International Journal of Statistics and Applied Mathematics 2023; 8(5): 07-13

ISSN: 2456-1452
Maths 2023; 8(5): 07-13

https://www.mathsjournal.com
Received: 10-06-2023
Accepted: 20-07-2023

Halagundegowda GR
Scientist-C, Statistics Section, Central Silk Board, Bengaluru, Karnataka, India

Abhishek Singh
Associate Professor, Department of Farm Engineering, IAS, BHU, Varanasi, Uttar Pradesh, India

Mohan Kumar TL
Assistant Professor, Department of Agricultural Statistics, UAS, Bengaluru, Karnataka, India

Naveena K
Scientist-B, Centre for Water Resources Development and Management, Kozhikode, Kerala, India

Corresponding Author:
Halagundegowda GR
Scientist-C, Statistics Section, Central Silk Board, Bengaluru, Karnataka, India

Evaluation of classification ability of support vector machine (SVM) in binary classification problems

Halagundegowda GR, Abhishek Singh, Mohan Kumar TL and Naveena K

Abstract
The objective of the study is to investigate the classification ability of SVM architecture including internal parameters and kernel types on farmers' classification based on drought coping strategies. Support vector machines with sigmoid kernel trick, having hyper parameter gamma=0.096 and classification type 1 with capacity C =2.40. The optimal value of hyper parameter has been computed by i=1000 number of iterations for tuning the model by random grid search optimization approach with sigmoid kernel trick and coefficient value 0.06. SVM with the Sigmoid kernel trick has 97 overall support vectors and 93 bounded support vectors. The classification summary of SVM Sigmoid depicts that overall, 93.3% of these cases were correctly classified by the model and the remaining 7% were wrongly classified. The SVM with RBF kernel trick, having hyperparam gamma=0.083 and classification type 1 with capacity C =1.20. The optimal value of the hyperparameter has been computed by i=1000 number of iterations for tuning the model by random grid search optimization approach with radial basis function kernel trick. SVM with RBF kernel trick has 82 overall support vectors and 74 bounded support vectors.

Keywords: Support vector machine, kernel tricks, random grid search, drought coping strategies, sigmoid, RBF, f1 score, etc.

1. Introduction

India ranks first among the rain-fed agricultural countries of the world in terms of both extent and value of produce. Rainfed agriculture is practiced in two-thirds of the total cropped area of 162 million hectares (66 percent). Rainfed agriculture supports 40 percent of India's population and contributes 44 percent to the national food basket. Nearly 50 percent of the total rural workforce and 60 percent of the cattle heads of the country are located in rainfed areas. Importance of the rainfed agriculture is obvious from the fact that 55 percent of rice, 91 percent of coarse grains, 90 percent of pulses, 85 percent of oilseeds, and 65 percent of cotton are grown in rainfed areas (GOK News, 2022). The State of Karnataka has 72 percent of the cultivable area is rainfed and only 28 percent is under irrigation (GOK News, 2022). The State is the second largest in terms of arid region, next only to Rajasthan in India (in terms of total geographical area prone to drought). Drought is a common phenomenon in the State of Karnataka.

When drought occurs in a particular area obviously affects crop and livestock production, In order to reduce the effect of drought on farm production and to stabilize farm income, farmers have to take some systematic measures such measures are called drought coping mechanisms. The objective of the study was to identify the determinants of the adoption of drought-coping mechanisms in order to balance and stabilize the farm income of the stakeholders. The study also helps to know how to mitigate the effect of drought on farmer's livelihoods.

Jeetendra Prakash Aryal *et al*. (2019) [7] applied a multivariate probit model for the simultaneous multiple adoption decisions and ordered probit models for assessing the factors affecting the level of adoption. The factors that determine the probability and level of adoption of multiple climate-smart agriculture practices include seeds of stress-tolerant varieties, minimum tillage, laser land leveling, site-specific nutrient management, and crop diversification.

Halagundegowda *et al*. (2017) [6] used discriminant analysis to examine the factors influencing on adoption of drought coping strategies, results are Farm Size (0.552), Extension Visits (0.574), Crop Diversification (0.321) and Crop Insurance (0.368) are relatively more important and positively influencing on discrimination of farmers group. Whereas variables like Age (-0.516) negatively influenced on discrimination of adopters and non-adopters.

