

Support Vector Machine to Monitor Greenhouse Plant with Gaussian Loss Function

Manfu Yan¹, Qing Zhang¹, and Jianhang Zhang²

¹Department of Mathematics, Tangshan Teachers' College, Tangshan Hebei 063000, China
3005@tstc.edu.cn

²Faculty of Engineering, National University of Singapore, Singapore 119278

Abstract. In this paper, it applies Gaussian loss function instead of ε -insensitive loss function in a standard SVRM to devise a new model and a new type of support vector classification machine whose optimization problem is easier to solve and has conducted effective test on open data set in order to apply the new algorithm to environment monitoring in greenhouse plants and the monitoring result is better than any other method available.

Keywords: Support Vector Machine, Gaussian Loss Function, Greenhouse Plants, Environment Monitoring.

1 Introduction

It can be concluded that, the classification problem is a special type of the regression problem from the mathematics language description [1]. Therefore, it is feasible to create classification algorithm by Support Vector Machine (SVM). A general Support Vector Regression Machine uses ε -insensitive loss function to create classification algorithm. However, the solution of the problem is very difficult to get in this method. As a result, by substituting Gaussian loss function for the ε -insensitive loss function, the dual problem is derived. After some simplification and transformation of equations, the resultant optimization problem is easy to solve. At last, this new method is applied to Iris open data set, in order to do the data collection and algorithm verification. In a word, the new algorithm can be applied to modern agriculture and solve practical problems, and it is proved to be achievable and effective.

2 SVR Model of Solving Classification Problems

A classification problem is that the training set is given as,

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (\mathcal{X} \times \mathcal{Y})^l \quad (1)$$

here $x_i \in \mathcal{X} = \mathbb{R}^n$, $y_i \in \mathcal{Y} = \{1, -1\}$, $i = 1, \dots, l$, then it could figure out one real-valued function $g(x)$ in order to deduce the corresponding value y under any mode x by using decision function

$$f(x) = \text{sgn}(g(x)) \tag{2}$$

By comparing the definition with that of classification problems, the latter can be considered as a kind of special regression problem, thus it is able to solve such problems by applying Support Vector Regression Machine.

2.1 The Original Optimization Problem and the Dual Problem

Now, try to consider the classification problem as regression problem since y_i takes value from $\{1, -1\}$, and instead of ε -insensitive loss function, Gaussian loss function is selected, the form of original optimization problem is

$$\min_{w, \xi, b} \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^l \xi_i^2 \tag{3}$$

$$s.t \quad ((w \cdot x_i) + b) - y_i \leq \xi_i, i = 1, 2, \dots, l \tag{4}$$

$$y_i - ((w \cdot x_i) + b) \leq \xi_i, i = 1, 2, \dots, l \tag{5}$$

$$\xi_i \geq 0, i = 1, 2, \dots, l \tag{6}$$

Problem (3)-(6) equal to

$$\min_{w, \xi, b} \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^l \xi_i^2 \tag{7}$$

$$s.t \quad \xi_i \geq |(w \cdot x_i) + b - y_i|, i = 1, 2, \dots, l \tag{8}$$

Apparently constraints in the problem can be formulated as equalities, then problem (7) and (8) equal to

$$\min_{w, b} \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^l (y_i((w \cdot x_i) + b) - 1)^2 \tag{9}$$

Let

$$\eta_i = 1 - y_i((w \cdot x_i) + b) \tag{10}$$

then the above problems can be expressed as

$$\min_{w, \xi, b} \quad \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^l \eta_i^2 \tag{11}$$

$$s.t \quad y_i((w \cdot x_i) + b) = 1 - \eta_i, i = 1, 2, \dots, l \tag{12}$$

The resulting problem is exactly the original optimization problem in the Least Squares Support Vector Machine.

Next, discuss properties of the resulting problem and its dual problem, and then create an algorithm based on them.

