

Research Article

CLASSIFICATION OF DATA USES TWIN SEPARABLE HYPERSPHERE

Mansour Zalani Sofla¹, Nahid Rabiei Somar², *Nader Rafati Maleki³ and Fatemeh Zalani Sofla³

¹Department of Mathematics, Eyvan-e-Gharb Branch, Islamic Azad University, Eyvan, Iran

²Department of Mathematical Sciences, the University of Guilan, Rasht-Iran

³Department of Mathematics, Tabriz Branch, Islamic Azad University, Tabriz, Iran

*Author for Correspondence

ABSTRACT

This paper formulates a twin support vector machine classifier by describing support vector machine and twin vector machine classifier that determines a pair of hypersphere by solving two related SVM-type programming problems such that this is more efficient than the classical SVM. In addition, the TSVH successfully avoids matrix inversion compared with the twin support vector machine (TSVM) which displays algorithms of SVM can be easily extended to this TSVH.

Keywords: Support Vector Machine, Kernel, Hypersphere, Classification, Hyperplane

INTRODUCTION

The algorithms of SVM for the first time was introduced by Vapnik (1995, 1998) denasalized a non-linear equation. Kernel based methods have proved to be a powerful tool for solving classification problems. The aim of support machine is finding the best function in order to classify, such that it can be distinguished member of two classes in to data set. Its application includes recognition of letters and numbers, voice and face, mentioned and disease and etc. see Muphy and Aha (1992), Joachims (1998) and Cao and Tai (2001).

TSVM aims at generating two non-parallel hyper-planes such that each one is closer to one class and is at least one for from the other class for any given binary data set. SVDD classification is one class classification based on support vector. It is supposed that is a hypersphere with a center and R radius and ζ_i error variables, it has the lowest radius and covers target class samples but using non-linear descriptions and kernel functions result in appropriation limits.

TSVH (Osuna *et al.*, 1997) is provided by describing support vector data. TSVH efforts to produce two hyperspheres which each hypersphere includes as many as possible samples of one class and is as far as possible from those of the other class. Similar to TSVM, the TSVH solves two smaller-sized QPPs rather than large one. TSVH has some advantages compared with TSVM.

We can find of (1) that mentioned problem is a programming problem of quadratic with linear constrains and it is solved by using (quadprog) box which is related to optimization in MATLAB software. In the next section, numerical results are presented in the chart and discuss method.

Support Vector Machine (SVM)

The algorithms of SVM for the first time was introduced by Vapnik (1963) and Vapnik and Corina (1995) denasalized a non-linear equation.

Separable and Linear Support Vector Machine

D set includes n of member which we have it and is expressed as following:

$$D = \{(x_i, y_i); x_i \in R^n, y_i \in \{-1, 1\}\}, i = 1, \dots, n$$

The goal of this equation is finding hyper planes which can separate by maximum distance from margined points that it separates points with $y_i = 1$ of points with $y_i = -1$.

Each hyper plan can be described by set of \mathbf{X} points such that it confirm this condition $w \cdot x + b = 0$, where w is normal vector that is vertical on hyperplane.(6)

We want to choose w such that produce great distance between parallel-hyper planes. These hyper planes are described using these relationships: $w \cdot x + b = 1$, $w \cdot x + b = -1$. for linear separable problem data, we can consider two hyper planes in the points margin, which they haven't been any similar point

Research Article

and for their distance, create a maximum that distance of two hyper planes of paper is $\frac{2}{\|w\|}$. Thus we

must minimize $\|w\|$. The closest training data to separable hyper planes termed support vector.

