Support Vector Machine Applied in the Classification of Fanjing Mountain’s Tea

Lina Yuan*, Huajun Chen and Jing Gong
College of Data Science, Tongren University, Tongren, China

*Corresponding author e-mail: 893422817@qq.com

Abstract. The problems of classification and regression are no longer state-of-the-art ones, however, with the popularization and extensive application of computers, especially, the rapid development of machine learning and data mining gives them fresh meaning. Nowadays, these problems become major problems that have triggered a broad range of research activities again. Due to its advantages, such as superior generalization performance, global convergence, sample dimension insensitivity, and so on, support vector machine (SVM) has made great progress in theory and application. This article compares SVM with BP neutral network to classify teas of Fanjing Mountains in view of accuracy and mean square error (MSE), and the simulation results show that the accurate rate of SVM is higher than that of BP neutral network, and the MSE of the former is lower than that of the latter.

1. Introduction
Support vector machine (SVM) is a universal learning machine, a novel technique in data mining, and a new tool for the problems of machine learning (ML) with the help of optimization methods, which is better to implement the idea of structural risk minimization (SRM). Theoretically, because of employing two planning optimization, the global optimal solution can be obtained, and the unavoidable problem of local minimum is solved. Thanks to utilizing kernel function, the problem of dimension is cleared up skillfully, making the complexity of the algorithm independent of the sample dimension, and it is very suitable for dealing with nonlinear problems. In addition, SVM applies the principle of SRM, thus, it has a great ability to extension. In recent years, SVM has found an increasingly wide utilization in various fields, especially in practical problems, for examples, image retrieval, the recognition of handwriting or 3-dimensional objectives, the classification of text, facial expression, ground cover, or microarray gene expression data, the prediction of traffic speed and travel time, and the diagnosis and prediction of breast cancer, which have achieved excellent effects. Therefore, SVM has attracted a great deal of attention lately, and become one of the fastest growing research directions in the early 21st Century. Accordingly, the study of theory and application of SVM has vital theoretical significance and application prospect, and is worth investigating.

Tea industry is a highly competitive and efficient industry in China’s agriculture, and one of the important agricultural exports of foreign exchange, which is the main component of rejuvenating the rural economy and increasing the income of farmers. Fanjing Mountains, located at the shared-border of Jiangkou County, Yinjiang County, and Songtao County, 2493 meters above sea level, and the main peak of the Wuling Mountains, has great regional area, and its forest cover rate is 95%. Furthermore, it
is one of the most unique landmarks in Guizhou, the Lingshan of the eastern on Qian, the kingdom of zoology, and the national natural protection zone, meantime, it is abundant to various kinds of teas, such as Fanjing Mountain’s tea, Green tea, Black tea, White tea, Dark tea, Oolong tea, Instant tea, and so on. This paper mainly researches SVM applied in the classification of Fanjing Mountains’ teas. The rest of this thesis is organized as follows. The four core ideas of SVM and multi-class SVMs are provided the details in Section II and Section III, respectively. Section IV presents simulation results for comparing the accuracy and mean square error (MSE) of SVM against BP neutral network. Finally, Section V concludes the paper.

2. The four core ideas of SVM

2.1. Maximum interval

In MLP, some patterns or knowledge are found from data, which is always required to consider the similarities among the data. For example, in the problem of classification, supposing the training set contains \( l \) samples: \((x_i, y_i), i = 1, 2, \ldots, l\), the input vector is \( x_i \in \mathbb{R}^n \), and the corresponding expectation output is \( y_i \in \{+1, -1\}, i = 1, 2, \ldots, l \), where +1 and -1 respectively represent the label identification of the two categories, and \( n \) is the input dimension. The goal of learning is to construct a decision function that classifies the test data as accurately as possible to infer the corresponding value of \( y \) for any pattern \( x \). An intuitive idea for solving this type of problem is to determine whether the new input \( x \) is similar to those of the positive class, or the negative class, thus to infer the attribution of \( x \). Then, as shown in Fig. 1 [6], how to select and divide the normal vector-\( \mathbf{w} \) of the straight line \( L \)? It is understood that for a given appropriate normal direction, there will be two extreme lines, and the distance between the two lines is called the normal direction of the corresponding "interval". As might be imagined, the normal direction to maximize the "interval" should be chosen. If \( \mathbf{u} \cdot \mathbf{w} \geq c \), then it is decided to be +. That is, if \( \mathbf{u} \cdot \mathbf{w} + b \geq 0 \), then +. For example,

\[
\mathbf{w} \cdot \mathbf{X}_+ + b \geq 1, \text{ and } \mathbf{w} \cdot \mathbf{X}_- + b \leq -1. \tag{1}
\]