Theorem 1. The Dual problem of problem (11)-(12) is

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j ((x_i \cdot x_j) + \frac{\delta_{ij}}{C}) - \sum_{i=1}^l \alpha_i \tag{13}$$

$$s.t \quad \sum_{i=1}^l \alpha_i y_i = 0 \tag{14}$$

where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{15}$$

Proof. Introducing the Lagrange function of problems (11)-(12)

$$L(w, b, \eta, \alpha) = \frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^l \eta_i^2 - \sum_{i=1}^l \alpha_i (y_i ((w \cdot x_i) + b) + \eta_i - 1) \tag{16}$$

where $\alpha \in R^l$ is the Lagrange multiplier vector, find the minimum of Lagrange function with respect to w, b, η , get the following KKT condition:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \tag{17}$$

$$\sum_{i=1}^l y_i \alpha_i = 0 \tag{18}$$

$$\eta = \frac{\alpha}{C} \tag{19}$$

$$y_i ((w \cdot x_i) + b) + \eta_i - 1 = 0, i = 1, 2, \dots, l \tag{20}$$

Substitute the above conditions into the Lagrange function and find the maximum of α , the dual problem (13) and (14) are obtained.

The several theorems below are about relations between solution of original problem (11)-(12) and that of dual problem (13)-(14) are all established:

Theorem 2. The solution (w^*, b^*, η^*) of original problem (11)-(12) exists and the solution is unique.

Theorem 3. Suppose (w^*, b^*, η^*) is the solution of original problem(11)-(12), then dual problem(13)-(14) must have solution $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ to satisfy

$$w^* = \sum_{i=1}^l \alpha_i^* y_i x_i \tag{21}$$

Proof. Concluded from Theorem 1 and Wolfe Theorem, if (w^*, b^*, η^*) is the solution of original problem (11)-(12), and the dual problem(13)-(14) must have solution which satisfies equation (21).

Theorem 4. Suppose $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ is an arbitrarily solution of dual problems(13)-(14), then the solution to (w, b) of original problem (11)-(12) exists and must be unique,

$$w^* = \sum_{i=1}^l \alpha_i^* y_i x_i \tag{22}$$

$$b^* = y_i (1 - \frac{\alpha_i^*}{C}) - \sum_{j=1}^l \alpha_j^* y_j (x_j \cdot x_i) \tag{23}$$

Its proof can be seen in [2]

2.2 The SVR Algorithm of Solving Classification Problems

For general nonlinear problems, put the input space R^n into a single mapping $\Phi(\cdot)$, which can transform it to a high-dimensional Hilbert space. In this space, the original optimization problem is constructed and its dual problem is obtained.

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j (\Phi(x_i) \cdot \Phi(x_j) + \frac{\delta_{ij}}{C}) - \sum_{i=1}^l a_i \tag{24}$$

$$s.t \quad \sum_{i=1}^l a_i y_i = 0 \tag{25}$$

Instead of the dual problem of the inner product $(\Phi(x_i) \cdot \Phi(x_j))$, the kernel function $K(x_i, x_j)$ is introduced, then the dual problem becomes,

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j (K(x_i, x_j) + \frac{\delta_{ij}}{C}) - \sum_{i=1}^l a_i \tag{26}$$

$$s.t \quad \sum_{i=1}^l a_i y_i = 0 \tag{27}$$

For $K(x_i, x_j) + \frac{\delta_{ij}}{C}$ in the objective function, it can be represented by a kernel function

$$\hat{K}(x_i, x_j) = K(x_i, x_j) + \frac{\delta_{ij}}{C} \tag{28}$$

In Hilbert space, Theorem 2-4 are still hold for the relationship between the solution of dual problem and that of the original problem, then the formula of the solution to b^* becomes

$$b^* = y_i (1 - \frac{\alpha_i^*}{C}) - \sum_{i=1}^l \alpha_i^* y_i K(x_j, x_i) \tag{29}$$

According to theorem 4, the following algorithm is established:

Algorithm 1. The SVR Algorithm for Solving Classification Problems

(i) Assume a known training set

$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (\mathcal{X} \times \mathcal{Y})^l$, here, $x_i \in \mathcal{X} = R^n, y_i \in \mathcal{Y} = \{-1, 1\}, i = 1, 2, \dots, l$

(ii) Choose a suitable positive C and a kernel $K(x, x')$;

(iii) Construct the problem and find the solution of

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j (K(x_i, x_j) + \frac{\delta_{ij}}{C}) - \sum_{i=1}^l a_i \quad (30)$$

$$s.t \quad \sum_{i=1}^l a_i y_i = 0 \quad (31)$$

The optimum solution is obtained $a^* = (a_1^*, \dots, a_l^*)^T$

(iv) To create decisive function

$$f(x) = \text{sgn}(\sum_{i=1}^l a_i^* y_i K(x_i, x) + b^*) \quad (32)$$

Here b^* is given by equation (29).

2.3 The Numerical Experiments

In order to verify the proposed Algorithm 1, a test was conducted on Iris data set [3]. The Iris data set is used to test the performance of classification algorithms. The data set contains the number of 150 sample points, which are divided into three categories, namely, I(Iris-setosa), II(Iris-versicolor) and III(Iris-virginica), there are 50 sample points in each type and each sample point has for properties.

There are three two-class classification problems, namely, Class I and II is the positive class, and Class III is the negative class; or Class I and III is the positive class, and Class II is the negative class; or Class II and III is the positive class, and Class I is the negative class. In each of the two-class classification problems there are 150 sample points, which has been randomly assigned to training set and testing set. The training set contains 50 positive points and 25 negative points, while the testing set contains 50 positive points and 25 negative points. The trainings are conducted by using Algorithm 1 and standard C-SVM. During the training process, the RBF Kernel function is adopted for the two algorithms. The parameter c is set to be 0.1, 1, 10, 100, 1000, 10000 and so on. The decisive functions gained in each training session are tested, and each testing results are recorded. Finally compute and compare the average testing accuracy, result is shown in the following table:

Table 1. Result Comparison Table

Classification	<i>c-svc</i>	Algorithm 1
{I, II}—III	95.6%	96.1%
{II, III}—I	100%	100%
{I, III}—II	97.5%	97.2%

From the above comparison results, obviously Algorithm 1 and *c-svc* share the similar testing accuracy rate.

3 The Application of SVR Classification Algorithm to the Environment Monitoring on the Greenhouse Plant Growth

The great progress in technology has brought a serious problem to the traditional agriculture, which is far from meeting the needs of the modern social development [4]. Therefore, improvement and revolution must be done to the traditional agriculture. A new cultivation method is developed through many years' experience, which is that people can control environmental factors so that the crops can grow in the most suitable environment. In addition, the growing seasons may be extended and the best output is gained. This agricultural mode is known as the greenhouse agriculture, or as factorial agriculture and greenhouse agriculture in developed countries. With its striking feature of being free from environmental constraint, the new agricultural mode enables the crops to grow under some pre-designed conditions is highly yielding and greatly effective. Hence it has been a trend all over the world. In a word, the research on environment monitoring is crucial, especially for the real-time environment monitoring.

3.1 Objective Conditions and Advantages of SVM-Based Environment Monitoring of Greenhouse Plants

The environmental factors for soilless greenhouse cultivation include temperature, humidity, CO₂, illumination, EC, and nutrition elements. After discussing with agricultural experts, choosing *l* greenhouses to be measured with *n* factors as the measurement parameter. The *i*-th parameter is marked as $[x]_i, i=1,2,\dots,n$, then $x=([x]_1, \dots, [x]_n)^T$ here, $x_i \in R^n$, and normal greenhouse is recorded as 1, otherwise -1, then let

$$T = \{ (x_1, y_1), \dots, (x_l, y_l) \}$$

Here $[x]_i \in R^n, i=1,2,\dots,n, y_i = 1, or -1$

It is just a trainingset of Support Vector Classification Machine. Thus the problem of environmental monitoring could be solved by SVM. (see 3.2).

