



PAPER • OPEN ACCESS

The Development and Application of Support Vector Machine

To cite this article: Zhao Jun 2021 *J. Phys.: Conf. Ser.* **1748** 052006

View the [article online](#) for updates and enhancements.

You may also like

- [A Method of Particle Swarm Optimized SVM Hyper-spectral Remote Sensing Image Classification](#)
Q J Liu, L H Jing, L M Wang et al.
- [SVM bearing fault diagnosis based on fast inter-class distance in the feature space and DMPSO algorithm](#)
Renwang Song, Baiqian Yu, Lei Yang et al.
- [Support vector machines to detect physiological patterns for EEG and EMG-based human-computer interaction: a review](#)
L R Quitadamo, F Cavrini, L Sbernini et al.



The Electrochemical Society
Advancing solid state & electrochemical science & technology

247th ECS Meeting
Montréal, Canada
May 18-22, 2025
Palais des Congrès de Montréal

Showcase your science!

**Abstracts
due
December
6th**

The Development and Application of Support Vector Machine

Zhao Jun

University of Jinan, Shandong, China

zhaojun202109@163.com

Abstract—Support Vector Machine(SVM) algorithm has the advantages of complete theory, global optimization, strong adaptability, and good generalization ability because of it on the basis of Statistical Learning Theory's(SLT). It is a new hot spot in machine learning research. This article first systematically studies some basic concepts of SVM and the optimization of SVM. In addition, this article also discusses the application of SVM in modern machining, protein prediction and face detection. Through these applications, the performance characteristics and advantages of SVM can be reflected. At the end of the article, some shortcomings of SVM are introduced, and the development trend of subsequent SVM is pointed out accordingly.

1. Introduction

Traditional statistics[1] study the situation when the amount of samples tends to be infinite. However, the amount of samples in our daily lives is usually limited. Different from traditional statistics, Statistical Learning Theory(SLT) is a theory that specializes in studying the laws of machine learning in the case of small samples. SLT provides a new framework in dealing with the general learning problem. In 1995, Vapnik et al[2] developed a novel potent method—Support Vector Machine(SVM) on the basis of SLT. SVM has considerable unique merits in tackling small samples, nonlinear and high-dimensional pattern recognition.

Because of these merits, SVM aroused general interest in the matter of machine learning. This paper is aimed at helping readers understand the advantages and disadvantages of SVM by introducing some basic concepts of SVM and its applications in various fields, so as to predict the future development direction of SVM.

2. Basic concept

In this part, some questions about classification will be raised. With the help of these questions, some concepts of SVM will be introduced, like hard margin, soft margin and kernel function. After understanding these concepts, we will know how SVM is produced and how SVM classifies data better.

2.1. Hard Margin

Here are some samples like :

$$D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_m, y_m)\}$$

where y_i equals to -1 or 1. There are lots of hyperplanes can separate these training samples in the figure 1. But which hyperplane is the best to separate these samples?



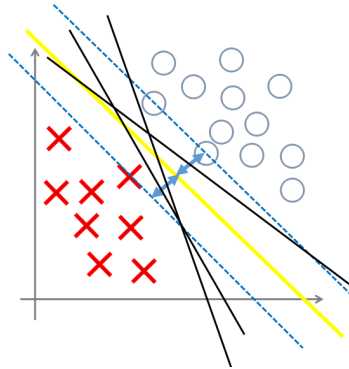


Figure1. Hyperplans that separate training samples

In figure1, it is clear that the yellow one is the best hyperplane, for this hyperplane is the most tolerant to the local disturbance of the training sample. In sample space, we can use $w^T x + b = 0$ (1) to express the hyperplane.

Therefore, we can use (2) to calculate the distance of every point in the sample to the hyperplane.

$$r = \frac{|w^T x + b|}{\|w\|} \quad (2)$$

If a specific hyperplane can correctly separate these samples, we will get:

$$\begin{cases} w^T x_i + b \geq 0 & y_i = +1 \\ w^T x_i + b \leq 0 & y_i = -1 \end{cases} \quad (3)$$

As we can see in the figure 1, there are some points just falling on the boundary, and they are what we call support vectors. They will make the equal sign in formula (3) hold. Margin is the sum of the distances between the support vectors on the two boundary edges and the hyperplane. For example, in figure 1, the margin is equal to the sum of the distance from an 'X' on the boundary to the hyperplane plus an 'O' on the other boundary to the hyperplane. It can be expressed as:

$$r = \frac{2}{\|w\|} \quad (4)$$

To solve this problem, we need to satisfy both the formula (3) and the maximum margin, so we can get:

$$\max_{w,b} \frac{2}{\|w\|^2} \quad (5)$$

$$\text{s.t. } y_i (w^T x_i + b) \geq 1 \quad i=1,2,3\dots m$$

Therefore, maximizing the margin equal to minimizing $\|w\|$, then we will get:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (6)$$

$$\text{s.t. } y_i (w^T x_i + b) \geq 1 \quad i=1,2,3\dots m$$

This is what we call SVM.

