Fuzzy Twin Support Vector Machine Based on Intra-class Hyperplane

Hanghang Zhang¹ and Haipeng Li²,³
¹Graduate Team, Engineering University of PAP, Shaanxi, China, Email: 1043009963@qq.com
²No.1 General Department, Xi’an Electronic Engineering Research Institute, Shaanxi, China
³National Lab of Radar Signal Processing, Xidian University, Shaanxi, China, Email: 317707690@qq.com

Abstract. Aiming at the sensitivity of twin support vector machines to noise and outliers, an improved fuzzy twin support vector machine is proposed. Aiming at the shortcomings of fuzzy membership based on the distance between a sample and its class center, this algorithm proposes a new fuzzy membership function that can effectively reflect sample uncertainty—based on intra-class hyperplane and sample affinity. When determining the membership function, the intra-class hyperplane replaces the class center, reducing the dependence on the sample set geometry and improving the model generalization ability. At the same time, considering the relation between each sample and the confusion degree of sample points, the sample affinity is applied to the design of membership function, so as to effectively distinguish the effective sample from noise and outliers and improve the classification accuracy. The experimental results show that compared with the twin support vector machine and the classic fuzzy support vector machine, the improved fuzzy twin support vector machine has better classification performance.

1. Introduction
Support Vector Machine (SVM) is a powerful binary classification machine learning method proposed by Vapnik et al. SVM is based on statistical learning theory such as Vapnik-Chervonenkis (VC) dimension theory and structural risk minimization principle. It overcomes the problems of small sample size, dimension disasters and over-learning. It has strong learning ability and generalization ability, and has been successfully applied in many fields. Solving convex quadratic programming problems, SVM is time-consuming. In order to overcome the problem of too long training time of SVM, scholars have made many improvements on the basis of SVM. In 2006, Mangasarian et al. proposed the Generalized Eigenvalue Proximal SVM (GEPSVM). GEPSVM eliminates the constraint that the hyperplanes should be parallel. Thus, by solving the eigenvectors of the smallest eigenvalues corresponding to the generalized eigenvalue problem, the problem translates into generating a non-parallel plane for each type of data. But the disadvantage is that the accuracy is not satisfactory for points at the intersection of the two planes. Therefore, in order to further improve, Jayadeva et al. proposed the Twin Support Vector Machine (TWSVM) in 2007. TWSVM generates a plane for each class sample, making each plane as close as possible to the samples belonging to its own class and as far as possible from the other class samples. The two planes of TWSVM may not be parallel. The problem is transformed into two smaller quadratic programming problems, so the training time is also reduced to 1/4 of SVM’s.
In the process of classification, the existence of data outliers and noise affects the generalization ability of classification model. The fuzzy membership function solves the problem of different characteristics of samples by setting different membership degrees for different sample points. Therefore, scholars proposed fuzzy twin support vector machine (FTWSVM) based on the above TWSVM. For the design of fuzzy membership function, there is no general criterion to follow. According to the fact that noise and outliers are generally far from the class center, Khemchandani R et al. use the distance relationship between samples and the class center to determine the fuzzy membership function. That is, the farther from the class center, the smaller the membership value. Considering the influence of geometric shape of sample set on the model, Du et al. used intra-class hyperplanes instead of class centers. A fuzzy membership function based on the distance between sample and intra-class hyperplane is proposed, which is more consistent with the classification principle of SVM. In this paper, the intra-class hyperplane is applied to the design of membership function, which reduces the dependence of the model on the geometry of the sample set. At the same time, considering the relationship between each sample, a new membership function design method is proposed by using sample affinity and confusion. The effect of noise and isolated points is effectively reduced, and the obtained model is more robust.