![Figure 1. The geometric meaning of linear classification for SVM.](image)

2.2. Decision-making formula

Suppose:

\[
y_i = \begin{cases} +1, & \text{for } y : + \\ -1, & \text{for } y : - \end{cases}
\]

The formula (1) is constrained by \( y_i \cdot (\mathbf{w} \cdot x_i + b) - 1 \geq 0 \), while \( x_i \) is right on the dividing line, so there is
\[ y_i \cdot (w \cdot x_i + b) - 1 = 0. \] \hspace{1cm} (2)

### 2.3. Objective function

It is known that the product of the two vectors is the projection, the maximum "interval" is

\[ d_{\text{max}} = (X_+ - X_-) \cdot \frac{w}{\|w\|} = \frac{1}{\|w\|} \left( w \cdot X_+ - w \cdot X_- \right). \] \hspace{1cm} (3)

(2) and (3) are combined to obtain the following formula:

\[ d_{\text{max}} = \frac{1}{\|w\|} (1 - b - (-1 - b)) = \frac{2}{\|w\|}. \] \hspace{1cm} (4)

From the formula (4): \( d_{\text{max}} \) only has to do with the slope- \( -\frac{1}{\|w\|} \). Therefore, the maximum “interval”- \( d_{\text{max}} \) can be transformed into maximizing \( \frac{1}{\|w\|} \), while the maximum \( \frac{1}{\|w\|} \) can be also translated into minimizing \( \|w\| \). For the convenience of mathematical calculation, the minimum- \( \|w\| \) can be rendered into minimizing \( \frac{1}{2} \|w\|^2 \), and then the whole process can be expressed as

\[ \max \frac{2}{\|w\|} \iff \max \frac{1}{\|w\|} \iff \min \|w\| \iff \min \frac{1}{2} \|w\|^2. \] \hspace{1cm} (5)

### 2.4. Optimum theory

\[ L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[ y_i \cdot (w \cdot x_i + b) - 1 \right]. \] \hspace{1cm} (6)

Firstly, the derivation of \( w \) and \( b \) in (6) is respectively:

\[ \frac{\partial L}{\partial w} = w - \sum_{i=1}^{m} \alpha_i y_i x_i. \] \hspace{1cm} (7)

\[ \frac{\partial L}{\partial b} = -\sum_{i=1}^{m} \alpha_i y_i. \] \hspace{1cm} (8)

Secondly, (7) and (8) are set to be equal to 0 as follows.

\[ w = \sum_{i=1}^{m} \alpha_i y_i x_i. \] \hspace{1cm} (9)

\[ \sum_{i=1}^{m} \alpha_i y_i = 0. \] \hspace{1cm} (10)
Thirdly, to substitute (9) and (10) into (6):

\[ L = \frac{1}{2} \sum_{i=1}^{m} \alpha_i y_i x_i - \sum_{i=1}^{m} \alpha_i y_i \sum_{j=1}^{n} \alpha_j y_j x_{ij} - \sum_{i=1}^{m} \alpha_i y_i b + \sum_{i=1}^{m} \alpha_i \]

(11)

Then, the formula (11) can be rewritten as:

\[ L = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i y_i \alpha_j y_j x_{ij}. \]

(12)

Where, \( x_i \cdot x_j \) is inner product in (12). If \( \sum_{i=1}^{m} \alpha_i y_i x_i \cdot u + b \geq 0 \), then +. That is, it does not require to know what kind of \( x_i \), and only obtains the dot product between \( x_i \) and \( u \) that can decide its category.

3. Multi-class SVMs

3.1. Directed acyclic graph-based multi-class SVM classifier (DAGSVM)

DAGSVM constructs a learning structure with a directed acyclic graph, and according to this chooses and combines the two categories of classifiers in one-against-one. Similar to the training method of one-against-one, \( M(M-1)/2 \) classifiers are also required. During the testing phase, the algorithm selects each SVM classifier with a directed no-loop graph containing the root node. It has \( N = M(M-1)/2 \) internal nodes and \( M \) leaf nodes in the directed no-loop graph, and each node corresponds to a classifier. For a testing sample \( x \), starting from the root node (using the decision function to evaluate its value), it moves to the left or right depending on its output value (Fig. 2 shows four-class problems of directed no-loop graph).

