 | -1 |
| 22.8 | 0 | 0 | 35 | 35 | 38 | 47 | 110 | -1 | -1 |
| 26 | 0 | 0 | 30.6 | 22.8 | 30.6 | 33 | 270 | 1 | 1 |
| 26 | 20 | 20 | 27 | 27 | 60 | 70 | 44 | 1 | 1 |
| 26 | 0 | 0 | 39 | 39 | 60 | 70 | 44 | -1 | -1 |
| 26.66 | 0 | 0 | 45 | 45 | 35 | 50 | 150 | 1 | 1 |
| 25 | 0 | 0 | 32.4 | 32.4 | 30 | 48 | 50 | 1 | 1 |
| 18.84 | 0 | 0 | 30 | 30 | 37.5 | 45 | 61 | -1 | -1 |
| 23.24 | 19.15 | 28.73 | 22.6 | 19.1 | 29 | 40 | 46 | -1 | -1 |
| 27 | 0 | 0 | 30 | 30 | 37.5 | 26 | 110 | 1 | 1 |
| 27 | 0 | 0 | 20 | 30 | 37.5 | 26 | 110 | 1 | 1 |

Table 2 - Performance of testing dataset

| d ($kN/m^3$) | $c_A$ (kPa) | $c_B$ (kPa) | $\phi_A(^0)$ | $\phi_B(^0)$ | $\psi_p(^0)$ | $\psi_f(^0)$ | H(m) | Actual class | Predicted class |
|---|---|---|---|---|---|---|---|---|---|
| 27 | 0 | 0 | 20 | 30 | 43 | 26 | 50 | 1 | 1 |
| 27 | 20 | 20 | 20 | 30 | 43 | 26 | 60 | 1 | 1 |
| 27 | 0 | 0 | 10 | 10 | 43 | 26 | 60 | -1 | -1 |
| 24 | 49 | 49 | 20 | 30 | 65 | 31 | 40 | 1 | 1 |
| 20 | 0 | 0 | 40 | 4 | 45 | 60 | 100 | -1 | -1 |
| 19.9 | 40 | 19 | 22 | 22 | 37 | 42 | 140 | -1 | -1 |
| 26.66 | 0 | 0 | 35 | 35 | 30 | 42 | 150 | 1 | 1 |
| 18.84 | 30.07 | 3.6 | 30 | 36.7 | 37.5 | 45 | 61 | -1 | 1 |
| 27 | 0 | 0 | 20 | 30 | 37.5 | 26 | 50 | 1 | 1 |
| 27 | 0 | 0 | 15 | 15 | 43 | 26 | 60 | -1 | -1 |

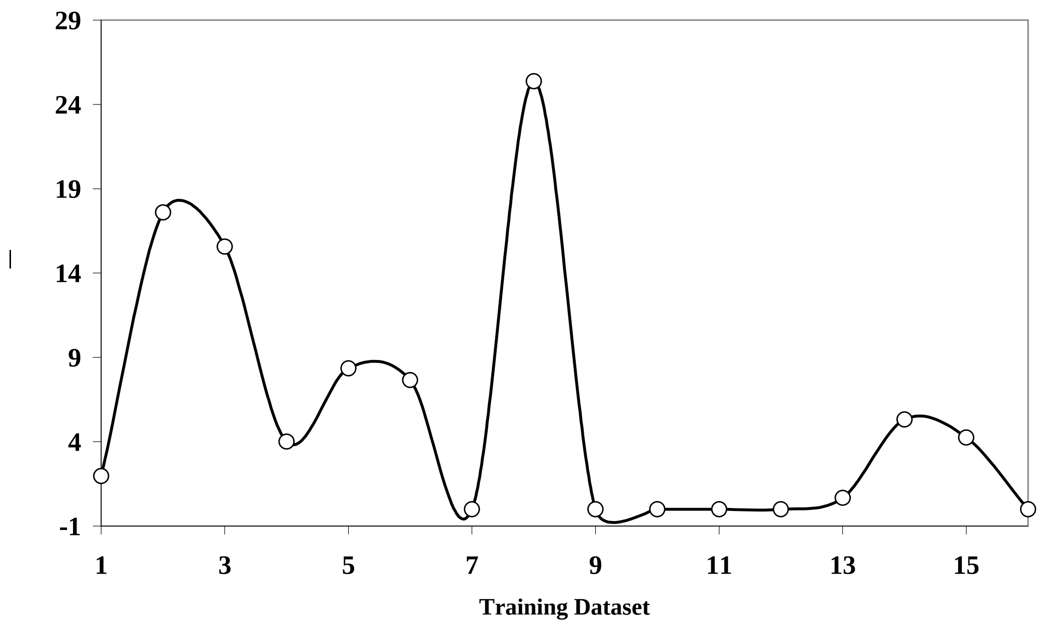


Fig. 6 - Values of $\alpha$

The following equation can be developed for the prediction of status(s) of rock slope (by putting $K(x_i, x) = \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2\sigma^2}\right\}$, $\sigma$=1, $l$=16 and b=0 in Eq. 13)

$$s = sign\left(\sum_{i=1}^{16} \alpha_i y_i \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2}\right\}\right) \qquad (15)$$

Figure 6 shows the value of $\alpha$. Practicing engineers can use the above equation for the prediction of stability of rock slope.