2. Twin Support Vector Machine
Given a training set of m training points \( T = \{(x_i, y_i), (x_2, y_2), \ldots, (x_m, y_m)\} \) in the n-dimensional real space \( \mathbb{R}^n \), \( x_i \in \mathbb{R}^n, i = 1, 2, \ldots, m \). \( x_i \) is the \( i \)th sample. let \( y_i \in \{+1, -1\} \) denote the class to which the \( i \)th sample belongs. \( A \in \mathbb{R}^{m \times n} \) represent the training samples belonging to classes +1. \( B \in \mathbb{R}^{m \times n} \) represent the training samples belonging to classes -1. Training datasets, TWSVM obtains two non-parallel separating hyperplanes through:

\[
x^T w_1 + b_1 = 0, x^T w_2 + b_2 = 0
\]

The two separated hyperplanes of TWSVM are obtained by solving the following pair of dual optimization problems:

\[
\begin{align*}
\max_{\alpha} & \quad \epsilon_1^T \alpha - \frac{1}{2} \alpha^T G (H^T H)^{-1} G^T \alpha \\
\text{s.t.} & \quad 0 \leq \alpha \leq c \epsilon_2
\end{align*}
\]

(2a)

\[
\begin{align*}
\max_{\gamma} & \quad \epsilon_2^T \gamma - \frac{1}{2} \gamma^T H (G^T G)^{-1} H^T \gamma \\
\text{s.t.} & \quad 0 \leq \gamma \leq c \epsilon_1
\end{align*}
\]

(2b)

where \( H = [A \quad e_1], G = [B \quad e_2] \). \( \alpha \in \mathbb{R}^m, \gamma \in \mathbb{R}^m \) are Lagrangian multiplier vectors. The purpose of this step is to solve the dual problem more easily, and to introduce the kernel function naturally, so as to extend it to the non-linear classification problem. The nonparallel hyperplanes (1) can be obtained from the solutions of \( \alpha^* \) and \( \gamma^* \) in formula 2a and formula 2b by:

\[
u = -(H^T H)^{-1} G^T \alpha^*, \quad v = (G^T G)^{-1} H^T \gamma^*
\]

(3)

where \( u = [w_1 \quad b_1]^T, v = [w_2 \quad b_2]^T \). Each new test data point \( x \in \mathbb{R}^n \) is assigned to a given class \( l \) by using the following formula depending on which plane is closest to that data point.

\[
f(x) = \arg \min_{l=1,2} |x^T w_l + b_l|
\]

(4)
3. Fuzzy Twin Support Vector Machine
In practical, the impact of different training samples on the hyperplane is often different. In FTWSVM, different weighted parameters are assigned to each sample based on fuzzy membership value, so as to enhance the function of normal samples and weaken the influence of noise and outliers.

The dual optimization problem of FTWSVM is:

\[
\begin{align*}
\max_{\alpha} & \quad e_+^T \alpha - \frac{1}{2} \alpha^T G(H^T H)^{-1} G^T \alpha \\
\text{st.} & \quad 0 \leq \alpha \leq c_1 s_2 \\
\max_{\gamma} & \quad e_-^T \gamma - \frac{1}{2} \gamma^T H(G^T G)^{-1} H^T \gamma \\
\text{st.} & \quad 0 \leq \gamma \leq c_2 s_1
\end{align*}
\]

(5a)

(5b)

\( s_1 \in R^{m_1} \) and \( s_2 \in R^{m_2} \) represents the fuzzy membership of positive and negative sample points respectively.

The classification process is similar to TWSVM. Comparing the dual problem between TWSVM and FTWSVM, it can be observed that the Lagrangian multipliers of the two are different. The upper bound of \( \alpha, \gamma \) of FTWSVM varies with the fuzzy membership \( s_i \). That is, each point has its own penalty factor, so as to distinguish different sample points, improve the classification accuracy, and make the algorithm more robust. Therefore, the design of fuzzy membership degree is the key to the performance of FTWSVM algorithm.