The data related to the adoption of any agriculture technology have qualitative response variables with two or more categories. Most of the studies used qualitative response models such as the logit model, probit model, and multivariate techniques like discriminant analysis to measure the farmer's perception towards the adoption of any agriculture technology. In this study, we have chosen a machine learning model such as Support Vector Machine (SVM), in order to assess the classification ability.

2. Materials and Methods

Nature and source of data

The current study utilizes both classification and prediction techniques. The household data was used to fit the classificatory statistical models and the data were recorded on Socio-characters of farmers of Kolar districts of Karnataka (India). The data is mainly related to coping strategies implemented against drought by the farmers of this region and was collected by employing the multi-stage sampling design. Multi-stage sampling design was employed in the present study for the selection of sample respondents. In the first stage, a rainfed crops growing district namely Kolar was selected. In the second stage, major rainfed crops growing talukas from each district were selected. Then during the third stage, major rainfed crops growing villages were selected from each taluk.

Table 1: The number of respondents and splitting ratio of training and test data set

Class	Total Respondents	Training set (80%)	Test set (20%)
Non-Adopters	62	50	12
Adopters	88	70	18
Total	150	120	30

The number of respondents in the training and test data set is reported in Table 1. In the first phase training data are used for building the model as well as estimating parameters using various classification techniques. Before training the model each attribute is normalized to zero mean and unit variances, which will improve the performance of the model as well as cut down the learning time. The following table gives the variable of interest and the unit of measurement used in this study.

Table 2: Variables Encoding Summary

Code	Variables	Measurement
Y	Adoption behavior	Y= 0 for Non-Adopters = 1 for Adopters
X_1	Age of the farmer	Number of years
X_2	Education of the farmer	Formal Years of Education
X_3	Household Size	Number of family members
X_4	Farm Size	Number of acre's
X_5	Farming Experience	Number of years
X_6	Animal Husbandry	Number of farm animals
X_7	Media Exposure	Number of sources exposed frequently
X_8	Extension Visits	Number of Visits
X_9	Crop Diversification	Number of Crops Grown in that year
X_{10}	Income Status	In Rupees (Rs.)
X_{11}	Worth of Liquidating Assets	In Rupees (Rs.)
X_{12}	Crop Insurance by Government	In Rupees (Rs.)

Support Vector Machine for Classification

Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

(i) Linear Support Vector Machine for Separable Data

Proposed an original linear SVM formulation for separable data. For a given binary classification problem, the objective is to estimate functions f with parameter vector θ such that $\{f(x;\theta): R^n \to \{-1,+1\}\}$ using a finite set of training data. Let the training data set consist of $\{x_i, y_i\}_{i=1}^{N}$ with input patterns $x_i \in R^n$ and their respective class labels $y \in \{-1,+1\}$. When the training data is linearly separable, a separating hyperplane (a hyperplane that separates the positive from the negative examples, this hyperplane is shown in Fig.1) of the form

$$w^T x + b = 0, \quad (1)$$

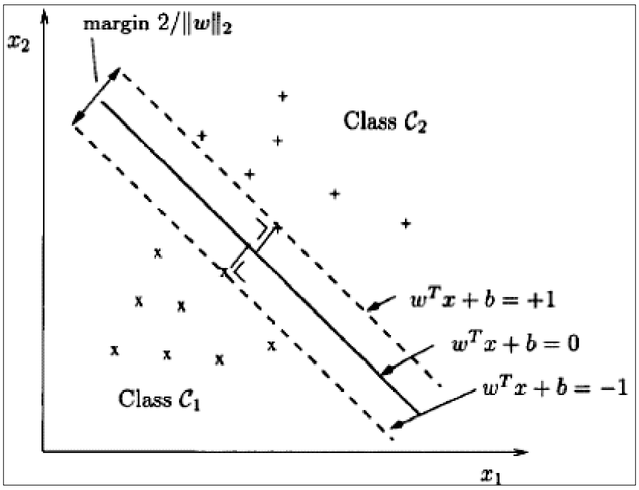

Fig 1: Construction of optimal hyperplane for binary classification problem illustrated for two-dimensional input space.