The problem of support vector machine (SVM) is formulated as follows:

$$\begin{aligned} \min & \frac{1}{2} w^T w \\ \text{s.t} & y_i (w^T \cdot x + b) \geq 1 \\ \text{for} & i = 1, 2, \dots, n \end{aligned} \tag{1}$$

This problem is constrained optimization problem which it is related-convex and quadratic, for solving this problem we form Lagrangian (3, 4) function as follows:

$$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_{i=1}^n \alpha_i (y_i (w \cdot x_i + b) - 1) \tag{2}$$

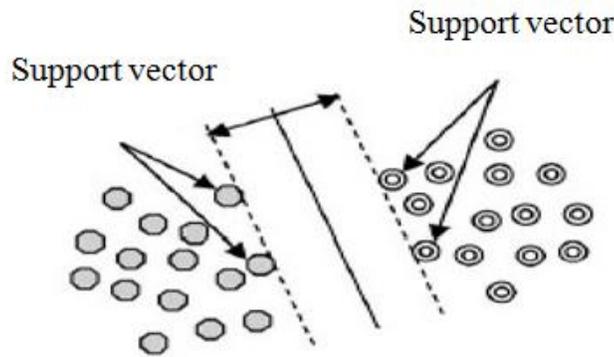


Figure 1:

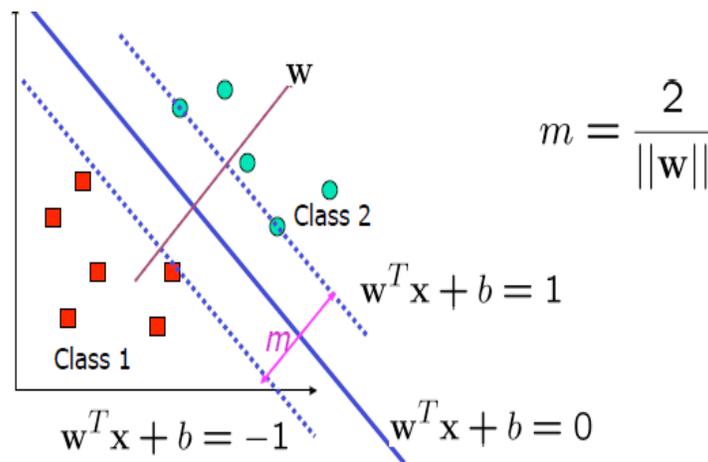


Figure 2:

For obtaining the answer of problem that is (w, b, α) , this answer must confirm Karush-Kuhn-Tucker (K.K.T)(4) condition and by the answer point, L derivation to the w, b, α must be equals to zero. By equaling this derivation to zero, we can obtain the follow equations:

Research Article

$$w = \sum_{i=1}^n \alpha_i y_i x_i \tag{3}$$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Thus dual problem can be expressed as following equation:

$$\max \quad -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i \tag{4}$$

subject to $0 \leq \alpha_i \leq C$

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Support Vector Machine in Inseparable Case

A group of ζ_i variables are defined for inseparable cases which termed lack variables, in this condition, optimization of support vector machine is specified as follows:

$$\min \frac{1}{2} w^T w + c \sum_{i=1}^n \zeta_i$$

s.t $y_i(w \cdot x_i + b) \geq 1 - \zeta_i$ (5)

$\zeta_i \geq 0 \quad i = 1, 2, \dots, n$

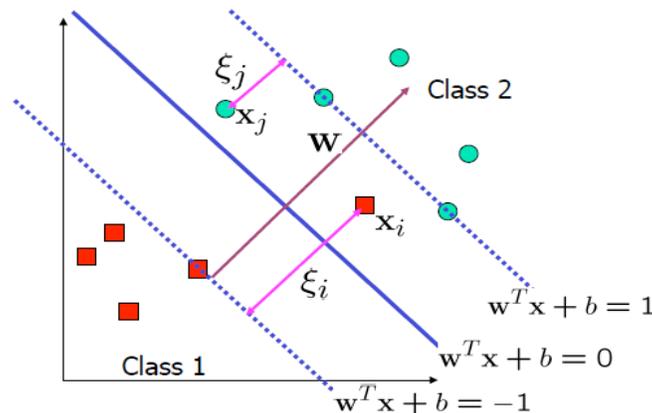


Figure 3:

For this problem we must make K.K.T condition in the answer point and so we obtain the below dual problem:

$$\max \quad -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i$$

subject to $0 \leq \alpha_i \leq C$ (6)

$$\sum_{i=1}^n \alpha_i y_i = 0$$

Research Article

Solving SVM (4,6)problem in inseparable situation is similar to solve it in the separable but there is a difference between α_i Lagrangian coefficients changes. In both separable and inseparable cases, difference function for classifying x input expressed by:

$$f(x) = \text{sgn}(wx + b) \tag{7}$$

Non-Linear Support Vector Machine

Separable hyper plane was presented by Vapnik (1995) which it was a linear class.

A strategy was proposed by Bernhard *et al.*, (1992) for producing a non-linear classification that was used a core for finding hyper-plane with more margine, resultant algorithm seems similar to the other and what is different that all point multiples with a non-linear core function as substituted.

$$k(x_i, x_j) = (x_i \cdot x_j)^d \quad \text{Homogeneous polynomial}$$

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d \quad \text{Heterogeneous polynomial}$$

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad \text{Goussian}$$

$$k(x_i, x_j) = \tanh(kx_i \cdot x_j + c) \quad \text{Hyperbolic tangent}$$

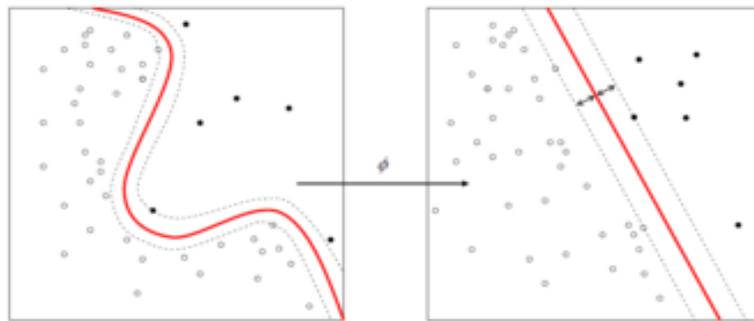


Figure 4:

Twin Support Vector Machine (TSVM)

Jayadeva *et al.*, (2007) proposed a twin support vector (TSVM) classifier for categorizing twin data. TSVM aims at generating two non-parallel hyper-planes such that each one is closer to one class and is at least one for from the other class for any given binary data set.

Strategy of solving smaller-sized pair of QPPs instead of a large one as in classical SVMs makes the learning speed of TSVM be approximately four times faster than classical SVM.

This is because TSVM differs from one fundamental way, in which a pair of smaller sized is solved, where as in the classic SVM, only a single QPPs is solved.

However, there exists some shortcoming in the TSVM. First, for non-linear TSVM, the objective functions of its dual QPPs require inversion of matrix of size (n+1)*(n+1) twice, where n is number of samples and includes computational costs for increasing their optimization. Second, TSVM uses pair of non-parallel hyper-planes to indicate descriptions of two classes of samples such that each hyper-plane is closest to data points of class and is at distance of at least 1 from the data points of the other one.

Twin Support Vector Machine (TSVM)

$$\begin{aligned} \min & \frac{1}{2} \sum_{i \in I^+} (w_+^T x_i + b_+)^2 + C_1 \sum_{j \in I^-} \zeta_j \\ \text{s.t.} & -w_+^T x_j - b_+ \geq 1 - \zeta_j \\ & \zeta_j \geq 0, j \in I^- \end{aligned} \tag{9}$$

