



SVM approach for non-parametric method in classification and regression learning process on feature selection with ε -insensitive region

M. Premalatha^{1*} and C. Vijayalakshmi²

Abstract

Machine Learning is considered as a subfield of Artificial Intelligence. Machine Learning is concerned with the development of techniques and design the methods which enable the computer to learn. The field of machine learning is concerned with constructing computer program that automatically improve its performance with experience. In today's machine learning applications, support vector machines are considered (SVM) one of the most robust and accurate methods among all well-known algorithms and also being developed at a fast pace. The aim of SVM is to find the best classification function, in a two-class learning task, and to distinguish between members of the two classes in the training data. Hence, the goal of machine learning is to find the output hypothesis that performed the correct classification of the training data, but the other earlier algorithms to find the hypothesis that accurate fit to the data. SVM requires that each data instance is represented as a vector of real numbers. Hence, if there are categorical attributes, convert them into numerical data, then we using m numbers to represent an m -category attribute. Only one of the m numbers is one, and others are zero. Machine learning has been applied in various field such as medical diagnosis, bioinformatics, detecting credit card fraud, classifying DNA sequences, speech and handwriting recognition, object recognition in computer vision, and robot locomotion. The objective is algorithmic approach for non parametric methods to tractable for high dimensional massive datasets.

Keywords

Machine Learning, Support Vector Machine, Classification, Regression.

AMS Subject Classification

62H30, 62G08, 97R40.

¹ Department of Mathematics, Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.

² School of Advance Science, Department of Mathematics, VIT University, Chennai, Tamil Nadu, India.

*Corresponding author: ¹ premalatha.m@rajalakshmi.edu.in; ² vijusesha2002@yahoo.co.in

Article History: Received 22 November 2018; Accepted 09 May 2019

©2019 MJM.

Contents

1	Introduction	276
2	SVM System Architecture	277
3	SVM Margin Mathematically.....	277
4	Linearly Separable	277
5	Support Vector Machine for Regression.....	277
6	Regression with ε -insensitive region	278
7	Algorithm for Classification and Regression	278
8	Three Components of the Generalization Error ..	278
9	Comparisons based on Feature selection	279

10	Conclusion	279
	References	279

1. Introduction

The VC dimension measures the large number of training attributes that should be helped the function class to learn perfectly, by obtaining zero error rates on the training data [1, 2]. It can be analysis to prove that the actual error on the future data is bounded by a sum of two terms. The first term is representing the training error, and the second term is representing if proportional to the square root of the VC dimension h . Thus, if we can minimize h , we can minimize the future error, as long as we also minimize the training

error. SVM can be easily extended to perform numerical calculations easily to obtain better results [3]. One of the approaches is to break a large optimization problem into a series of smaller problems. The series of smaller problem that process iterates until all the decomposed optimization problems are solved easily to get the results [4, 5].

2. SVM System Architecture

Data divided in to two sub classes one is training data and another set is testing data. The training data set used to create the model but the testing data set used for testing the model [6]. For a linearly separable dataset, a classification function represents linearly that corresponds to a separating hyperplane $f(x)$, it should be passes through the middle of the two classes and also separating as two class. Once this function is determined, new data x_n a linear classification function corresponds to a separating hyperplane $f(x)$ that passes through the middle of the two classes, separating the two. Once this function is determined, new data x_n can be classified by simply testing the $sgn[f(x_n)]$; x_n belongs to the positive class if $f(x_n) > 0$.

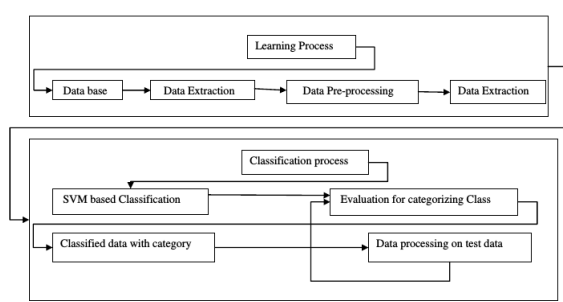


Figure 1. SVM Learning Process

3. SVM Margin Mathematically

The SVM goal is to correctly classify all the data during calculation time. For mathematical calculation as follows, The set of training vectors belonging to two separate classes, $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $x \in R^n$, $y \in \{-1, 1\}$ with hyperplane $(w \cdot x) + b = 0$ Optimally hyperplane if it is separated without error and the distance between the closest vectors to the hyperplane gives maximal [7].

4. Linearly Separable

In the linearly separable set contain Z training points, in that set each input x_i has D dimensional and it is belongs to one of two classes $y_i = -1$ or $+1$ $\{x_i, y_i\}$ where $i = 1, 2, 3 \dots Z$ $y_i \in \{-1, 1\}$ $x \in R^D$. A hyperplane on graphs of $x_1, x_2 \dots X_D$ when $D > 2$. This type of hyperplane can be formulated by $w \cdot x + b = 0$ where w is normal to the hyperplane, $\frac{b}{\|w\|}$ is the

perpendicular distance from the hyperplane to the origin. The aim of Support Vector Machines (SVM) is to orientate this hyperplane in such a way as to be as far as possible from the closest members of both classes [8].