Can be fitted to correctly classify training patterns, where weight vector w is normal to the hyperplane and defines its orientation, b is the bias term and T is the transpose. From equation 1 a linear classifier (decision function) is given by:

$$y(x) = sign(w^T x + b), \quad (2)$$

Which classifying class C₂ ($y_i = +1$ if $w^T x + b \geq 0$)and class C₁($y_i = -1$ if $w^T x + b \leq 0$)patterns. Let $H_+(w^T x + b = +1)$ and $H_-(w^T x + b = -1)$ the shortest distance from the separating hyperplane to the closest class 2 (class 1), then the margin of the hyperplane is defined as the sum of$H_+$and H_-, i.e. $H_+ + H_-$. An optimal hyperplane for a linearly separable set of training data is defined as the linear decision function with the maximal margin between the vectors of two classes, as is shown in Fig.1. The support vector algorithm will construct this optimal separating hyper plane. It is shown that the optimal hyper plane will have good generalization abilities, and only a relatively small amount of training data is needed to construct this plane. The set of margin-determining training vectors is called the support vectors.

Assume all training data satisfy

$$\begin{cases} w^T x_i + b \geq +1, if\ y_i = +1 \\ w^T x_i + b \leq -1, if\ y_i = -1 \end{cases} \quad i = 1,2, \dots N, \quad (3)$$

This can be combined into a single set of equalities:

$$y_i(w^T x_i + b) - 1 \geq 0, i = 1,2, \dots N, \quad (4)$$

Where N is the training data size
To find the optimal separating hyperplane, it is necessary to maximize the margin $H_+ + H_-$. suppose x_1 and x_2 with $y_i = +1$ and $y_i = -1$ are Class C₂ and Class C₁ patterns closest to the hyperplane of H_+ and H_- respectively. For maximal separation, the hyperplane should be as far away as possible from each of them. By letting$\|.\|$ be the l_2 norm of a vector, we can get

$$w^T x_1 + b = +1 \text{And} w^T x_2 + b = -1$$

$$w^T(x_1 - x_2) = +2$$

$$\frac{w}{\|w\|}.(x_1 - x_2) = \frac{2}{\|w\|}$$

Where $\|w\| = \sqrt{w^T w} = \sqrt{w_1^2 + w_2^2 + \cdots + w_n^2}$ is the l_2 norm of a vector.

Maximizing the margin is equivalent to maximizing $2/\|w\|$, which is in turn the same as solving

$$J_P \min_{w,b} \frac{1}{2}\|w\|^2 \text{ or} J_P \min_{w,b} \frac{1}{2} w^T w \quad (5)$$

Subject to constraints

$$y_i(w^T x_i + b) \geq 1, i = 1,2, \dots N, \quad (6)$$

Constructing the optimal hyperplane is therefore convex Quadratic Programming (QP) problem. We call this optimization problem is in the primal formulation. Therefore SVM formulations are done within the context of a convex QP optimization problem. The general methodology is to start formulating the problem in the primal weight space as a constrained optimization problem. Then Lagrange multipliers and Kursh-Kühn-Tucker (KKT) complimentary conditions are used to find the optimal solution. Under the condition for optimality, the above QP problem is finally solved in the dual space of the Lagrange function.
The Lagrangian multipliers$\alpha_i \geq 0$, are introduced for each of the constrains in Equations 6 to get the following Lagrangian:

$$L(w, b;\ \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i(y_i(w^T x_i + b) - 1) \quad (7)$$

The objective is to minimize equation (7) with respect to w and b under the requirement that the derivatives of the Lagrangian with respect to all the $\alpha_i$vanish. This must be subject to the constrains that the Lagrangian multipliers *remain* non-negative. Since all the constrains are linear in convex quadratic optimizations, it is possible to equivalently solve the dual optimization problem of maximizing equation 7, such that the gradient of L with respect to w and b vanishes, and it is required that $\alpha_i \geq 0$. That is

$$\frac{\partial}{\partial b} L(w, b;\ \alpha) = 0 \text{ and} \frac{\partial}{\partial w} L(w, b;\ \alpha) = 0 \quad (8)$$

and thus

$$\sum_{i=1}^{N} \alpha_i y_i = 0 \text{ and } w = \sum_{i=1}^{N} \alpha_i y_i x_i \quad (9)$$

