

Article

An Effective Multiclass Twin Hypersphere Support Vector Machine and Its Practical Engineering Applications

Qing Ai ^{1,2,*} , Anna Wang ^{2,*}, Aihua Zhang ³ and Wenhui Wang ² and Yang Wang ²

¹ School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China

² College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; angleboy@foxmail.com (W.W.); wangyang0531@163.com (Y.W.)

³ College of Engineering, Bohai University, Jinzhou 121000, China; jsxixi_zah@163.com

* Correspondence: lyaiqing@126.com (Q.A.); wanganna@mail.neu.edu.cn (A.W.)

Received: 27 September 2019; Accepted: 17 October 2019; Published: 20 October 2019



Abstract: Twin-KSVC (Twin Support Vector Classification for K class) is a novel and efficient multiclass twin support vector machine. However, Twin-KSVC has the following disadvantages. (1) Each pair of binary sub-classifiers has to calculate inverse matrices. (2) For nonlinear problems, a pair of additional primal problems needs to be constructed in each pair of binary sub-classifiers. For these disadvantages, a new multi-class twin hypersphere support vector machine, named Twin Hypersphere-KSVC, is proposed in this paper. Twin Hypersphere-KSVC also evaluates each sample into 1-vs-1-vs-rest structure, as in Twin-KSVC. However, our Twin Hypersphere-KSVC does not seek two nonparallel hyperplanes in each pair of binary sub-classifiers as in Twin-KSVC, but a pair of hyperspheres. Compared with Twin-KSVC, Twin Hypersphere-KSVC avoids computing inverse matrices, and for nonlinear problems, can apply the kernel trick to linear case directly. A large number of comparisons of Twin Hypersphere-KSVC with Twin-KSVC on a set of benchmark datasets from the UCI repository and several real engineering applications, show that the proposed algorithm has higher training speed and better generalization performance.

Keywords: K-SVCR; Twin-KSVC; 1-vs-1-vs-rest; twin hypersphere support vector machine

1. Introduction

Support vector machine (SVM) [1,2], as a computationally powerful tool for classification, have already applied in wide engineering problems [3–8]. The SVM has three elements that make it so successful, including structural risk minimization (SRM) principle, kernel trick and dual theory. However, SVM has to solve a large-sized quadratic programming problem (QPP), which greatly limits its applications. To improve learning complexity of SVM, Jayadeva et al. proposed twin SVM (TSVM) [9]. Unlike SVM that seeks an optimal separating hyperplane which maximizes the margin of two classes of samples, TSVM constructs two nonparallel proximal hyperplanes, each of which is close to the corresponding class as possible, and keeps as far away as possibly from the opposite class. The strategy makes TSVM only need to solve two smaller QPPs, instead of one larger QPP as in SVM. Due to its high learning speed, TSVM has attracted interest in recent years. Many improvements have also been proposed [10–28], where the twin hypersphere SVM (THSVM) [25–28] is an excellent improvement of TSVM. Unlike TSVM, THSVM seeks a pair of hyperspheres, instead of two nonparallel hyperplanes, to depict two classes of samples. Compared with TSVM, THSVM has better classification performance.

The SVM and TSVM can only solve the binary classification problems, however, many practical engineering problems are often related to multi-class classification in the real world. Currently, we usually use 1-vs-rest and 1-vs-1 strategy to resolve multi-class classification problems. In the 1-vs-1 SVM, $K(K-1)/2$ binary SVM sub-classifiers are constructed. Each sub-classifier can be trained by using two classes of samples. Because only two classes are considered for each sub-classifier in the 1-vs-1 SVM and the rest samples are not involved, the 1-vs-1 SVM may get unfavorable classification results. The 1-vs-rest SVM needs to construct K binary SVM sub-classifiers. Each sub-classifier can be trained by using all the samples; thus, the 1-vs-rest SVM may lead to class imbalance problems. For above drawbacks of 1-vs-1 SVM and 1-vs-rest SVM, Angulo et al. proposed a new support vector classification-regression machine for K class (K-SVCR) [29]. $K(K-1)/2$ binary SVM sub-classifiers are constructed, each of which is trained with all the samples and evaluates each sample into 1-vs-1-vs-rest structure. K-SVCR avoids the class imbalance problems and information loss. Compared with 1-vs-1 SVM and 1-vs-rest SVM, K-SVCR achieves better generalization performance. Twin-KSVC [30,31], being an effective extension of K-SVCR, is based on TSVM and also evaluates each sample into 1-vs-1-vs-rest structure. Twin-KSVC achieves higher learning speed in comparison with K-SVCR. However, Twin-KSVC has the following disadvantages:

- Each pair of sub-classifiers has to calculate inverse matrices, which is extraordinarily time-consuming for the large-scale engineering problems.
- For nonlinear problems, each pair of sub-classifiers needs to construct a pair of additional primal problems, instead of directly applying the kernel trick to linear case as in SVM.

For the disadvantages of Twin-KSVC, in this paper, we propose a Twin Hypersphere-KSVC, inspired by THSVM. The Twin Hypersphere-KSVC also evaluates each sample into 1-vs-1-vs-rest structure, as in Twin-KSVC. However, our Twin Hypersphere-KSVC does not seek two nonparallel hyperplanes in each pair of binary sub-classifiers as in Twin-KSVC, but a pair of hyperspheres. Compared with Twin-KSVC, Twin Hypersphere-KSVC avoids computing inverse matrices, and for nonlinear problems, can apply the kernel trick to linear case directly.

This paper is outlined as follows. We briefly review the related multi-class classification algorithms in Section 2. In Section 3, the Twin Hypersphere-KSVC is proposed in detail. The experimental results on a set of benchmark datasets and several real engineering problems are presented in Section 4 and the conclusions are drawn in last section.

2. Related Works

In this paper, we consider a multi-class classification problem with a training dataset $D = \{x_p^k \in R^d | k = 1, \dots, K, p = 1, \dots, n_k\}$, where K is the number of classes and n_k is the number of the samples of the k -th class. The size of training dataset is $n = n_1 + \dots + n_K$. Denote, for convenience, by X_k the sets of samples of the k -th class, i.e., $X_k = \{x_p^k | p = 1, \dots, n_k\}$.

2.1. Review of K-SVCR Multi-Classifier

The K-SVCR multi-classifier [29] is based on decomposition-reconstruction strategy. $K(K-1)/2$ binary SVM sub-classifiers are constructed, each of which evaluates each sample into 1-vs-1-vs-rest structure. The classification result of K-SVCR is shown in Figure 1a intuitively.

The sub-classifier $f^{ij}(x)$ for two focused classes i and j in K-SVCR seeks an optimal hyperplane

$$w^{ij} \cdot x + b^{ij} = 0, \quad (1)$$

where $w^{ij} \in R^d$ is the normal vector and $b^{ij} \in R$ is the bias term. The optimal hyperplane can be obtained by resolving the following QPP:

$$\begin{aligned}
& \min \frac{1}{2} \|w^{ij}\|^2 + c_1 \left(\sum_{p=1}^{n_i} \eta_p^{ij} + \sum_{p=1}^{n_j} \eta_p^{ij*} \right) + c_2 \sum_{p=1}^{n_{\bar{ij}}} (\xi_p^{ij} + \xi_p^{ij*}), \\
& \text{s.t. } w^{ij} \cdot x_p^i + b^{ij} \geq 1 - \eta_p^{ij}, \quad p = 1, \dots, n_i, \\
& \quad w^{ij} \cdot x_p^j + b^{ij} \leq -1 + \eta_p^{ij*}, \quad p = 1, \dots, n_j, \\
& \quad -\varepsilon - \xi_p^{ij*} < w^{ij} \cdot x_p^{\bar{ij}} + b^{ij} \leq \varepsilon + \xi_p^{ij}, \quad p = 1, \dots, n_{\bar{ij}}, \\
& \quad \eta_p^{ij} \geq 0, \quad p = 1, \dots, n_i, \quad \eta_p^{ij*} \geq 0, \quad p = 1, \dots, n_j, \\
& \quad \xi_p^{ij} \geq 0, \quad \xi_p^{ij*} \geq 0, \quad p = 1, \dots, n_{\bar{ij}},
\end{aligned} \tag{2}$$

where $x_p^{\bar{ij}} \in D - X_i - X_j$, $n_{\bar{ij}} = n - n_i - n_j$, $\tilde{\xi}_p^{ij}$, $\tilde{\xi}_p^{ij*}$ and η_p^{ij} , η_p^{ij*} are slack variables, the parameter ε is restricted to $[0, 1)$.

