International Journal of Advances in Engineering Research http://www.ijaer.com

(IJAER) 2018, Vol. No. 16, Issue No. I, July **e-ISSN: 2231-5152, p-ISSN: 2454-1796**

MACHINE LEARNING AND SUPPORT VECTOR MACHINES (SVM): DEVELOPING PERFORMANCE GAUGE EFFICACY OF STANDARD SVM AND NU-SVM

Tejasdeep Singh Sahdev

Heritage School, Rohini, New Delhi

ABSTRACT

Analyzed hypothetically, v-SVM was observed to be over-subject to each preparation test, regardless of whether the examples have same esteem. This reliance would result in more opportunity for preparing, more help vectors and more choice time. With the end goal to defeat this issue, we propose another v-SVM. This new v-SVM increases each slack variable in the target work by a weight factor and consequently processes each weight actor by the quantity of comparing tests with the same incentive before preparing. Hypothetical investigation and the aftereffects of trials demonstrate that the new v-SVM has a similar order accuracy rate as the standard v-SVM and the new v-SVM is quicker than the v-SVM in preparing and choice if the preparation sets have same esteem tests.

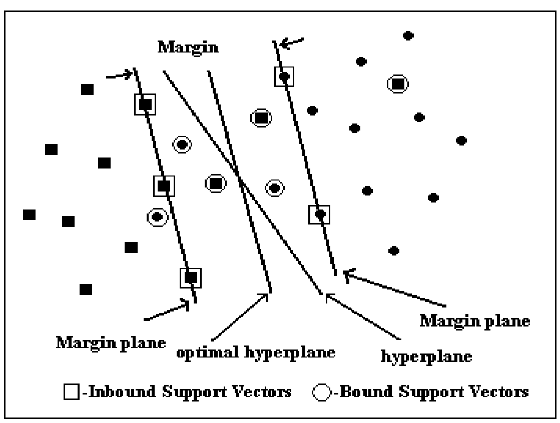

Figure 1. Principles of Support Vector Machines

ν-SUPPORT VECTOR MACHINE

Given $Z = \{(x_i, y_i) : x_i \in R^N, y_i \in \{+1, -1\}, i = 1, \dots, m\}$ an arrangement of preparing tests, where each information vector, is the mark of the class that Xi has a place with. With the end goal to look for the ideal hyperplane that best isolates the two classes from one another with the most stretched out edge, we have to take care of the accompanying advancement issue:

$$\min \tau(w, \xi, \rho) = \frac{1}{2} w^T w - \nu\rho + \frac{1}{m} \sum_{i=1}^{m} \xi_i \, , \qquad (1)$$

$$\text{s.t} \quad y_i (w^T x + b) \geq \rho - \xi_i \, , \qquad (2)$$

$$\text{and} \quad \xi_i \geq 0, i = 1, \dots, m \, , \qquad (3)$$

$$\text{and} \quad \rho \geq 0 \, . \qquad (4)$$

By Introducing Lagrange multipliers α_i, β_i, δ,
We have:

$$L(w, \xi, b, \rho, \alpha, \beta, \delta) = \frac{1}{2} w^T w - \nu\rho + \frac{1}{m} \sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} (\alpha_i (y_i (w^T \phi(x_i) + b) - \rho + \xi_i) + \beta_i \xi_i) - \delta\rho. \qquad (5)$$

The corresponding dual Lagrangian is:

$$\max W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j k(x_i, x_j), \qquad (6)$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq \frac{1}{m}, i = 1, \dots, m \, , \qquad (7)$$

$$\text{and} \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \qquad (8)$$

$$\text{and} \quad \sum_{i=1}^{m} \alpha_i \geq \nu \, . \qquad (9)$$

By solving the above dual Lagrangian, we obtain:

$$w = \sum_{i=1}^{m} \alpha_i y_i \phi(x_i) \, . \qquad (10)$$

The resulting decisionfunction can be shown as:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i k(x_i, x) + b\right) . \qquad (11)$$

Analysis of ν-SVM

For ν-SVM, the Karush-Kuhn-Tucker conditions of the primal problem ((1)-(4)) can be stated

$$\alpha_i (w^T \phi(x_i) + b - \rho + \xi_i) = 0 \, , \qquad (12)$$

$$\beta_i \xi_i = \left(\frac{1}{m} - \alpha_i\right) \xi_i = 0 \, , \qquad (13)$$

$$\delta\rho = 0 \, . \qquad (14)$$

Therefore, there are four cases as follows:

1) If , according to (13), we obtain 0iα=ix 0iξ=. In this case, is correctly classified.