By substituting equation (8) in(9), the dual form of the optimization problem is derived, which is also QP determined by

$$J_D \max_{\alpha} = \sum_i^N \alpha_i - \frac{1}{2} \sum_i^N \sum_j^N \alpha_i \alpha_j\, y_i y_j x_i^T x_j \quad (10)$$

Subject to

$$\alpha_i \geq 0 \text{ and} \sum_{i=1}^{N} \alpha_i y_i = 0\,;\ i = 1,2, \dots N \quad (11)$$

Thus by solving the dual QP problem, the coefficients α_i are obtained. These coefficients are then used to calculate w from equation (11). The vector w will be a solution to the problem (11). The decision function from equation (11) can be rewritten as

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i y_i x^T x_i + b) \quad (12)$$

The decision surface in equation 12 is determined by the N Lagrangian multipliers α_i. these multipliers are either zero or positive. The subset of zero multipliers will have no effect on the decision function and can be omitted. It is the set of positive multipliers that influences the classification, and their corresponding training vectors are called the *support vectors*. The last unknown parameter b is determined by taking average of $b = y_i - \sum_{i=1}^{N} \alpha_i y_i x^T x_i$ for all support vectors.

(ii) Feature Function and Kernel Functions

The idea of transforming the input space into a feature space of a higher dimension by using feature functions $\varphi(x_i)$and then performing a linear classification in that higher dimensional space is central to SVM. However, the feature space may have a very high dimensionality, even infinite. An obvious consequence is to avoid the inner product of feature functions $\varphi(x_i)$. Fortunately, a method was developed to generate a mapping into a high-dimensional feature space with kernels. The rationale that prompted the use of kernel functions is to enable computations to be performed in the original input space rather than the high-dimensional (even infinite) feature space. Using this approach, the SVM algorithm avoids the evaluation of the inner product of the feature functions.

Under certain conditions, an inner product in the feature space has an equivalent kernel in the input space. For any symmetric continuous function $K(x,z)$ satisfying Mercer's condition, there exists a Hilbert space H, a map $\varphi(x): R^n \to R^{n_{\mathcal{H}}}$ and a positive number λ_i such that one can write $K(x,z) = \sum_{i=1}^{n_{\mathcal{H}}} \lambda_i \varphi(x)\varphi(z)$, where $x, z \in R^n$ and $n_{\mathcal{H}}$ is the dimension of H (which can be infinite-dimensional). Mercer's condition requires that

$$\int K(x,z)g(x)g(z)dxdz \geq 0 \quad (13)$$

Then *Kernel function* can be expressed as the inner product (often called the dot product) which is given as $K(x,z) = \varphi(x)^T\varphi(z)$, where $\varphi(x) = \sqrt{\lambda_i}\varphi_i(x)$ and $\varphi(z) = \sqrt{\lambda_i}\varphi_i(z)$. The application of kernel function $(x,z))$ is often called the kernel trick.

Table 3: Some typical choices of kernel function are

Kernel Type	Expression
Linear SVM	$K(x,x_i) = x_i^T x$
Polynomial of degree d	$K(x,x_i) = (x_i^T x + k)^d$
Radial Basis Function (RBF)	$K(x,x_i) = exp\{-\|x-x_i\|^2/2\sigma^2\}$ *or* Equivalently $K(x,x_i) = exp\{-\gamma\|x-x_i\|^2\}$
Multi-Layer Perceptron (MLP)	$K(x,x_i) = \tanh(k_1 x_i^T x + k_2)$

The Mercer condition holds for all σ or γ values in the RBF kernel case and k values in the polynomial case but not for all possible choices of k_1, k_2 in the MLP case.

Table 4: Classification Ability Measures:

		Predicted Class	
		Class=Yes	Class=No
Actual Class	Class=Yes	True Positive	False Negative
	Class=No	False Positive	True Negative

True positives and true negatives are the observations that are correctly predicted and need to be maximized, we need to minimize the false positives and false negatives.