Introduce Lagrange function to simplify the complexity of this problem, we will get:

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

$\alpha = (\alpha_1, \alpha_2, \alpha_3 \dots \alpha_m)$, take the partial derivative of w and b in Equation (7) and set them equal to 0 and then put the answers in Equation (7), we will get a dual problem:

$$w(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (8)$$

$$\text{s.t. } y_i[(w \cdot x_i) + b] - 1 \geq 0$$

$$\sum_{i=1}^m y_i a_i = 0, \quad a_i \geq 0, \quad i=1,2,3 \dots m$$

This is a Quadratic Programming problem with inequality constraints. It will follow KKT (Karush-Kuhn-Tucker).

Therefore we can get the model:

$$f(x) = \sum_{i=1}^m a_i y_i x_i^T x + b$$

According to this result, we can get a significant conclusion: the model of SVM is only determined by the support vector!

2.2. Soft Margin

However, most data in our lives are linearly inseparable, as shown in figure 2:

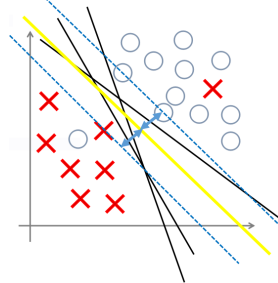


Figure 2. Hyperplanes that can not separate training samples correctly

How do we solve these kind of problems? To solve this problem, we should allow SVM make exceptions for some wrong samples. So we need to introduce slack variables ξ_i .

ξ_i indicates the degree of error made by each sample, $\xi_i=0$, it means no error, the larger ξ_i , the greater the error.

So formula (6) can be rewritten as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (9)$$

$$\text{s.t. } y_i(w^T x_i + b) \geq 1 - \xi_i$$

$$i=1,2,3 \dots m, \quad \xi_i \geq 0$$

This is what we call soft margin SVM.

$$\text{where } \xi_i = \begin{cases} 0 & \text{if } y_i(w^T x_i + b) \geq 1 \\ 1 - y_i(w^T x_i + b) & \text{otherwise} \end{cases}$$

It can be simplified to: $\xi_i = \max(0, 1 - yy')$

Because the constrain is always true when it is based on $\xi_i = \max(1 - yy', 0)$. Therefore, the formula (9) does not need to be constrained any more.

To optimize this problem, we can also use lagrangian multiplier method, so we will get:

$$L_{(w,b,\alpha,\xi,\mu)} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i + \sum_{i=1}^m \alpha_i (1 - \xi_i - y_i (w^T x_i + b)) - \sum_{i=1}^m \mu_i \xi_i \quad (10)$$

Make the partial derivative of L to w, b, μ equal to 0, and then we put these answers into Equation (10), we will get a dual problem:

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (11)$$

$$\text{s.t.} \quad \sum_{i=1}^m \alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq C \quad i=1,2,3,\dots,m$$

Similarly as hard margin, for the soft margin SVM, its will also follow KKT.

2.3. Kernel Function

In 2.1 and 2.2 we discussed some samples that be correctly separated by a hyperplane. In the figure 3, there are some samples which can not be separated by any hyperplane in the original sample space, how do we deal with this kind of problem?

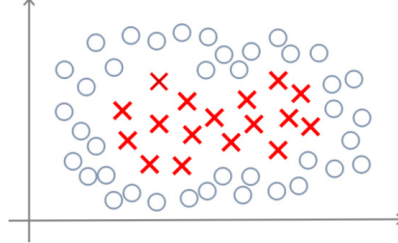


Figure3. Samples that can not be separated by heperplane

To solve this kind of problem, we can map the sample to a higher dimensional space. Maybe we can find some hyperplanes there to separate these samples correctly.

Let $\phi(x)$ denote the eigenvector after x is mapped, so the the model corresponding to the division of the hyperplane in the feature space can be expressed as:

$$f(x) = w^T \phi(x) + b \quad (12)$$

then formula (6) can be rewritten as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (13)$$

$$\text{s.t.} \quad y_i (w^T \phi(x_i) + b) \geq 1 \quad i=1,2,3,\dots,m$$

Similarly as (7), use lagrangian multiplier method, we will get a dual problem:

$$w(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\phi(x_i)^T \phi(x_j)) \quad (14)$$

$$\text{s.t.} \quad \sum_{i=1}^m y_i \alpha_i = 0 \quad \alpha_i \geq 0, i=1,2,3,\dots,m$$

Usually, when we calculate $\phi(x_i)^T \phi(x_j)$, it is very difficult, so we can define a function:

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$

Therefore, then formula (14) can be rewritten as:

$$w(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (15)$$

$$\text{s.t. } \sum_{i=1}^m y_i \alpha_i = 0 \quad \alpha_i \geq 0, i = 1, 2, 3, \dots, m$$

After calculation we will get:

$$f(x) = w^T \phi(x) + b = m \sum_{i=1}^m \alpha_i y_i k(x_i, x_j) + b \quad (16)$$

Function such as $k(\cdot, \cdot)$ is what we call kernel function. Here are some common kernel functions (use k to represent $k(x_i, x_j)$):

Table 1. SOME COMMONLY USED KERNEL FUNCTIONS

Name:	Expression:
Liner Kernel	$k = x_i^T x_j$
Polynomial Kernel	$k = (x_i^T x_j)^d$
Exponential Kernel	$k = \exp(-\frac{\ x_i - x_j\ }{\sigma})$
Gaussian Kernel	$k = \exp(-\frac{\ x_i - x_j\ ^2}{2\sigma^2})$
Sigmoid Kernel	$k = \tanh(\beta x_i^T x_j + \theta)$

Besides, we can also use Hybrid Kernel, which means that we can use multiple kernel functions at the same time, for example:

$$k = (x_i^T x_j)^d \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$$

3. The application of SVM

3.1. Application in Modern Machining

Bearing fault diagnosis is a study hotspot in the field of rotating machinery state monitoring. Sugumaran et al. used multi-class SVM to develop the fault diagnosis of rolling bearing in EDM[3]. They used Gaussian Kernel as a kernel function. This study used a kernel-based neighborhood score multi-class SVM for classification, and used a decision tree to select good features from all the features. The result of their study have shown that it is effective in diagnosing bearing fault conditions.

3.2. Application in Predicting Protein

The function of a specific membrane-bound protein is always associated with how it binds to the lipid bilayer and there are 5 types of membrane protein. Therefore, there is no denying that the speed of the process of determining the function of a new protein will be greatly increased if there is an automatic method to identify the type of membrane protein. Cai, Yu-Dong, et al trained a model to realize this idea[4]. In their study, they used the component of the amino acid as the input of SVM, and they used Gaussian Kernel as kernel function. After SVM being trained, they uses three professional testing methods to verify whether the training model has the function of predicting membrane proteins. The prediction accuracy of the three tests were 80.4%, 96.2%, and 85.4%. The results are better than the traditional method 'covariant discriminant algorithm', which indicate that SVM has the function of identify the membrane protein types.

3.3. Application in Face Detection

Face detection has a wide range of applications as a computer vision task. Face detection also has many potential applications in human-machine interfaces, monitoring systems, census systems, etc. Osuna, Edgar et al trained a SVM model to distinguish between faces and non-face patterns[5]. They used a 2nd-degree Polynomial Kernel as kernel function and inputted a database of 361 pixel patterns about face and non-face. The result of this experiment was that their model run 30 times faster than the previous system which was proposed by others. This also revealed that SVM can be used as an effective method for face detection.

4. Conclusion

In the previous discussion, we learned that the superiority of SVM made it a significant development in the fields of modern machining, predicting protein and face detection. We also learned that the model of SVM is only determined only by the support vectors, so it is very convenient for us to train.

But SVM also has some drawbacks[6], for instance, the training efficiency of the existing SVM for large-scale sample data sets of practical problems cannot reach the ideal training efficiency. Therefore the future development of SVM may concentrate more on how to further improve the SVM algorithm.

In addition, despite the fact that SVM has outstanding advantages in theory, compared with theoretical research, applied research on application still lags behind. Therefore, how to apply SVM more in people's daily life and explore new application areas of SVM will be the emphasis of future research.

ACKNOWLEDGEMENT

This paper was completed with the care and help of my teaching assistant from topic selection to writing. At my professor's class, he told me how to organize the structure of my paper and answered me some practical questions about SVM. Thanks to my professor and teaching assistant, without their help I can not complete this paper.

REFERENCE

- [1] Zhang Xuegong. On Statistical Learning Theory and Support Vector Machines[J]. Acta Automatica Sinica, 2000, 26(1): 32-42.
- [2] Vapnik V N. The nature of statistical learning[J]. Theory, 1995..
- [3] Sugumaran V, Sabareesh G R, Ramachandran K I. Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine[J]. Expert Systems with Applications, 2008, 34(4): 3090-3098.
- [4] Cai Y D, Ricardo P W, Jen C H, et al. Application of SVM to predict membrane protein types[J]. Journal of theoretical biology, 2004, 226(4): 373-376.
- [5] Osuna E, Freund R, Girosit F. Training support vector machines: an application to face detection[C]//Proceedings of IEEE computer society conference on computer vision and pattern recognition. IEEE, 1997: 130-136.
- [6] Ding Shifei, Qi Bingjuan, Tan Hongyan. Summary of support vector machine theory and algorithm research[D]., 2011.