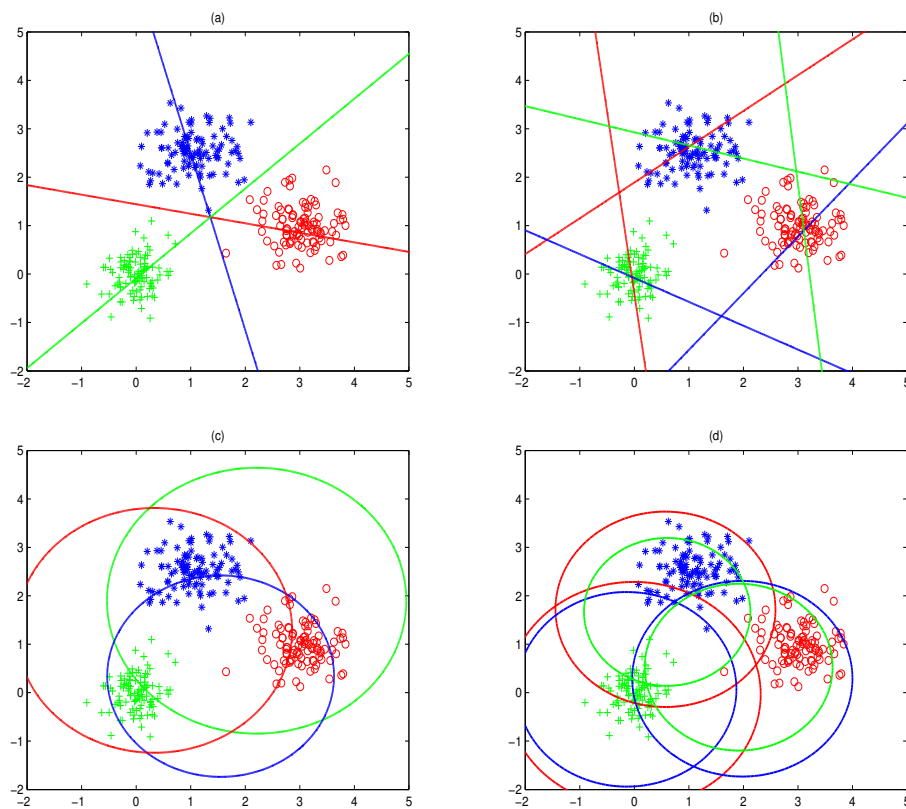


Figure 1. (a) K-SVC. (b) Twin-KSVC. (c) THKSVM. (d) Twin Hypersphere-KSVC.

For a testing sample x , the sub-classifier $f^{ij}(x) = w^{ij} \cdot x + b^{ij}$ determines its class by

$$F^{ij}(x) = \begin{cases} -1, & \text{if } f^{ij}(x) < \varepsilon, \\ 1, & \text{if } f^{ij}(x) > \varepsilon, \\ 0, & \text{else.} \end{cases} \tag{3}$$

For the testing sample x , the final label can be determined by vote rule.

2.2. Review of Twin-KSVC Multi-Classifier

Twin-KSVC [30,31] is an improvement of K-SVC. The Twin-KSVC constructs $K(K-1)/2$ pairs of binary TSVM sub-classifiers, which evaluates each sample into 1-vs-1-vs-rest structure. The classification result is intuitively presented in Figure 1b.