True Positives (TP): These are the correctly predicted positive values which means that the value of the actual class is "Yes" and the value of the predicted class is also "Yes". E.g.: if the actual class value indicates Adopters and the predicted class tells you the same thing.

True Negatives (TN): These are the correctly predicted negative values which means that the value of the actual class is "No" and the value of the predicted class is also "No". For: if the actual class value indicates Non Adopters and the predicted class tells you the same thing.

False Positives (FP): When the actual class is "No" and the predicted class is "Yes". For: if the actual class says Non Adopters but the predicted class tells you that Adopters.

False Negatives (FN): When the actual class is "Yes" but the predicted class in "No". E.g.: if the actual class value indicates Adopters and the predicted class tells you those farmers are Non Adopters. After these four parameters we can calculate Accuracy, Precision, Recall, and F1 score.

Accuracy: Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observations to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positives and false negatives are almost the same.

Accuracy = TP+TN/TP+FP+FN+TN

Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answers is of all passengers that were labeled as survived, how many actually survived? High precision relates to the low false positive rate.

Precision = TP/TP+FP

Recall (Sensitivity): Recall is the ratio of correctly predicted positive observations to all observations in actual class - yes. The question recall answers are: Of all the passengers that truly survived, how many did we label? Ex: We have got recall of 0.631 which is good for this model as it's above 0.5.

Recall = TP/TP+FN

F1 score: The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar costs. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

F1 Score = 2*[(Recall * Precision)/(Recall + Precision)]

Result and Discussion

Support Vector Machine with Sigmoid Kernel trick

Support Vector Machines (SVMs) is a generalized portrait classification algorithm based on statistical learning theory and developed to perform binary classification problem initially. Here we employed binary class SVM by solving a single optimization problem. The hyper-parameters of this model are estimated using a very efficient random grid search technique. Both the sigmoid kernel and RBF kernel method of SVM were fitted to the data, although many procedures are available in traditional statistical classification, the usefulness depends on assumptions and circumstances.

Table 5 explains the details about the model summary and specifications of the SVM model, including the number of support vectors and their types, the kernels, and their parameters. The model was constructed to classify the binomial dependent variable by including 12 predictors and there are two major classification types in SVM such as classification type 1 and classification type 2, The capacity value indicates the tradeoff between two boundaries, which is the hyperparameter of the predictive model, the current research takes up classification type 1 with capacity C =2.40.

Table 5: Model Summary

Model Specifications	Value
Number of independents	12
	Classification type 1(Capacity= 2.40)
SVM type	Sigmoid (Gamma= 0.096, Coefficient= 0.06)
Kernel type	
Number of SVs	97 (93 bounded)
Number of SVs (0)	48
Number of SVs (1)	49

In classification problems, only two hyper-parameters are needed to be defined by user i.e. the trade-off between model capacity and training error represented by C (Capacity) and the kernel parameter γ (Gamma). These hyperparameters are directly coded with real values within a given search space to randomly generate M number of initial particles of set S. Search space of hyper-parameters C,γ are respectively restricted to ranges of $[C_{min}, C_{max}]$, $[\gamma_{min}, \gamma_{max}]$are randomly generated. The entire operation has been system-inbuilt algorithms in STATISTICA 8 version software. The current research has identified the capacity parameter C by randomly grid search method and the model provides saturated result at the value C=2.40. For finding the optimal value of hyper parameters traditionally various optimization techniques such as particle swarm optimization, genetic algorithms, ant colony optimization techniques, simulated annealing algorithms, etc. but the current research have taken a random grid search algorithm by specifying the range of values between minimum and maximum with particle incremental value as an interval.

For instance, large C forces the SVM classification algorithm to reduce the training errors, which in turn can be accomplished by increasing the machine capacity and as a consequence may reduce the margin. This is contrary to the main objective of margin maximization and also does not guarantee a good generalization performance of the classifier. Therefore, the selection of optimal hyper-parameters is an important step in SVM modeling. Further, several kernel functions $k(x_i, x_j)$are available in the literature, like the polynomial function, sigmoid kernel, Gaussian kernel, and radial basis function (RBF).