## 4. CONCLUSIONS

This paper describes SVM model for prediction of stability of rock slope. The developed SVM model gives promising result for the prediction of stability of rock slope. An equation has been also developed for the prediction of status of rock slope. Once the model is developed and trained, it requires only a small fraction of computational time and gives promising result. Moreover, the model can always be updated to yield better results, as new data becomes available. In summary, this paper gives a robust models based on SVM for prediction of stability of rock. SVM model can be also applied for solving different problems in rock mechanics.

### *References*

Boser, B.E., Guyon, I.M, and Vapnik, V.N. (1992). A Training Algorithm for Optimal Margin Classifiers, In D. Haussler, editor, 5th Annual ACM Workshop on COLT, Pittsburgh, PA, ACM Press, pp. 144-152.

Cortes, C. and Vapnik, V.N. (1997). Support Vector Networks, Machine Learning, 20, pp. 273-297.

Cristianini, N. and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machine, University press, Cambridge, London.

Fletcher, R.(1987). Practical Methods of Optimization, Wiley, Chichester, Newyork.

Goh, A.T.C. and Goh, S.H. (2007). Support Vector Machines: Their Use in Geotechnical Engineering as Illustrated using Seismic Liquefaction Data, Computers and Geotechnics, 34(5), pp. 410-421.

Goodman, R.E. and Kieffer, D.S. (2000). Behavior of Rock in Slope, J. Geotech Geoenviron Eng., 126(8), pp. 675–84.

Gualtieri, J.A., Chettri, S.R., Cromp, R.F. and Johnson. L.F. (1999). Support Vector Machine Classifiers as Applied to AVIRIS Data, In the Summaries of the Eighth JPL Airbrone Earth Science Workshop.

Hack. R., Price, D. and Rengers, N. (2003). A New Approach to Rock Slope Stability - A Probability Classification (SSPC), Bull Eng Geol Environ, 62, pp. 167- 184 and erratum, pp. 185–185.

Hoek, E. and Bray, J.W. (1981). Rock Slope Engineering, 3rd ed, Institute of Mining and Metallurgy, London.

Jaeger, J.C. (1971). Friction of Rocks and Stability of Rock Slopes, Geotechnique, 21(2), pp. 97–134.

Osuna, E, Freund, R., Girosi, F. (1997). An Improved Training Algorithm for Support Vector Machines, Proc. IEEE Workshop on Neural Networks for Signal Processing 7, Institute of Electrical and Electronics Engineers, New York, pp. 276–285.

Sakellatiou, M.G. amd Ferentinou, M.D. (2005). A Study of Slope Stability Prediction using Neural Networks, International Journal of Geotechnical and Geological Engineering, 23, pp. 419-445.

Samui, P, Kurup, P. and Sitharam, T.G. (2008). OCR Prediction using Support Vector Machine Based on Piezocone Data, Journal of Geotechnical and GeoEnvironmental Engineering, 134, 6, pp. 894-898.

Samui, P. (2008). Support Vector Machine Applied to Settlement of Shallow Foundations on Cohesionless Soils, Computers and Geotechnics, 35, 3, pp. 419-427.

Siad, L. (2003). Seismic Stability Analysis of Fractured Rock Slopes by Yield Design Theory, Soil Dynam Earthquake Eng., 23, pp. 203–12.

Smola, A.J. and Scholkopf, B. (2004). A Tutorial on Support Vector Regression, Statistics and Computing, 14, pp. 199-222.

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory, Springer ,New York.

Vapnik, V. N. (1998). Statistical Learning Theory, Wiley, New York.

Yarahmadi, Bafghi, A.R. and Verdel, T. (2005). Sarma-Based Key-group Method for Rock Slope Reliability Analyses, Int J Numer Anal Meth Geomech, 29, pp. 1019–43.

Zanbak, C. (1983). Design Charts for Rock Slopes Susceptible to Toppling, J Geotech Eng Div., ASCE, 190(8), pp. 1039–62.