Research Article

$$\min \frac{1}{2} \sum_{i \in I^-} (w_-^T x_j + b_-)^2 + C_2 \sum_{i \in I^+} \zeta_i$$

$$s.t \quad w_-^T x_i - b_- \geq 1 - \zeta_i \tag{10}$$

$$\zeta_i \geq 0, i \in I^+$$

The dual formulation of (9), (10)

$$\max \sum_{j \in I^-} \alpha_j - \frac{1}{2} \sum_{j_1, j_2 \in I^-} \alpha_{j_1} \alpha_{j_2} z_{j_1}^T (\sum_{i \in I^+} z_i z_i^T)^{-1} z_{j_2}$$

$$s.t \quad 0 \leq \alpha_j \leq C_1, j \in I^- \tag{11}$$

$$\max \sum_{i \in I^+} \alpha_i - \frac{1}{2} \sum_{i_1, i_2 \in I^+} \alpha_{i_1} \alpha_{i_2} z_{i_1}^T (\sum_{j \in I^-} z_j z_j^T)^{-1} z_{i_2}$$

$$s.t \quad 0 \leq \alpha_i \leq C_2, i \in I^+ \tag{12}$$

The problem decision function of TSVM(3) is measured when X new input has the lowest distance of each hyperplanes.

$$f(x) = \operatorname{argmin}\{d_{\pm}(x)\} \tag{13}$$

$$d_{\pm}(x) = \frac{|w_{\pm}^T x + b_{\pm}|}{\|w_{\pm}\|} \tag{14}$$

Support Vector Data Description (SVDD)

SVDD (4) classification is one class classification based on support vector. It is supposed that is a hypersphere with a center and R radius and ζ_i error variables, it has the lowest radius and covers target class samples but using non-linear descriptions and kernel functions result in appropriation limits.

$$\min R^2 + C \sum_{i=1}^n \zeta_i$$

$$\|x_i - a\|^2 \leq R^2 + \zeta_i$$

$$\zeta_i \geq 0$$

$$i = 1, \dots, n \tag{15}$$

Which $C \sum_{i=1}^n \zeta_i$ clause indicates error rate and C parameter responsible for balancing between hypersphere and error rate we design a Lagrangian function for solving optimization problem.

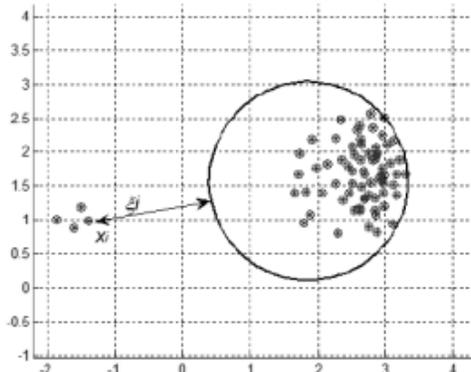


Figure 6:

Research Article

After substituting these formulas in binary Lagrangian function, the optimization problem of the SVDD is obtained by:

$$\begin{aligned} \max \quad & \sum_{i=1}^n \alpha_i x_i x_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j x_i x_j \\ \text{s.t} \quad & \sum_{i=1}^n \alpha_i = 1, \quad 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n \end{aligned} \tag{16}$$

To calculate R^2 , it needs enough that a distance of voluntary of sample on the sphere obtains from the center of hyper-sphere which is expressed as following:

$$R^2 = \|x - a\|^2 = x x - 2 \sum_{i=1}^n \alpha_i (x x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j (x_i x_j) \tag{17}$$

This hypersphere acts according to this signal function:

$$\begin{aligned} f(x) &= \text{sgn}(R^2 - \|\varphi(x) - a\|^2) \\ &= R^2 - \sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) \\ &\quad + 2 \sum_i \alpha_i \alpha_j k(x_i, x_j) \end{aligned}$$

Twin Support Vector Hypersphere

TSVH (4) is provided by describing support vector data .TSVH efforts to produce two hyperspheres which each hypersphere includes as many as possible samples of one class and is as far as possible from those of the other class.

Similar to TSVM, the TSVH solves two smaller-sized QPPs rather than large one.

TSVH has some advantages compared with TSVM.

First, the formulation of TSVH differs from TSVM formulation. Particularly, it is avoided inversion matrix in TSVM that this issue lead to the TSVH to obtain lower cost and can be easily extended to large-scale problems with efficient algorithm.