4. Improvement of Fuzzy Membership Function
There are many methods to construct fuzzy membership function. The classical design method is based on the distance relationship between the sample point and the center of the class. However, this method reduces the effect of samples that are close to the hyperplane but far from the class center, and is not friendly to the sample set with non-spherical distribution. In this paper, a distance function based on intra-class hyperplane is proposed. The distance from sample point to intra-class hyperplane is replaced by the distance from the center of the class, which reduces the dependence of the algorithm on the geometric shape of the sample set and more effectively reflects the contribution of the sample to the classification surface. At the same time, the design method based on distance cannot distinguish support vector and outlier effectively. Moreover, the self-classification of \( k \) neighboring sample points in cross-region is not considered. Therefore, a new fuzzy membership function is proposed in this paper, which combines the sample affinity with the distance to the intra-class hyperplane.

4.1. Distance Function based on Intra-class Hyperplane
The positive and negative class sample class centers are \( C^+, C^- \), and \( \bar{w} = C^+ - C^- \) is a normal vector. The hyperplanes passing through A and S are \( H^+ : \bar{w}^T (x - C^+) = 0 \) and \( H^- : \bar{w}^T (x - C^-) = 0 \), respectively. Then the distance between the sample point and its own intra-class hyperplane in the class are \( d_i^+ = \frac{|\bar{w}^T (x_i - C^+)|}{\|\bar{w}\|} \) and \( d_i^- = \frac{|\bar{w}^T (x_i - C^-)|}{\|\bar{w}\|} \), respectively. \( C^+ = \frac{1}{m_1} \sum_{i=1}^{m_1} x_i^+ \), \( C^- = \frac{1}{m_2} \sum_{i=1}^{m_2} x_i^- \).

Let \( D^+ = \max\{d_i^+\} \), \( D^- = \max\{d_i^-\} \).

The distance function from the sample point to the intra-class hyperplane is: \( q_i^\pm = 1 - \frac{d_i^\pm}{D^+ + \delta} \) where \( \delta > 0 \) and small enough to make \( 0 < q_i^\pm \leq 1 \). The closer \( q_i^\pm \) is to 1, the closer the sample point is
to the hyperplane; the closer to 0, the farther the sample point is from the sample point. Because $d_i^+\text{ and } D^+\text{ contain } \|\mathbf{v}\|, \text{ eliminate } \|\mathbf{v}\| \text{ to simplify calculation.}$

For the non-linear case, the mapping function $\phi(\cdot)$ is introduced. The positive and negative class centers of sample points in high-dimensional feature space $H$ are respectively:

$$C_{m+} = \frac{1}{m_1} \sum_{i=1}^{m_1} \phi(x_i^+)$$

$$C_{m-} = \frac{1}{m_2} \sum_{i=1}^{m_2} \phi(x_i^-)$$

Using kernel tricks, the distance from the positive sample point to the hyperplane is:

$$d_i^+ = \left| (\phi(C^+) - \phi(C^-))^T (\phi(x_i^+) - \phi(C^+)) \right|$$

$$= \left| \frac{1}{m_1} \sum_{j=1}^{m_1} k(x_i^+, x_j^+) - \frac{1}{m_2} \sum_{j=1}^{m_2} k(x_i^+, x_j^-) - \frac{1}{m_1} \sum_{j=1}^{m_1} k(x_i^+, x_j^+) + \frac{1}{m_2} \sum_{j=1}^{m_2} k(x_i^+, x_j^-) \right|$$

(6)

where $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. Similarly, the distance from the negative sample point to the hyperplane is:

$$d_i^- = \left| \frac{1}{m_2} \sum_{j=1}^{m_2} k(x_i^-, x_j^-) - \frac{1}{m_1} \sum_{j=1}^{m_1} k(x_i^-, x_j^+) - \frac{1}{m_2} \sum_{j=1}^{m_2} k(x_i^-, x_j^-) + \frac{1}{m_1} \sum_{j=1}^{m_1} k(x_i^-, x_j^+) \right|$$

(7)