2) If , then and 01/imα<<x (())Tiiiywxbφρξ+−+=0iξ=. In this case, ilies on the margin plane and ixis called an in-bound support vector 0.

3) In most cases, ρ
is greater than zero, according to (14), we know. This results in constraint (9) reducing to an equality condition

$$\sum_{i=1}^{m} \alpha_i = \nu \qquad (15)$$

Suppose $N_{BSV+}$and $N_{BSV-}$are the number of bound support vectors in the positive class and the negative class respectively, and N_{SV+}andN_{sv-} are the number of all kind of support vectors in the positive class and the negative class respectively, and $m_+$and m- are the number of data points in the positive class and the negative class respectively, we have:

$$\frac{2N_{BSV+}}{m} \le \nu \le \frac{2N_{SV+}}{m}. \qquad (16)$$

$$\frac{2N_{BSV-}}{m} \le \nu \le \frac{2N_{SV-}}{m}. \qquad (17)$$

Multiplying (16), (17) by and respectively, and substituting for ,

$$\frac{N_{BSV+}}{m_+} \le \frac{m_+ + m_-}{2m_+}\nu \le \frac{N_{SV+}}{m_+}. \qquad (18)$$

$$\frac{N_{BSV-}}{m_-} \le \frac{m_+ + m_-}{2m_-}\nu \le \frac{N_{SV-}}{m_-}. \qquad (19)$$

From (18) and (19), we know:

1) Omitting a data point, even if it has the same value as another point, would increase the upper bound of the fraction of bound support vectors and the lower bound of the fraction of support vectors of the class that it belongs to. In other words, omitting a data point would change the optimal hyperplane.

Besides the above characteristics, we can know intuitively that ν-SVM has another characteristic as follows:

2) If the training sets have same value samples, there would be same support vectors. In this case, decision is slower.

A NEW ν-SUPPORT VECTOR MACHINE

Description of the new ν-SVM

We have known from the last section that ν-SVM would produce same support vectors in the event that the preparation sets have the same esteem tests, so we propose another ν-SVM to take care of this issue and diminishing the time required for preparing and choice.

The primal problem in the new ν-SVM is

$$\min \tau(w,\xi,\rho)=\frac{1}{2}w^T w-\frac{m-n}{m}\nu\rho+\frac{1}{m-n}\sum_{i=1}^{m-n}d_i\xi_i\,, \qquad (20)$$

$$\text{s.t} \quad y_i(w^T x+b)\geq \rho-\xi_i\,, \qquad (21)$$

$$\text{and} \quad \xi_i\geq 0, i=1,\ldots,m-n\,, \qquad (22)$$

$$\text{and} \quad \rho\geq 0\,. \qquad (23)$$

In above formulations, m is the quantity of preparing tests, n is the quantity of preparing tests every one of which has an indistinguishable incentive from an example in the m-n tests. That is to say, m-n is the number of training samples each of which has a unique value d_i is the duplicate factor of sample X_i, and d_i>=1 The term $d_i\xi_i$in (20) is the error loss resulted from misclassifying x_i

By Introducing Lagrange multipliers α,βandδ , we have:

$$L(w,\xi,b,\rho,\alpha,\beta,\delta)=\frac{1}{2}\|w\|^2-\frac{m}{m-n}\nu\rho+\frac{1}{m-n}\sum_{i=1}^{m-n}d_i\xi_i-\sum_{i=1}^{m-n}(\alpha_i(y_i(w^T\phi(x_i)+b)-\rho+\xi_i)+\beta_i\xi_i)-\delta\rho \qquad (24)$$

With the same method as that in ν-SVM, we obtain dual Lagrangian

$$\max W(\alpha)=-\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j k(x_i,x_j), \qquad (25)$$

$$\text{s.t.} \quad 0\leq\alpha_i\leq\frac{1}{m-n}d_i, i=1,\ldots,m-n\,, \qquad (26)$$

$$\text{and} \quad \sum_{i=1}^{l}\alpha_i y_i=0\,, \qquad (27)$$

$$\text{and} \quad \sum_{i=1}^{l}\alpha_i\geq\frac{m}{m-n}\nu\,. \qquad (28)$$

The decision function of the new ν-SVM is

$$f(x)=\text{sgn}(\sum_{i=1}^{m-n}y_i\alpha_i k(x_i,x)+b)\,. \qquad (29)$$

Computation of n and the weights d_i

At First, n is initialized to zero. After reading a training sampleix, the new ν-SVM searches if ix is the same as any other samples which have been read and stored in the samples list. If there is a same sample $x_j$which has been stored in the samples list, then n is increased by 1, the corresponding $d_j$is increased by 1 and ix is ignored, otherwise ix is stored in the sample list and is set to 1. After all samples are read, all samples in the samples list are not duplicate.