Nowadays, the normal data ranges for the environment monitoring of greenhouse plants are given. In practice, adjusting those parameters based on the standard. However, it is found that although all the data are in range, the problem is still come out

sometimes. It indicates that the interval control method has a large deviation and it is not reliable. For example, as temperature is adjusted, it may interfere with the humidity, so the environment control is not effective and sometimes make it worse. The advantage of SVM is that it do not have to set up the ranges for each parameter beforehand; instead, taking into account of all the parameters together, resulting in the much improved accuracy of monitoring. Assume the output is -1, which means the current environment is not the one wanted, so adjusting the corresponding parameters in the computer until the result is 1. This operation is obviously easy and effective with the aid of computer.

3.2 The Application to Environment Monitoring of Celery Cultivation

Three years' research on celery is conducted in Shouguang, Shandong and Fengnan, Hebei. 100 greenhouses are randomly taken as our sample and measured 15 parameters, which are suggested by agricultural experts. The parameters include temperature, humidity, illumination, and so on. A data set of 100 15-dimension vectors is obtained, at the same time, let agricultural experts determine whether the celery greenhouse is normal or not [5]. The 100 15-dimension vectors are given values of -1 and 1 according to the result is abnormal or not. In this way, the trainingset of the SVM is obtained.

3.2.1 Data Pre-retreatment

From the data, it is noticed that some parameters are in a narrow range, for example, the nitrogen concentration is between 0.004 and 0.01, but some parameters are in a wide range, like illumination is between 200 and 1200 luxes. Thus, standardization of those parameters is needed. The maximum-minimum standardized method is used. For example, nitrogen $[x_i]$ is between 0.004 and 0.01, so here uses maximum-minimum formula,

$$[x_i]' = \left([x_i] - \min_{j=1,2,\dots,100}([x_j]) \right) / \left(\max_{j=1,2,\dots,100}([x_j]) - \min_{j=1,2,\dots,100}([x_j]) \right) \quad (33)$$

Standardize the data set to D' .

The data set is randomly splitted into 2 subsets by the ratio of 7:3. One subset is trainingset T and the number of training point is l (here $l = 70$), the other subset is test set S , the test point is m (here, $m = 30$). The number of the positive points in training set T is T_+ , that for the negative points is T_- . Similarly, the number of the positive points in test set S is S_+ , the number of negative points is S_- . The number of positive points is 74, and the number of negative points is 26. There is an imbalance between these two kinds of points. Therefore, set up penalty parameters C_+ and C_- for both positive and negative points when using SVM. C_+ and C_- could be determined by the following formula,

$$C_+ = CT_+ / l, \quad C_- = CT_- / l, \quad (34)$$

Here $c > 0$, which has been given in advance.

3.2.2 The Choice of Model

Suitable models of SVM are needed to be chosen for the above classificatory problem. Three models were choose, the first one is weighed proximal SVM model [6], and the original problem is

$$\min_{w, \eta, b} \frac{1}{2} (\|w\|^2 + b^2) + \frac{1}{2} C_+ \sum_{y_i=1} \eta_i^2 + \frac{1}{2} C_- \sum_{y_i=-1} \eta_i^2 \tag{35}$$

$$\text{s.t. } y_i ((w \cdot x_i) + b) \geq 1 - \eta_i, i = 1, \dots, l, \tag{36}$$

and its dual problem is

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (K(x_i, x_j) + 1) + \frac{1}{2C_+} \sum_{y_i=1} \alpha_i^2 + \frac{1}{2C_-} \sum_{y_i=-1} \alpha_i^2 - \sum_{i=1}^l \alpha_i \tag{37}$$

The second model is support vector machine algorithm 1, the optimization problem to be solved is (30) - (31).