To selecting the variables w and b so that our training data can be described by:

$$x_i \cdot w + b = \begin{cases} +1; & y_i = +1 \\ -1; & y_i = -1 \end{cases} \quad (4.1)$$

Combined equations represent as

$$y_i(x_i \cdot w + b) - 1 \geq 0 \text{ for all } i \quad (4.2)$$

The points that lie closest to the separating hyper plane, i.e. the Support Vectors, then the two planes H_1 and H_2 that these points lie on can be described as follows

$$x_i \cdot w + b = \begin{cases} +1; & H_1 \\ -1; & H_2 \end{cases} \quad (4.3)$$

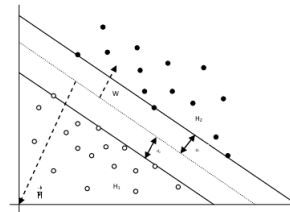


Figure 2. Separating Hyperplane linearly separable classes

Here d_1 and d_2 represent the distance from H_1 and H_2 to the hyperplane to it. The hyper plane's equidistance from H_1 and H_2 means that $d_1 = d_2$ - a quantity known as the SVM's margin. In order to orientate the hyperplane to be as far from the Support Vectors as possible, we need to maximize this margin. The margin is equal to $\frac{1}{\|w\|}$ and maximizing it subject to the constraint in (4.2) is equivalent to finding Objective function

$$\min \|w\|$$

such that

$$y_i(x_i \cdot w + b) - 1 \geq 0 \forall i$$

It is possible to perform Quadratic programming (QP)

Objection function

$$\min \frac{1}{2} \|w\|^2$$

subject to the constraint

$$y_i(x_i \cdot w + b) - 1 \geq 0 \forall i \quad (4.4)$$

5. Support Vector Machine for Regression

Classify new unseen examples x' into one of two categories $y' = \pm 1; i = 1, 2 \dots N$

$$y_i = x_i \cdot w + b = -1 \quad (5.1)$$

$$y_i \in \mathfrak{R}, x \in \mathfrak{R}^D$$



6. Regression with ϵ -insensitive region

The regression SVM will use a penalty function than before not allocating a penalty. If the predicted value y_i is less than a distance ϵ away from the actual value t_i , if $|t_i - y_i| < \epsilon$. In figure 3, represents the region bound by $y_i \pm \epsilon$ for all i is called ϵ -insensitive region [9–11].

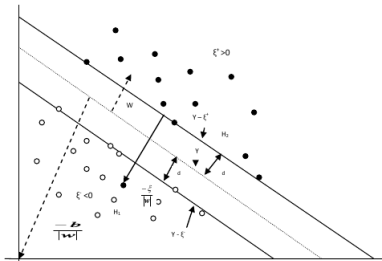


Figure 3. Regression analyses with ϵ -insensitive region

By modifying to the penalty function is that output variables which are lies outside the region are given one of two slack variable penalties depending on whether they lay above ξ^+ or below ξ^- the region where $\xi^+ > 0$ and $\xi^- < 0$ for all i

$$r_i \leq y_i + \epsilon + \xi^+ \quad (6.1)$$

$$r_i \geq y_i - \epsilon - \xi^- \quad (6.2)$$

SVM regression error function

$$C \sum_{i=1}^N (\xi_i^+ + \xi_i^-) + \frac{1}{2} \|w\|^2 \quad (6.3)$$

Minimize the subject to the constraints

$$\xi^+ \geq 0, \xi^- \geq 0$$

7. Algorithm for Classification and Regression

Classification	Regression
Create H , $H_{ij} = y_i y_j \phi(x_i) \cdot (x_j)$	
Misclassification: Selecting a suitable value for the parameter C	Misclassification: Large the insensitive loss region by selecting the parameters values C and ϵ
Find α , Maximized $\sum_{i=1}^N \alpha_i - \frac{1}{2} \alpha^T H \alpha$ Subject to the constraints $0 \leq \alpha_i \leq C$ $\sum_{i=1}^N \alpha_i y_i = 0$	Find α^+ and α^- maximized $\sum_{i=1}^N (\alpha_i^+ - \alpha_i^-) r_i - \epsilon \sum_{i=1}^N (\alpha_i^+ - \alpha_i^-) - \frac{1}{2} \sum_{i,j} (\alpha_i^+ - \alpha_i^-) (\alpha_j^+ - \alpha_j^-) \phi(x_i) \cdot (x_j)$ subject to constraint $0 \leq \alpha_i^+ \leq C; 0 \leq \alpha_i^- \leq C$ $\sum_{i=1}^N (\alpha_i^+ - \alpha_i^-) = 0$
Calculate $w = \sum_{i=1}^N \alpha_i y_i \phi(x_i)$	Calculate $w = \sum_{i=1}^N [\alpha_i^+ (\phi(x_i)) - \alpha_i^- (\phi(x_i))]$
Evaluate the support vectors sets S $0 \leq \alpha_i \leq C$	Evaluate the support vectors sets S $0 \leq \alpha_i \leq C$ and $\xi_i = 0$
Calculate $b = \frac{1}{N_s} \sum_{s \in S} (y_s - \sum_{n \in S} \alpha_n y_n \phi(X_n) \cdot \phi(X_s))$	Calculate $b = \frac{1}{N_s} \sum_{s \in S} [r_i - \epsilon - \sum_{m \in S} (\alpha_i^+ - \alpha_i^-) x_i \cdot x_m]$
Each new point x' is classified by evaluating $y' = \text{sgn}(w \cdot \phi(x') + b)$	Each new Point x' is classified by evaluating $= \sum_{i=1}^N [(\alpha_i^+ - \alpha_i^-) (\phi(x_i) \cdot \phi(x'))] + b$