Second, TSVH uses pair of hyperspheres to describe samples, such as each hypersphere covers as many as possible samples of one class, whereas, it is far from another class Samples classifying TSVH is obtained by solving below pair of QPPs:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i \in I^+} \|\varphi(x_i) - c_+\|^2 - \nu_1 R_+^2 + C_1 \sum_{j \in I^-} \xi_j \\ \text{s.t} \quad & \|\varphi(x_j) - c_+\|^2 \geq R_+^2 - \xi_j, \\ & R_+^2 \geq 0, \quad \xi_j \geq 0, \quad i \in I^- \end{aligned} \tag{19}$$

Research Article

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i \in I^-} \left\| \varphi(x_j) - c_- \right\|^2 - v_2 R_-^2 + C_2 \sum_{j \in I^-} \xi_j \\ \text{s.t.} \quad & \left\| \varphi(x_i) - c_- \right\|^2 \geq R_-^2 - \xi_i, \\ & R_-^2 \geq 0, \quad \xi_i \geq 0, \quad i \in I^+ \end{aligned} \tag{20}$$

This model finds two hyperspheres (one for each class) and their points are classified correspond to the hypersphere which a given point is closest to.

The first benchmark is related to the sum of squared distances from the center of hypersphere to points of one class. The second benchmark maximizes the squared radius of this hypersphere which makes this hypersphere be as large as possible.

Dual first problem:

$$\begin{aligned} \max \quad & \frac{1}{2} \sum_{i \in I^+} k(x_i, x_i) - \frac{1}{2(I^+ - 2v_1)} \sum_{i_1, i_2 \in I^+} k(x_{i_1}, x_{i_2}) - \sum_{j \in I^-} \alpha_j k(x_{j_1}, x_{j_2}) \\ & - \frac{2}{I^+ - 2v_1} \sum_{j_1, j_2 \in I^-} \alpha_{j_1} \alpha_{j_2} k(x_{j_1}, x_{j_2}) + \frac{2}{I^+ - 2v_1} \sum_{i \in I^+, j \in I^-} \alpha_j k(x_j, x_i) \\ \text{s.t.} \quad & \sum_{j \in I^-} \alpha_j = v_1, \quad 0 \leq \alpha_j \leq C_1, \quad j \in I^- \end{aligned} \tag{21}$$

Similarly dual of second problem is obtained.

For x new input sample which obtained based on-ort class of two hyperspheres single function expressed as follows:

$$f(x) = \arg \min_{+,-} \left\{ \frac{\left\| \varphi(x) - c_+ \right\|^2}{R_+^2}, \frac{\left\| \varphi(x) - c_- \right\|^2}{R_-^2} \right\}. \tag{22}$$

We can find of (1) that mentioned problem is a programming problem of quadratic with linear constrains and it is solved by using (quadprog) box which is related to optimization in MATLAB software. In the next section, numerical results are presented in the chart and discuss method.

RESULTS AND DISCUSSION

Numerical Results

In this section, accuracy and speed of presented method are considered by data which have been extracted of (1) equation. For calculating, 2013 a MATLAB program and one system equipped ASUS and K43 B processing and 4 GB rom were used. In this research, we have data that some of them considered as education data and some data are experimental, but first data is normalized that is these data are suffered and then it is done an average, the next stage, training data and their test are considered to. Of course, for predicating process, factual data must be compared with previous data. Averaging data improves prediction. A good kernel function and selecting C parameter is a major parameter in improving plan accuracy. In the following table numerical results have obtained and presented:

Dataset	SVM	
	Accuracy%	Time(s)
Sonar	80.82	1106.5
Spect	71.83	1202.31
Ionosphere	89.23	1904.74

Research Article

Haberman	74.22	736.10
fertilityDiagnosis	88.07	188.51

Conclusion

In this paper considers support vector machine and one of its challenges is exist of complexity in cancellation .twin support vector machine includes pair of QPPs that in each QPP, objective function equals sepecific class and their limitations are determined by other samples of class. In comparison with large-sized QPP ,TSVH lead to two smaller-sized QPP which maximizes speed of TSVM four times faster classic SVM for learning its hyperspheres. In contrast with TSVM fomulation which parallel class data points get clustered around each hypersphere, TSVH uses pair of hyper sphere in the SVDD to indicates two classes, only problem which there is about TSVH and SVDD is related to lack of convex of studied cases.