4.2. Sample Affinity

d_ {ij}$ is expressed as the distance between sample point $x_i$ and $x_j$. Sort $d_ {ij}$ from large to small, and take $k$ samples nearest to $A$, denoted as $x_1, x_2, \ldots, x_k$, then the distance between $x_i$ and $x_j (j = 1, 2, \ldots, k)$ is $d_{i1}, d_{i2}, \ldots, d_{ik}$. Sample affinity can be expressed as $t_i = \sum_{j=1}^{k} \frac{1}{d_{ij} + a}$ where $a > 0$ and small enough,

$$d_{ij} = \sqrt{k(x_i, x_j) - 2k(x_i, x_j) + k(x_j, x_j)}$$

Normalize $t_i$:

$$p_i = \frac{t_i}{T_i}, \quad T_i = \max\{t_1, t_2, \ldots, t_m\}$$

(8)

However, $p_i$ can only reflect the distance relationship between samples, and does not take into account the influence of the category of the sample points. So it is adjusted. If k nearest neighbor samples are of the same class without confusion, they remain unchanged. If there is confusion between the two classes of samples, $p_i$ is appropriately reduced. If k nearest neighbor samples are heterogeneous samples, the sample point is considered to be noise, and the assignment is 0, which reduces its influence on the classification hyperplane. The specific expression is as follows:

$$p_i = (1 - \frac{l}{k}) \frac{t_i}{T_i}$$

(9)

$l$ is the number of heterogeneous samples in the k neighbor samples.

4.3. Design of New Fuzzy Membership Function

A new fuzzy membership function is obtained by combining the distance function between sample point and intra-class hyperplane and the sample affinity:
According to equation (10), when the sample affinity is constant, the farther the sample is from the intra-class hyperplane, the smaller the membership degree $s_i$ is. When the distance function $q_i$ is constant, the higher the sample density is, the lower the degree of confusion is, and the higher the membership $s_i$ is. At the same time, unlike the distance function based on the class center, the model can reduce the dependence on the shape of the sample set by replacing the class center with the intra-class hyperplane. The new membership function can better reflect the uncertainty of samples.

5. Experimental Results and Analysis

In order to evaluate the performance of the proposed algorithm (New-FTWSVM), this paper carries out simulation experiments through 2 artificial data sets and 8 UCI data sets. It compares with SVM, TWSVM and class-center distance-based FTWSVM (FTWSVM-D) in terms of classification accuracy and classification time. The experimental environment is processor Intel core i5, CPU 2.60GHz, 4.00 GB RAM, based on 64-bit R 3.4.4.

5.1. Artificial Datasets

The simulation data are randomly generated by computer, which are normal sample data set and non-spherical sample data set. The simulation data is randomly generated by the computer, which are a normal sample data set and a non-spherical sample data set, respectively. Both data sets contain 470 two types of two-dimensional samples and 30 noise points, of which 200 are used as training sets and 300 are used as test sets. Figure 1 and 2 are schematic diagrams of sample distribution. For comparison, all algorithms use a Gaussian kernel. Set penalty parameter $c_1 = c_2 = 10$, $\sigma = 10$, $k=10$ and $\delta = a = 0.05$. Ten-fold cross validation method was used to select the model. The experimental results are shown in Table 1.

![Figure 1. Normal sample dataset.](image1)

![Figure 2. Non-spherical sample dataset.](image2)

| Datasets     | Classification Performance | SVM  | TWSVM | FTWSVM-D | New-FTWSVM |
|--------------|----------------------------|------|-------|----------|------------|
| Normal Sample| Accuracy/%                 | 85.33| 86.00 | 87.66    | 88.33      |
|              | Time/s                     | 4.26 | 1.15  | 1.31     | 1.38       |
| Non-spherical Sample | Accuracy/%             | 80.00| 82.33 | 84.66    | 89.00      |
|              | Time/s                     | 4.55 | 1.23  | 1.35     | 1.44       |
As can be seen from Table 1, the classification accuracy of New-FTWSVM is the highest. Especially in the non-spherical sample classification results, the accuracy is higher than SVM, TWSVM, FTWSVM-D are 9.00%, 6.66%, 4.33%, respectively. It shows that New-FTWSVM has stronger effect on the classification of non-spherical samples and can effectively reduce the influence of noise. The classification time is slightly longer than TWSVM and FTWSVM-D. This is because calculating fuzzy membership, especially sample tightness, requires a certain amount of computation, but is superior to SVM.