3.2. Crammer-Singer (directly structure classifiers)

The multi-class SVMs of Crammer-Singer provide a direct method for solving \( M \)-class problems, which is two-class SVM. A form of natural extension of the quadratic programming model in the following:
\[
\min \varphi(w, \xi) = \frac{1}{2} \sum_{m=1}^{M} \|w_m\|^2 + C \sum_{i=1}^{l} \sum_{m \neq y_i} \xi_i^m
\] (13)

The constraint conditions are:

\[
\begin{aligned}
(w_{y_i}, \xi_i) + b_{y_i} &\geq (w_m, x_i) + b_m + 2 - \xi_i^m, \\
\xi_i^m &\geq 0, \ i = 1, \ldots, l, \ m, y_i \in \{1, \ldots, M\}, \ m \neq y_i
\end{aligned}
\] (14)

Thus, the decision function of the following \(M\)-class SVM classifier is as follows:

\[
f(x) = \arg \max \left[ (w_i \cdot x) + b_i \right], i = 1, 2, \ldots, M
\] (15)

Fig. 3 shows the SVM classifier of Crammer-Singer on the two-dimensional plane.

4. Simulation results

In this section, we assume that data is generated randomly, and respectively adopt the two methods of SVM and BP neutral network to classify thirteen teas of Fanjing Mountains, i.e., Fanjing Mountain’s tea, Green tea, Black tea, White tea, Dark tea, Oolong tea, Instant tea, Cuiying tea, Tribute tea, Ampelopsis, Maofeng tea, Spring tea, Dragon-well tea. Through simulation by MATLAB, it can be known that the accuracy of classification by SVM and BP neural network is 99.361% and 98.291%, separately, and their mean square error (MSE) 0.001366 and 0.0036992, respectively, as shown in Fig. 4. Therefore, the accurate rate of SVM is higher than that of BP neutral network, and the MSE of the former is lower than that of the latter.

5. Conclusion

Although SVM acts as a new kind of ML method, which has excellent qualities that many traditional ML methods do not have, there are still a lot of imperfections for SVM at present. Hence, how to further enhance and develop SVM applied in theory and practical application, is the goal that researchers have been pursuing.
Acknowledgments
This work was financially supported by the project of platform talent in Guizhou in 2016 (Project Number: [2016] 5611), education and cooperation for talent team word in Guizhou in 2015 (Project Number: [2015] 67), the Collaborative Fund Project of Science and Technology Agency in Guizhou Province Marked by the word LH on 7480 [2014], and partly supported by the National Natural Science Foundation of China (NO.61562703).

References
[1] Stephen Marsland, Machine Learning, An Algorithmic Perspective, Second Edition: CRC Press, Oct. 2014.
[2] A. S. Nugroho, A. B. Witarto and D. Handoko, Support Vector Machine: Springer US, 2016.
[3] Huang X L, Shi L, Suykens J A K, "Support vector machine classifier with pinball loss," IEEE transactions on Pattern Analysis and Machine Intelligence, 2014, 36 (5): 984-997.
[4] Wang K N, Zhu W X, Zhong p, "Robust support vector regression with generalized loss function and applications," Neural Processing Letters, 2015, 41: 89-106.
[5] M. Rasmussen, J. Rieger and K. N. Webster, "Approximation of reachable sets using optimal control and support vector machines," Journal of Computational and Applied Mathematics, 2017, 311: 68-83.
[6] H. Jiang, K. Huang, and R. Zhang, "Field support vector regression," in Proc. Int. Conf. Neural Inf. Process., Nov. 2017.
[7] P. Songsiri, T. Phetkaew and B. Kijsirikul, "Enhancement of multiclass support vector machine construction from binary learners using generalization performance," Neurocomputing, vol. 151, pp. 434-448, 2017.
[8] Jing Gong, "An analysis of big data based on the ecological, tourism resources of Fanjing Mountains," Database & Information Manage, 2017, 11: 69-71 (In Chinese).