The current research has taken up with Sigmoid kernel and having hyperparam gamma=0.096, which was achieved by selecting the range of the parameter $[\gamma_{min}, \gamma_{max}]$. The γ_{min}=0 and γ_{max}=10 with the incremental value =0.2 by providing this information to the random grid search end up with the saturated result of gamma γ=0.096 and coefficient value mainly depends on how exactly the position and orientation of hyperplane, which depends on the weights which assign to the input vectors. It ranged from 0 to 3 with the incremental value of 0.02 winding up with the saturated final result of 0.06. Set the iteration number (t) from 1 to a maximum number of iterations and evaluate inertia weight $w(t)$ generation by generation according to the model equation. The current research has taken i=1000 iterations for tuning the model.

The formulation of SVM learning is based on the principle of structural risk minimization which attempts to minimize both the generalization bound and the empirical error instead of minimizing only empirical error implemented in traditional models. This technique is said to be independent of the dimensionality of feature space as a generalization is obtained by maximizing the margin, which corresponds to the minimization of the weight vector in a canonical framework. The training examples that are closest to the maximum margin hyperplane are called "support vectors", which can be sparse, and lie on the boundary and as such summarize information required to separate the data. The current research has 97 overall support vectors and 93 bounded support vectors and coming to category wise, there are 48 support vectors are in the side of the non-adopters category and around 49 support vectors are in the side of the adopter's category. The result indicates there is a sufficient number of vectors is at boundaries which makes the hyper plane for effectively classifying the cases into respective classes.

Table 6: Classification Matrix

Sample	Observed	Predicted		
		Adopters	**Non Adopters**	**Percent Correct**
Training	Adopters	62	8	88.5%
	Non-Adopters	8	42	84.0%
	Overall Percent			86.2%
Testing	Adopters	17	1	94.4%
	Non-Adopters	1	11	91.6%
	Overall Percent			93.3%

Table 6 shows that the cells on the diagonal of the cross-classification of cases are correct predictions for each sample. The cells of the diagonal of the cross-classification of cases are incorrect predictions of the cases used to create the model, 62 of the 70 farmers who previously adopted the drought coping strategies are classified correctly. 42 of the 50 non-adopters are classified correctly. Overall, 86.2% of the training cases are classified correctly, corresponding to the 13.8% incorrect shown in the model summary table. A better model should correctly identify a higher percentage of the cases.

Classifications based on the cases used to create the model tend to be too "optimistic" in the sense that their classification rate is inflated. The testing sample helps to validate the model; here 93.3% of these cases were correctly classified by the model. This suggests that overall our model is in fact correct and efficient in prediction and classification.

II. Support Vector Machine with RBF Kernel trick

Here, the result explains the detail about the model summary and specifications of the SVM model, including the number of support vectors and their types, the kernels, and their parameters. The current research has experimented most commonly used RBF kernel function.

$$k(x_i, x_j) = exp\{-\|x - x_i\|^2/2\sigma^2\}$$

to train nonlinear SVM model. It has only one hyper-parameter that needs to be pre-determined and yields good performance under general conditions. Training nonlinear SVR with RBF kernel function requires optimization of two hyper-parameters, *viz.* (i) Regularization parameter$C$, which balances the complexity and approximation accuracy of the model, and (ii) Kernel bandwidth parameter σ, which represents variance of RBF kernel function. However, here we can notice that in the above RBF kernel function can be equivalently represented as$k(x_i, x_j) = exp\{-\gamma\|x - x_i\|^2$, where$\gamma = 0.5/\sigma^2$. This is necessary due to SVM package requirements since it works with gamma values and not directly with the width σ. Here we adopt $k(x_i, x_j) = exp\{-\gamma\|x - x_i\|^2$ form RBF kernel function, which requires γ value to be estimated instead of σ.

Table 7: Model Summary

Model Specifications	Value
Number of independents	12
SVM type	Classification type 1 (Capacity= 1.200)
Kernel type	Radial Basis Function (Gamma= 0.083)
Number of SVs	82 (74 bounded)
Number of SVs (0)	41
Number of SVs (1)	41

Table 7 explains the model summary, the fitted model has 12 predictors and used the classification type 1 by using capacity parameter C=1.200, the capacity parameter was computed by random grid approach and the current research has taken up with Radial basis function as kernel tricks and having hyper parameter gamma=0.083, which has achieved by selecting the parameters range[$\gamma_{min}, \gamma_{max}$]. The γ_{min}=0 and γ_{max}=10 with the incremental value =0.02. It has 82 overall support vectors and 74 bounded support vectors and coming to category, there are 41 support vectors are on the side of the non-adopters category and around 49 support vectors are on the side of the adopters category. The result indicates there is a sufficient number of vectors is at boundaries which makes the hyperplane for effectively classifying the cases into respective classes.