5.2. UCI Datasets
Eight standard datasets from the UCI database were selected for experiments. The parameters are the same as the experimental settings of the artificial datasets. The experimental results are shown in Table 2. It can be seen that the classification accuracy of New-FTWSVM is better than SVM, TWSVM and FTWSVM-D. The classification time is greatly reduced compared with SVM, which shows the superiority of the algorithm.

| Datasets       | Classification Performance | SVM   | TWSVM | FTWSVM-D | New-FTWSVM |
|---------------|---------------------------|-------|-------|----------|------------|
| Heart disease | Accuracy/%                | 82.16 | 81.69 | 85.60    | 87.02      |
| 270x13        | Time/s                    | 5.43  | 1.86  | 2.23     | 2.61       |
| Breast cancer | Accuracy/%                | 67.89 | 68.34 | 70.48    | 74.87      |
| 683x10        | Time/s                    | 66.45 | 15.27 | 17.35    | 18.96      |
| Ionosphere    | Accuracy/%                | 78.32 | 78.99 | 80.56    | 82.33      |
| 351x34        | Time/s                    | 27.55 | 8.12  | 8.72     | 9.14       |
| Australian    | Accuracy/%                | 84.69 | 84.63 | 85.37    | 87.59      |
| 690x14        | Time/s                    | 75.39 | 17.95 | 19.71    | 21.08      |
| German        | Accuracy/%                | 73.46 | 72.34 | 77.60    | 78.92      |
| 1000x24       | Time/s                    | 143.24| 37.70 | 41.39    | 44.27      |
| Votes         | Accuracy/%                | 95.49 | 95.62 | 97.10    | 97.80      |
| 435x16        | Time/s                    | 35.72 | 9.04  | 9.46     | 9.77       |
| Pima-Indian   | Accuracy/%                | 81.55 | 81.98 | 84.19    | 87.36      |
| 768x8         | Time/s                    | 58.67 | 16.31 | 17.02    | 17.74      |
| ILPD          | Accuracy/%                | 80.67 | 84.89 | 87.40    | 88.64      |
| 583x10        | Time/s                    | 42.60 | 11.21 | 12.89    | 13.28      |

6. Conclusions and Future Work
In this paper, a new design method of fuzzy membership function is proposed based on the intra-class hyperplane and sample affinity, and it is applied to FTWSVM. By assigning different membership degrees to each sample point, the dependence of the model on the geometry of the sample set is reduced, and the generalization ability and robustness of the model are improved. The experimental results show that the proposed algorithm is superior to TWSVM and classic FTWSVM when the classification time is feasible. Since the algorithm involves multiple parameters, the next step is to research and optimize it.

7. Reference
[1] Cortes C and Vapnik V 1995 Support-vector networks Machine Learning, Vol. 20 No. 3, pp. 273-297.
[2] Mangasarian O L and Wild E W 2006 Multisurface Proximal Support Vector Machine Classification via Generalized Eigenvalues IEEE Transactions on Pattern Analysis &Machine Intelligence, Vol. 28 No. 1, pp. 69-74.
[3] Jayadeva, Khemchandani R and Chandra S 2007 Twin Support Vector Machines for pattern classification *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol. 29 No. 5, pp. 905.

[4] Khemchandani R, Jayadeva, and Chandra S 2008 Fuzzy Twin Support Vector Machines for Pattern Classification *WORLD SCIENTIFIC*, pp. 131-142.

[5] DU Z, LIU S and QI X 2009 Fuzzy Support Vector Machine with New Membership Function *Journal of System Simulation*, No. 7, pp. 1901-1903.

[6] Ding S, Zhang N, Zhang X and Wu F 2017 Twin support vector machine: theory, algorithm and applications *Neural Computing and Applications*, Vol. 28 No. 11, pp. 3119-3130.