The sub-classifiers $f^i(x)$ and $f^j(x)$ for two focused classes i and j in Twin-KSVC seek a pair of hyperplane

$$w^i \cdot x + b^i = 0 \text{ and } w^j \cdot x + b^j = 0 \quad (4)$$

where $w^{i(j)} \in R^d$ and $b^{i(j)} \in R$ are the normal vector and the bias term of the corresponding hyperplane, respectively. The two hyperplanes can be obtained by resolving the QPPs as follows:

$$\begin{aligned} \min & \frac{1}{2} \sum_{p=1}^{n_i} (w^i \cdot x_p^i + b^i)^2 + c_1 \sum_{p=1}^{n_j} \eta_p^i + c_2 \sum_{p=1}^{n_{\bar{j}}} \zeta_p^i, \\ \text{s.t.} & -(w^i \cdot x_p^j + b^i) + \eta_p^i \geq 1, \quad p = 1, \dots, n_j, \\ & -(w^i \cdot x_p^{\bar{j}} + b^i) + \zeta_p^i \geq 1 - \varepsilon, \quad p = 1, \dots, n_{\bar{j}}, \\ & \eta_p^i \geq 0, \quad p = 1, \dots, n_j, \\ & \zeta_p^i \geq 0, \quad p = 1, \dots, n_{\bar{j}}, \end{aligned} \quad (5)$$

$$\begin{aligned} \min & \frac{1}{2} \sum_{p=1}^{n_j} (w^j \cdot x_p^j + b^j)^2 + c_3 \sum_{p=1}^{n_i} \eta_p^j + c_4 \sum_{p=1}^{n_{\bar{i}}} \zeta_p^j, \\ \text{s.t.} & (w^j \cdot x_p^i + b^j) + \eta_p^j \geq 1, \quad p = 1, \dots, n_i, \\ & (w^j \cdot x_p^{\bar{i}} + b^j) + \zeta_p^j \geq 1 - \varepsilon, \quad p = 1, \dots, n_{\bar{i}}, \\ & \eta_p^j \geq 0, \quad p = 1, \dots, n_i, \\ & \zeta_p^j \geq 0, \quad p = 1, \dots, n_{\bar{i}}, \end{aligned} \quad (6)$$

where $\eta_p^{i(or j)}$ and $\zeta_p^{i(or j)}$ are slack variables.

For a testing sample x , the sub-classifiers $f^i(x) = x^T w^i + b^i$ and $f^j(x) = x^T w^j + b^j$ assign its class by

$$F^{ij}(x) = \begin{cases} -1, & \text{if } f^j(x) < 1 - \varepsilon, \\ 1, & \text{if } f^i(x) > -1 + \varepsilon, \\ 0, & \text{else.} \end{cases} \quad (7)$$

For the testing sample x , the final label can be also determined by vote rule.

2.3. Review of THKSVM Multi-Classfier

THKSVM (Twin Hypersphere Multiclass Support Vector Machine) [32] integrates THSVM and 1-vs-rest structure. THKSVM constructs K hyperspheres in the training stage, whose classification result is intuitively shown in Figure 1c.

The sub-classifier for the focused classes i in THKSVM seeks a hypersphere

$$\|x - a_i\|^2 = R_i^2 \quad (8)$$

where $a_i \in R^d$ and $R_i \in R$ are the center and the radius of the corresponding hypersphere, respectively. The hypersphere can be constructed by resolving the following QPP:

$$\begin{aligned} \min & \frac{1}{2} \sum_{p=1}^{n_{\bar{i}}} \|x_p^{\bar{i}} - a_i\|^2 - v_1 R_i^2 + c_1 \sum_{p=1}^{n_i} \eta_p^i, \\ \text{s.t.} & \|x_p^i - a_i\|^2 \geq R_i^2 - \eta_p^i, \\ & R_i^2 \geq 0, \quad \eta_p^i \geq 0, \quad p = 1, \dots, n_i, \end{aligned} \quad (9)$$

where $x_p^{\bar{i}} \in D - X_i$, $n_{\bar{i}} = n - n_i$ and $\eta_p^i \geq 0$ are slack variables.