Computation complex analysis of the new ν-SVM

Suppose the size of training set is m, and n samples in the set are the same as others. In other words, each of these n samples has the same value as a sample among the m-n samples. And suppose SMO algorithm[8] is used to solve the dual Lagrangian of ν-SVM and that of the new ν-

SVM. In training phase, the time complexity of ν-SVM and the new ν-SVM is $O(m^3)$ and $O(n^2+(m-n)^3)$ respectively. Therefore, the more same value samples in the training set, the more time saved if the new ν-SVM instead of ν-SVM is used. From (11) and (29), we can know that if ν-SVM has duplicate support vectors, then new ν-SVM is faster than ν-SVM in decision.

The space complexity of the new ν-SVM is larger than that of ν-SVM because it has to store the duplicate factors.

Bounds of the new ν-SVM

With the same method as in section 2.2, we obtain:

$$\frac{1}{m-n}\sum_{x_i \in BSV} d_i \le \frac{m}{m-n}\nu \le \frac{1}{m-n}\sum_{x_i \in SV} d_i \,. \qquad (30)$$

Multiplying (30) by $\frac{m-n}{m}$, we obtain:

$$\frac{\sum_{x_i \in BSV} d_i}{m} \le \nu \le \frac{\sum_{x_i \in SV} d_i}{m} \,. \qquad (31)$$

Because $\sum_{x_i \in BSV} d_i$ equals the number of bound support

$$\frac{\sum_{x_i \in BSV+} d_i}{m+} \le \frac{m_+ + m_-}{2m_+}\nu \le \frac{\sum_{x_i \in SV+} d_i}{m+} \,. \qquad (32)$$

$$\frac{\sum_{x_i \in BSV-} d_i}{m-} \le \frac{m_+ + m_-}{2m_-}\nu \le \frac{\sum_{x_i \in SV-} d_i}{m-} \,. \qquad (33)$$

the number of bound support vectors in the positive class, the number of support vectors in the positive class, the number of bound support vectors in the negative class, the number of support vectors in the negative class respectively, so the upper bound of the fraction of bound support vectors and the lower bound of support vectors in each class are not changed.

EXPERIMENTS

In this segment, we utilize some examination results to think about ν-SVM and the new ν-SVM exhibited in this paper. Examinations on ν-SVM were finished with libsvm[9], and tests on the new ν-SVM were finished with another program dependent on lib svm. Preparing tests and testing tests are every one of the two dimensional.

Consequences of preparing tests to appear in Table 1 and Results of testing experiments are shown in Table 2.

Table 1. Results of Training Experiments

ID	*A*	*B*	*ν-SVM*			*New ν-SVM*	
			C	*D*	*E*	*C*	*D*
1	*1000*	*614*	*0.5*	*957*	*572*	*0.28*	*385*
2	*2000*	*805*	*2.8*	*200*	*43*	*1.68*	*157*
3	*3000*	*124*	*6.6*	*138*	*24*	*7.6*	*114*
4	*4000*	*383*	*14.0*	*185*	*35*	*9.5*	*150*

In Table 1, Column A, B, C, D and E denote the number of training samples, the number of duplicate samples, time(seconds) for training, the number of support vectors and the number of duplicate support vectors, respectively.

Table 2. Results of Test Experiment s

ID	*A*	*B*	*C*
1	*50*	*40.6*	*30.7*
2	*50*	*20.3*	*17.2*
3	*50*	*16.5*	*14.8*
4	*50*	*18.9*	*17.1*

In Table 2, Column A, B and C represents test samples, time (milliseconds) for decision with ν-SVM and time (milliseconds) for decision with the new ν-SVM, respectively. Experiments in Table2 use the model generated by corresponding experiments in Table 1.
From experiment's output, we can infer that the new ν-SVM is faster than ν-SVM in both preparing and choice if the preparation set has to be copy and tests.

CONCLUSIONS

e have analyzed the characteristics of ν-SVM and presented a new ν-SVM. Results of theoretical analysis have shown that ν-SVM was over-subject to each preparation test. The new ν-SVM proposed in this paper uses a weight factor to represent the duplicate times of a training sample and compute the weight factor automatically. The new ν-SVM is faster than ν-SVM in training if the training sets have same-value samples, and faster in decision if the results of training of ν-SVM have same-value support vectors. The new ν-SVM has the same classification precision as ν-SVM.