The third one is weighed standard SVM model [7]. The original problem is

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C_+ \sum_{y_i=1} \xi_i + C_- \sum_{y_i=-1} \xi_i \tag{38}$$

$$\text{s.t. } y_i ((w \cdot x_i) + b) \geq 1 - \xi_i, i = 1, \dots, l, \tag{39}$$

$$\xi_i \geq 0, i = 1, \dots, l, \tag{40}$$

and its dual problem is

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{j=1}^l \alpha_j \tag{41}$$

$$\text{s.t. } \sum_{i=1}^l \alpha_i y_i = 0 \tag{42}$$

$$0 \leq \alpha_i \leq C_+, \quad y_i = 1 \tag{43}$$

$$0 \leq \alpha_i \leq C_-, \quad y_i = -1 \tag{44}$$

Proper parameters are going to be chosen after defining the three above models. The parameters include kernel function $K(x, x')$ and C_+, C_- , and the parameters in kernel function as well. Here, the radial basis kernel function is choosen,

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right)$$

Now the parameters to be determined are C_+, C_- and σ .

For each model, the best parameters is chosen by the method of lattice. In other words, the ranges of C_+ and C_- , are $\{0.1,1,10,100,1000,10000\}$ and the ranges of σ , which is $\{0.1,0.2,0.5,1,2,5\}$. Thereby, they have constituted a group of parameters, (C_+, C_-, σ) . Loo deviations were calculated for each group of parameters [8]. The group with minimal Loo value is the group of best parameters $(C_+, C_-, \bar{\sigma})$. For weighed proximal SVM model and it is $(\bar{C}_+ = 10, \bar{\sigma} = 2)$; for Algorithm 1 its group of best parameters is $(\bar{C}_+ = 100, \bar{C}_- = 100, \bar{\sigma} = 5)$; for weighed standard SVM, it is $(\bar{C}_+ = 10, \bar{\sigma} = 1)$.

3.2.3 Comparison of Results

Substituting three groups of best parameters into the corresponding models, decisive functions are obtained, which can be used to decide the points in test set, as shown in the following table.

Table 2. Result Comparison Table

Checkup Results / Checkup Items \ Models	C-SVM	Algorithm 1	PSVC
Result Precision	90%	95%	86%
Percentage of False Report	3%	0%	12.5%
Percentage of Checkup Result	66.7%	83.3%	66.7%

The Result Precision is the ratio of items of those checked correctly to all the items in the sample test set; the percentage of False Report is the ratio of false reported items to the number of real usual items; Percentage of Checkup Result is the percentage of the found real false items in all the real items.

From the above experiment results, Algorithm 1 leads to the best result precision among three kinds of models.

4 Conclusion

Based on the above discussion, the classification problem is treated as a special regression problem, and the ε -insensitive loss function is substituted to Gaussian loss function, so that the optimization problem is easy to solve. This classification algorithm introduces a new way of solving classification problem, and the new algorithm has been applied to practical greenhouse plant environment monitoring. It not only solves the practical problem, but the effectiveness is varified and comparison with old method shows the advantages of the new method.

References

- [1] Deng, N., Tian, Y.: Support Vector Machine – Theory, Algorithm and Expansion, pp. 63–64. Science Press, Beijing (2009)
- [2] Yan, M.: Support Vector Machines for Classification and Its Application. China Agricultural University, Beijing (2005) (Doctor's Degree Paper)
- [3] Deng, N., Tian, Y.: Optimal Method of Data Processing – Support Vector Machine. Science Press, Beijing (2004)
- [4] Wang, X.: The Problems and solutions of Public Vegetable Base, vol. (3), pp. 1–3 (2010)
- [5] Zhang, D.: Celery. China Celery (1), 15 (2010)
- [6] Fung, G., Magansarian, O.L.: Proximal Support Vector Machine Classification. In: KDD 2001, San Francisco, CA USA (2001)
- [7] Yang, Z., Liu, G.: Principle and Application of Uncertain Support Vector Machine, pp. 148–151. Science Press, Beijing (2007)
- [8] Vapnik, V., Chapelle, O.: Bounds on Error Expectation for Support Vector Machines. Neural Computation 12(9) (2000)