The class of a testing sample x can be determined by

$$\text{Class } k = \arg \max_{i=1, \dots, K} \frac{\|x - a_i\|^2}{R_i^2}. \quad (10)$$

3. Twin Hypersphere-KSVC

Twin Hypersphere-KSVC, inspired by THSVM and 1-vs-1-vs-rest structure, constructs $K(K-1)/2$ pairs of hyperspheres in the training stage. For two focused classes i and j , Twin Hypersphere-KSVC seeks a pair of hypersphere (a_i, R_i) and (a_j, R_j) , where a_i (R_i) and a_j (R_j) are respectively the centers (radii) of the corresponding hyperspheres. Each hypersphere covers the corresponding class as many as possibly, keeps as far away as possibly from another class, contains the rest of samples as little as possibly, and the radius of the hypersphere is as small as possible. The Twin Hypersphere-KSVC are intuitively presented in Figure 1d.

3.1. Linear Case

For the linear case, each pair of hyperspheres (a_i, R_i) and (a_j, R_j) for two focused classes i and j in Twin Hypersphere-KSVC is constructed by resolving the following QPPs:

$$\begin{aligned} \min R_i^2 - \frac{v_1}{n_j} \sum_{p=1}^{n_j} \|x_p^j - a_i\|^2 + \frac{c_1}{n_i} \sum_{p=1}^{n_i} \eta_p^i + \frac{c_2}{n_{ij}} \sum_{p=1}^{n_{ij}} \xi_p^i, \\ \text{s.t. } \|x_p^i - a_i\|^2 \leq R_i^2 + \eta_p^i, \quad p = 1, \dots, n_i, \\ \|x_p^{ij} - a_i\|^2 \geq R_i^2 - \xi_p^i, \quad p = 1, \dots, n_{ij}, \\ \eta_p^i \geq 0, \quad p = 1, \dots, n_i, \\ \xi_p^i \geq 0, \quad p = 1, \dots, n_{ij}, \\ R_i^2 \geq 0, \end{aligned} \quad (11)$$

$$\begin{aligned} \min R_j^2 - \frac{v_2}{n_i} \sum_{p=1}^{n_i} \|x_p^i - a_j\|^2 + \frac{c_3}{n_j} \sum_{p=1}^{n_j} \eta_p^j + \frac{c_4}{n_{ij}} \sum_{p=1}^{n_{ij}} \xi_p^j, \\ \text{s.t. } \|x_p^j - a_j\|^2 \leq R_j^2 + \eta_p^j, \quad p = 1, \dots, n_j, \\ \|x_p^{ij} - a_j\|^2 \geq R_j^2 - \xi_p^j, \quad p = 1, \dots, n_{ij}, \\ \eta_p^j \geq 0, \quad p = 1, \dots, n_j, \\ \xi_p^j \geq 0, \quad p = 1, \dots, n_{ij}, \\ R_j^2 \geq 0. \end{aligned} \quad (12)$$

where $\eta_p^{i(j)}$ and $\xi_p^{i(j)}$ are slack variables.

The Lagrangian function L for the QPP (11) is given by:

$$\begin{aligned} L = R_i^2 - \frac{v_1}{n_j} \sum_{p=1}^{n_j} \|x_p^j - a_i\|^2 + \frac{c_1}{n_i} \sum_{p=1}^{n_i} \eta_p^i + \frac{c_2}{n_{ij}} \sum_{p=1}^{n_{ij}} \xi_p^i + \sum_{p=1}^{n_i} \alpha_p (\|x_p^i - a_i\|^2 - R_i^2 - \eta_p^i) \\ - \sum_{p=1}^{n_{ij}} \beta_p (\|x_p^{ij} - a_i\|^2 - R_i^2 + \xi_p^i) - \sum_{p=1}^{n_i} s_p \eta_p^i - \sum_{p=1}^{n_{ij}} q_p \xi_p^i - \lambda R_i^2. \end{aligned} \quad (13)$$

The Karush-Kuhn-Tucker (KKT) conditions are satisfied as follows:

$$2 \frac{v_1}{n_j} \sum_{p=1}^{n_j} (x_p^j - a_i) - 2 \sum_{p=1}^{n_i} \alpha_p (x_p^i - a_i) + 2 \sum_{p=1}^{n_{ij}} \beta_p (x_p^{ij} - a_i) = 0, \quad (14)$$

$$1 - \sum_{p=1}^{n_i} \alpha_p + \sum_{p=1}^{n_{ij}} \beta_p - \lambda = 0, \quad (15)$$