Table 8: Classification Matrix

Sample	Observed	Predicted		
		Adopters	Non-Adopters	Percent Correct
Training	Adopters	65	5	92.83%
	Non-Adopters	6	44	88.0%
	Overall Percent			90.42%
Testing	Adopters	13	5	72.2%
	Non-Adopters	1	11	91.6%
	Overall Percent			81.44%

Table 8 shows that the cells on the diagonal of the cross-classification of cases are correct predictions for each sample. The cells off the diagonal of the cross-classification of cases are incorrect predictions of the cases used to create the model, 65 of the 70 farmers who previously adopted the drought coping strategies are classified correctly. 44 of the 50 non-adopters are classified correctly. Overall, 90.42% of the training cases are classified correctly, corresponding to the 9.58% incorrect shown in the model summary table. A better model should correctly identify a higher percentage of the cases.

Table 9: Classification ability Measures

SVM MLP	Training	Testing	SVM RBF	Training	Testing
Accuracy	86.22	93.33	Accuracy	90.42	81.44
Recall	0.8857	0.9444	Recall	0.9285	0.7222
Precision	0.8857	0.9444	Precision	0.9154	0.9285
F1	0.8857	0.9444	F1	0.9219	0.8125

Table 9 depicts the classification measures computed for two variant models of SVM, All measures are good for both models, and here the accuracy rate is interpreted if the model has high accuracy then our model is best, and it's more than 70% in both cases. The Recall and Precision also work well for both the model, which shows more than 0.50, among these two models SVM MLP poses well compared to SVM RBF for testing sets. F1 Score also poses a good score for both the models; overall both models have potential classification ability. Further comparative performance analysis shows SVM MLP is better than SVM RBF.

4. Conclusion

Support vector machines with sigmoid kernel trick, having hyperparameter gamma=0.096 and classification type 1 with capacity C =2.40. The optimal value of the hyperparameter has been computed by i=1000 number of iterations for tuning the model by random grid search optimization approach with sigmoid kernel trick and coefficient value 0.06. SVM with Sigmoid kernel trick has 97 overall support vectors and 93 bounded support vectors and coming to category wise, there are 48 support vectors are in the side of the non-adopters category and around 49 support vectors are in the side of the adopters category. The classification summary of SVM Sigmoid depicts that overall, 86.2% of the training cases are classified correctly, corresponding to the 14.8% incorrect classification. The testing sample helps to validate the model; here 93.3% of these cases were correctly classified by the model and the remaining 7% were wrongly classified.

The SVM with RBF kernel trick, having hyperparam gamma=0.083 and classification type 1 with capacity C =1.20. The optimal value of the hyperparameter has been computed by i=1000 number of iterations for tuning the model by random grid search optimization approach with radial basis function kernel trick. SVM with RBF kernel trick has 82 overall support vectors and 74 bounded support vectors and coming category-wise, there are 41 support vectors are in the side of the non-adopters category and around 41 support vectors are in the side of the adopter's category. The classification summary of SVM RBF depicts that overall, 90.42% of the training cases are classified correctly, corresponding to the 9.58% incorrect classification. The testing sample helps to validate the model; here 81.44% of these cases were correctly classified by the model and the remaining 19.66% were wrongly classified.