ACKNOWLEDGEMENT

We are grateful to Islamic Azad University, Tabriz branch authorities, for their useful collaboration.

REFERENCES

- Bernhard E Boser, Isabelle M Guyon and Vladimir N Vapnik (1992).** A Training Algorithm for Optimal Margin Classifiers, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*.
- Cao L and Tay FEH (2001).** Financial forecasting using support vector machines. *Neural Computing and Applications* **10**(2) 184-192.
- Jayadeva Khemchandani R and Chandra S (2007).** Twin support vector machines for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(5) 905–910.
- Joachims T (1998).** Text categorization with support vector machines: learning with many relevant features, *European Conference on Machine Learning, Chemnitz (Springer)* **10** 137-142.
- Muphy PM and Aha DW (1992).** UCI Repository of Machine Learning Databases. Department of Information and Computer Science (University of California at Irvine) Irvine, CA.
- Osuna E, Freund R and Girosi F (1997).** Training support vector machines: an application to face detection, *Proceedings of the 1997 Conference Computer Vision and Pattern Recognition[C]*. (IEEE Computer Society), Washington 130- 136.
- Vapnik VN (1995).** *The Natural of Statistical Learning Theory* (Springer) New York.
- Vapnik VN (1998).** *Statistical Learning Theory* (Wiley) New York.

Appendices

Algorithm 1: SVM

```

%min 0.5*w'w +C*1'*[y1;y2]
% s.t
- %Aw - 1*b - y1 <= -1
  %Bw + 1*b - y2 <= -1
%
function [w,b]=QuadprogSVM (A,B)
%clear all
clc
C=0.34;
[m1,n]=size(A);
[m2,~]=size(B);
H=[eye(n) zeros(n,1) zeros(n,m1+m2) ; (
  zeros(1,n) 0 zeros(1,m1+m2) ;(

```

Research Article

```

zeros(m1+m2,n) zeros(m1+m2,1) zeros(m1+m2,[(
f= [zeros(n,1);(
; 0
C*ones(m1+m2,1);[(
d=-ones(m1+m2,1);(
AA=[A;B;[
DD=eye(m1+m2);(
for iii=m1+1:m1+m2
    DD(iii,iii)=-1;
end
D=[-DD*AA -DD*ones(m1+m2,1) -eye(m1+m2);[(
lb=[-inf*ones(n,1);-inf;zeros(m1+m2,1);[(
opts = optimoptions('quadprog','Algorithm','active-set','Display','off;('
[x,fval] = quadprog(H,f,D,d,[],[],lb,[],[],opts;(
%
%min 1/2*x'*H*x + f*x
%s.t
%D*x <= d
%lb = x
%

w=x(1:n;(
b=x(n+1;(
y=x(n+2:end,1;(
E=C*ones(m1+m2,1).'*y;
disp('fval;('=
fprintf('%g + %g = %g\n',fval-E,E,fval);(

```

Algorithm 2: accuracy

```

clc
mu=.000 ;
delta=1e-3;
Cbig=1e10;
v=.1;
tic
if mu~=0
D=[A;B;[
]md,nd]=size(D;(
A=Rec_Kernel(A,D',mu;(
B=Rec_Kernel(B,D',mu;(
end
]rowsA,nn] = size(A;(
]rowsB,~] = size(B;(
breakInd = rowsA;
samples = rowsA + rowsB;
dataMatrix = [A;B;[
classLabels = ones(rowsA,1);(
classLabels = [classLabels;-ones(rowsB,1);[(
sampleInd = 1:samples;
nFolds =10;

```