$$\frac{c_1}{n_i} - \alpha_p - s_p = 0, \quad p = 1, \dots, n_i, \quad (16)$$

$$\frac{c_2}{n_{ij}} - \beta_p - q_p = 0, \quad p = 1, \dots, n_{ij}, \quad (17)$$

$$\alpha_p (\|x_p^i - a_i\|^2 - R_i^2 - \eta_p^i) = 0, \quad p = 1, \dots, n_i, \quad (18)$$

$$\beta_p (\|x_p^{ij} - a_i\|^2 - R_i^2 + \xi_p^i) = 0, \quad p = 1, \dots, n_{ij}, \quad (19)$$

$$s_p \eta_p^i = 0, \quad p = 1, \dots, n_i, \quad (20)$$

$$q_p \xi_p^i = 0, \quad p = 1, \dots, n_{ij}, \quad (21)$$

$$\lambda R_i^2 = 0. \quad (22)$$

From (14), (15) and (22), we can obtain

$$a_i = \frac{1}{1 - v_1} \left(\sum_{p=1}^{n_i} \alpha_p x_p^i - \frac{v_1}{n_j} \sum_{p=1}^{n_j} x_p^j - \sum_{p=1}^{n_{ij}} \beta_p x_p^{ij} \right) \quad (23)$$

By denoting

$$v_i = \frac{1}{1 - v_1} \quad (24)$$

and substituting (16)–(23) into (13), the dual optimal problem of (11) is obtained as follows:

$$\begin{aligned} \max \quad & \sum_{p=1}^{n_i} \alpha_p x_p^i \cdot x_p^i - \sum_{p=1}^{n_{ij}} \beta_p x_p^{ij} \cdot x_p^{ij} + 2v_i \frac{v_1}{n_j} \sum_{p=1}^{n_j} \sum_{p_2=1}^{n_i} \alpha_{p_2} x_{p_1}^j \cdot x_{p_2}^i - 2v_i \frac{v_1}{n_j} \sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_{ij}} \beta_{p_2} x_{p_1}^j \cdot x_{p_2}^{ij} \\ & - v_i \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_i} \alpha_{p_1} \alpha_{p_2} x_{p_1}^i \cdot x_{p_2}^i + 2v_i \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_{ij}} \alpha_{p_1} \beta_{p_2} x_{p_1}^i \cdot x_{p_2}^{ij} - v_i \sum_{p_1=1}^{n_{ij}} \sum_{p_2=1}^{n_{ij}} \beta_{p_1} \beta_{p_2} x_{p_1}^{ij} \cdot x_{p_2}^{ij} \\ \text{s.t.} \quad & 1 - \sum_{p=1}^{n_i} \alpha_p + \sum_{p=1}^{n_{ij}} \beta_p = 0, \\ & 0 \leq \alpha_p \leq \frac{c_1}{n_i}, \quad p = 1, \dots, n_i, \\ & 0 \leq \beta_p \leq \frac{c_2}{n_{ij}}, \quad p = 1, \dots, n_{ij}. \end{aligned} \quad (25)$$

By defining $\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{n_i} \end{pmatrix}$ and $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{n_{ij}} \end{pmatrix}$, the optimal problem (25) can be reformulated as

$$\begin{aligned} \max \quad & -v_i (\alpha^T \quad \beta^T) \begin{pmatrix} X_i^T X_i & -X_i^T X_{ij} \\ -X_{ij}^T X_i & X_{ij}^T X_{ij} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \text{diag}(X_i^T X_i) + 2v_i \frac{v_1}{n_j} X_i^T X_j e_j \\ -\text{diag}(X_{ij}^T X_{ij}) - 2v_i \frac{v_1}{n_j} X_{ij}^T X_j e_j \end{pmatrix}^T \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \\ \text{s.t.} \quad & \begin{pmatrix} e_i^T & -e_{ij}^T \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 1, \\ & 0e_i \leq \alpha \leq \frac{c_1}{n_i} e_i, \\ & 0e_{ij} \leq \beta \leq \frac{c_2}{n_{ij}} e_{ij}. \end{aligned} \quad (26)$$

According to the KKT conditions (16)–(21), we can obtain R_i^2 by the following formula:

$$R_i^2 = \|x^* - a_i\|^2, \quad (27)$$

where $x^* \in S_{i1} \cup S_{i2}$, $S_{i1} = \{x_p^i | 0 < \alpha_p < \frac{c_1}{n_i}\}$ and $S_{i2} = \{x_p^{ij} | 0 < \beta_p < \frac{c_2}{n_{ij}}\}$.