8. Three Components of the Generalization Error

- The approximation error measures, by exact solution can be approximated by a function implementable by our learning system,
- The estimation error measures, by accurately to determine the best function implementable by our learning system using a finite training set instead of the unseen testing data.
- The optimization error measures, by closely we can



compute the function that best satisfies the given information in our finite training set [12].

9. Comparisons based on Feature selection

Table 1. Classification accuracy for NB vs KNN vs SVM

Feature	NB	KNN	SVM
10	49.69 ± 0.13	58.3 ± 0.23	61.79 ± 0.19
50	60.42 ± 0.15	74.22 ± 0.15	79.09 ± 0.17
100	77.9 ± 0.19	80.8 ± 0.09	85.58 ± 0.11
500	80.8 ± 0.13	82.17 ± 0.9	86.60 ± 0.1
1000	80.89 ± 0.11	82.19 ± 0.09	86.37 ± 0.09
5000	79.25 ± 0.06	82.98 ± 0.06	91.59 ± 0.05

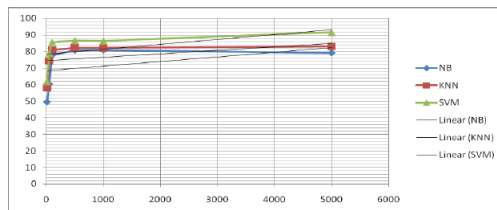


Figure 4. Classification accuracy

SVM obtains the best performance. Accuracy is improved with an increase in the numbers of features until some level.

10. Conclusion

Support Vector Machines acts as one of the best approach to data modelling and data mining. They combine generalization control as a technique and dimensionality. In classification problems generalization control is obtained by maximizing the margin, which corresponds to minimization of the weight vector in a standard framework. The minimization of the weight vector can be used as a criterion in regression problems, with a modified loss function. The generalization error is dominated by the approximation and estimation errors. The optimization error can be reduced to insignificant levels since the computation time is laborious. The main driver for the evolution of SVM solvers is that can quickly reduce the optimization error comfortably below the expected approximation. Mat Lab Tools can be used for Support Vector (SV) machines for function estimation these algorithms typically reduce the learning step memory constraint and accuracy of optimization variables, practitioners should choose different optimization methods. The main advantage of scaling is to avoid attributes in greater numerical ranges dominating those in smaller numerical ranges and also avoid numerical difficulties during the calculation.

References

- [1] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, *Data Min. Knowl. Discov.*, 2 (1998), 121–167.
- [2] H.H. Zhang, Variable selection for support vector machines via smoothing spline ANOVA, *Statist. Sinica*, 16 (2006), 659–674.
- [3] C.M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer (2006).
- [4] K. Fukumizu, F.R. Bach and M.I. Jordan, *Kernel dimension reduction for regression*, Technical report, Dept. Statistics, Univ. California, Berkeley (2006)
- [5] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, LIBLINEAR: A library for large linear classification, *Journal of Machine Learning Research*, 9 (2008), 1871–1874.
- [6] Ming Huang and Vojislav Kecman. Gene extraction for cancer diagnosis by support vector machines, *Artificial Intelligence in Medicine*, 35 (2005), 185–194.
- [7] Ming Zeng, Yu Yang and Junsheng Cheng, A generalized Gilbert algorithm and an improved MIES for one-class support vector machine, *Knowledge-Based Systems* 90 (2015), 211–223
- [8] Van Khoa Le and Pierre Beauleroy, Path for Kernel Adaptive One-Class Support Vector Machine, *IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (2015), 503–508.
- [9] Yang Liu and Gareth Pender, A flood inundation modelling using v-support vector machine regression model, *Engineering Applications of Artificial Intelligence*, 46 (2015), 223–231.
- [10] Andreas Christmann and Ding-Xuan Zhou, Learning rates for the risk of kernel-based quantile regression estimators in additive models, *Analysis and Applications* (2016).
- [11] Shuguang He, Wei Jiang and Houtao Deng, A distance-based control chart for monitoring multivariate processes using support vector machines, *Annals of Operations Research* (2016).
- [12] Yanyan Chen, Kuaini Wang and Ping Zhong, One-Class Support Tensor Machine, *Knowledge-Based Systems*, 96 (2016), 14–28.