5. References

1. Bharadwaj A, Dahiya S, Jain R. Discretization-based Support Vector Machine (D-SVM) for Classification of Agricultural Datasets. International Journal of Computer Applications. 2012;40(1):0975-8887.
2. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, *et al*. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sciences. 2017;22:961-975. DOI: 10.1016/j.tplants.2017.08.011.
3. Devadas R, Denham RJ, Pringle M. Support Vector Machine Classification of Object-Based Data for Crop Mapping, Using Multi-Temporal Landsat Imagery. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; c2012. p. XXXIX-B7.
4. González-Camacho JM, Ornella L, Pérez-Rodríguez P, Gianola D, Dreisigacker S, Crossa J, *et al*. Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. Plant Genome. 2018;11:1-15. DOI: 10.3835/plantgenome2017.11.0104.
5. Guo X, Meng C. Research on Support Vector Machine in Image Denoising, International Journal of Signal Processing, Image Processing and Pattern Recognition. 2015;8(2):19-28.
6. Halagundegowda GR, Abhishek Singh, Meenakshi HK. Discriminant analysis for classification of farmers based on adoption of drought coping mechanisms, Agriculture Update. 2017;12:2643-2648.
7. Jeetendra Prakash Aryal ML, Sapkota JTB, Rahut DB, Rai M, Jat HS, *et al.* Learning adaptation to climate change from past climate extremes: Evidence from recent climate extremes in Haryana, India, International Journal of Climate Change Strategies and Management; c2019. DOI: 10.1108/IJCCSM-09-2018-0065
8. Jog SM, Dixit M, Rajgopalan A, Ranade SD. Balanced Vs Imbalanced Training Data: Classifying Rapid eye Data with Support Vector Machines. International Journal of Innovative Research in Electrical, Electronics,

Instrumentation and Control Engineering. 2016;4(6):158-163.

9. Kaiyi W, Chunjiang Z, Tian ZF, Juan WX, Hong C, Fang YY, *et al*. Application of a support vector machine to HACCP in the animal meat industry, New Zealand Journal of Agricultural Research. 2010;50(5):743-748.
10. Robothama H, Castillo J, Bosch P, Perez-Kallens J. A comparison of multi-class support vector machine and classification tree methods for hydro acoustic classification of fish-schools in Chile, Fisheries Research. 2011;111:170-176.
11. Sahu SK, Prasad MBNV, Tripathy BK. A Support Vector Machine Binary Classification and Image Segmentation of Remote Sensing Data of Chilika Lagloon, International Journal of Research in Information Technology. 2015;3(5):191-204.
12. Surase R, Kale K. Multiple Crop Classification Using Various Support Vector Machine Kernel Functions, International Journal of Engineering Research and Applications**.** 2015;5(1):103-109.
13. Surase R, Kale K. Performance evaluation of support vector machine and maximum likelihood classifier for multiple crop classification, International Journal of Remote Sensing & Geoscience (IJRSG). 2015;4(1):62-69.
14. Ustuner M, Sanli FB, Abdikan S. Balanced Vs Imbalanced Training Data: Classifying Rapid eye Data with Support Vector Machines. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2016;41(B7):12-19.
15. Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation, and signal processing, In Mozer M, Jordan M, Petsche T. (Eds) Advances in Neural Information Processing Systems. 1997;9:281-287. Cambridge, MA, MIT Press.
16. Wu L, Wen Y. Weed/corn seedling recognition by support vector machine using texture features. African Journal of Agricultural Research. 2009;4(9):840-846.
17. Yang Z, Gub XS, Lianga XY, Linga LC. Genetic algorithm-least squares support vector regression based predicting and optimizing model on carbon fibre composite integrated conductivity. Materials and Design. 2011;31**:**1042-1049.
18. Yao T, Cheng J, Xu B, Zhang M, Hu Y, Zhao J, *et al.* Support vector machine (SVM) classification model based rational design of novel tetronic acid derivatives as potent insecticidal and acaricidal agents. The Royal Society of Chemistry. 2015;5:12-19.
19. Yao XJ, Panaye A, Doucet JP, Chen HF, Zhang RS, Fan BT, *et al.* Comparative classification study of toxicity mechanisms using support vector machines and radial basis function neural networks. Analytica Chimica Acta. 2005;535:259-273.
20. Zhao M, Fu C, Ji L, Tang K, Zhou M. Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. Expert Systems with Applications. 2011;38:5197-5204.
21. Zolfaghari K, Shang S, McNairn H, Li J, Homyouni S. Using Support Vector Machine (SVM) for Agriculture Land Use Mapping with SAR Data: Preliminary Results from Western Canada, Agriculture and Agri-Food Canada. 2013;9(5):12-17.