By denoting $v_j = \frac{1}{1-v_2}$, the dual optimal problem of (12) can be obtained as follows:

$$\begin{aligned} \max \quad & \sum_{p=1}^{n_j} \theta_p x_p^j \cdot x_p^j - \sum_{p=1}^{n_{ij}} \gamma_p x_p^{ij} \cdot x_p^{ij} + 2v_j \frac{v_2}{n_i} \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_j} \theta_{p_2} x_{p_1}^i \cdot x_{p_2}^j - 2v_j \frac{v_2}{n_i} \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_{ij}} \gamma_{p_2} x_{p_1}^i \cdot x_{p_2}^{ij} \\ & - v_j \sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_j} \theta_{p_1} \theta_{p_2} x_{p_1}^j \cdot x_{p_2}^j + 2v_j \sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_{ij}} \theta_{p_1} \gamma_{p_2} x_{p_1}^j \cdot x_{p_2}^{ij} - v_j \sum_{p_1=1}^{n_{ij}} \sum_{p_2=1}^{n_{ij}} \gamma_{p_1} \gamma_{p_2} x_{p_1}^{ij} \cdot x_{p_2}^{ij}, \\ \text{s.t.} \quad & 1 - \sum_{p=1}^{n_j} \theta_p + \sum_{p=1}^{n_{ij}} \gamma_p = 0, \\ & 0 \leq \theta_p \leq \frac{c_3}{n_j}, p = 1, \dots, n_j, \\ & 0 \leq \gamma_p \leq \frac{c_4}{n_{ij}}, p = 1, \dots, n_{ij}. \end{aligned} \quad (28)$$

By defining $\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_{n_j} \end{pmatrix}$ and $\gamma = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{n_{ij}} \end{pmatrix}$, the (28) can be reformulated as

$$\begin{aligned} \max \quad & -v_j (\theta^T \quad \gamma^T) \begin{pmatrix} X_j^T X_j & -X_j^T X_{ij} \\ -X_{ij}^T X_j & X_{ij}^T X_{ij} \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \end{pmatrix} + \left(\text{diag}(X_j^T X_j) + 2v_j \frac{v_2}{n_i} X_j^T X_i e_i \right)^T \begin{pmatrix} \theta \\ \gamma \end{pmatrix}, \\ \text{s.t.} \quad & \begin{pmatrix} e_j^T & -e_{ij}^T \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \end{pmatrix} = 1, \\ & 0e_j \leq \theta \leq \frac{c_3}{n_j} e_j, \\ & 0e_{ij} \leq \gamma \leq \frac{c_4}{n_{ij}} e_{ij}. \end{aligned} \quad (29)$$

We can compute R_j^2 by the following formula:

$$R_j^2 = \|x^* - a_j\|^2, \quad (30)$$

where $x^* \in S_{j1} \cup S_{j2}$, $S_{j1} = \{x_p^j | 0 < \theta_p < \frac{c_3}{n_j}\}$ and $S_{j2} = \{x_p^{ij} | 0 < \gamma_p < \frac{c_4}{n_{ij}}\}$.

3.2. Nonlinear Case

We extend the linear Twin Hypersphere-KSVC to the nonlinear case by directly considering the nonlinear map $\varphi : R^d \rightarrow H$ (H is a high-dimensional Hilbert space), instead of the kernel generated surfaces in Twin-KSVC.

$$\begin{aligned}
& \min R_i^2 - \frac{v_1}{n_j} \sum_{p=1}^{n_j} \|\varphi(x_p^j) - a_i\|^2 + \frac{c_1}{n_i} \sum_{p=1}^{n_i} \eta_p^i + \frac{c_2}{n_{ij}} \sum_{p=1}^{n_{ij}} \xi_p^i, \\
& \text{s.t. } \|\varphi(x_p^i) - a_i\|^2 \leq R_i^2 + \eta_p^i, \quad p = 1, \dots, n_i, \\
& \quad \|\varphi(x_p^{\bar{i}j}) - a_i\|^2 \geq R_i^2 - \xi_p^i, \quad p = 1, \dots, n_{ij}, \\
& \quad \eta_p^i \geq 0, \quad p = 1, \dots, n_i, \\
& \quad \xi_p^i \geq 0, \quad p = 1, \dots, n_{ij}, \\
& \quad R_i^2 \geq 0,
\end{aligned} \tag{31}$$

$$\begin{aligned}
& \min R_j^2 - \frac{v_2}{n_i} \sum_{p=1}^{n_i} \|\varphi(x_p^i) - a_j\|^2 + \frac{c_3}{n_j} \sum_{p=1}^{n_j} \eta_p^j + \frac{c_4}{n_{ij}} \sum_{p=1}^{n_{ij}} \xi_p^j, \\
& \text{s.t. } \|\varphi(x_p^j) - a_j\|^2 \leq R_j^2 + \eta_p^j, \quad p = 1, \dots, n_j, \\
& \quad \|\varphi(x_p^{\bar{i}j}) - a_j\|^2 \geq R_j^2 - \xi_p^j, \quad p = 1, \dots, n_{ij}, \\
& \quad \eta_p^j \geq 0, \quad p = 1, \dots, n_j, \\
& \quad \xi_p^j \geq 0, \quad p = 1, \dots, n_{ij}, \\
& \quad R_j^2 \geq 0.
\end{aligned} \tag{32}$$

According to the dual theory, one can get the dual optimal problems of (31) and (32) as follows:

$$\begin{aligned}
& \max -v_i (\alpha^T \quad \beta^T) \begin{pmatrix} K(X_i, X_i) & -K(X_i, X_{ij}) \\ -K(X_{ij}, X_i) & K(X_{ij}, X_{ij}) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\
& \quad + \begin{pmatrix} \text{diag}(K(X_i, X_i)) + 2v_i \frac{v_1}{n_j} K(X_i, X_j) e_j \\ -\text{diag}(K(X_{ij}, X_{ij})) - 2v_i \frac{v_1}{n_j} K(X_{ij}, X_j) e_j \end{pmatrix}^T \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \\
& \text{s.t. } \begin{pmatrix} e_i^T & -e_{ij}^T \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 1, \\
& \quad 0e_i \leq \alpha \leq \frac{c_1}{n_i} e_i, \\
& \quad 0e_{ij} \leq \beta \leq \frac{c_2}{n_{ij}} e_{ij},
\end{aligned} \tag{33}$$

$$\begin{aligned}
& \max -v_j (\theta^T \quad \gamma^T) \begin{pmatrix} K(X_j, X_j) & -K(X_j, X_{ij}) \\ -K(X_{ij}, X_j) & K(X_{ij}, X_{ij}) \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \end{pmatrix} \\
& \quad + \begin{pmatrix} \text{diag}(K(X_j, X_j)) + 2v_j \frac{v_2}{n_i} K(X_j, X_i) e_i \\ -\text{diag}(K(X_{ij}, X_{ij})) - 2v_j \frac{v_2}{n_i} K(X_{ij}, X_i) e_i \end{pmatrix}^T \begin{pmatrix} \theta \\ \gamma \end{pmatrix}, \\
& \text{s.t. } \begin{pmatrix} e_j^T & -e_{ij}^T \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \end{pmatrix} = 1, \\
& \quad 0e_j \leq \theta \leq \frac{c_3}{n_j} e_j, \\
& \quad 0e_{ij} \leq \gamma \leq \frac{c_4}{n_{ij}} e_{ij},
\end{aligned} \tag{34}$$

where $K(.,.)$ is a kernel matrix.

3.3. Decision Rule

For a testing sample x , each pair of sub-classifiers determines its label by

$$F_{ij}(x) = \begin{cases} 1 & \|\varphi(x) - a_i\|^2 / R_i^2 \leq \|\varphi(x) - a_j\|^2 / R_j^2 \\ -1 & \|\varphi(x) - a_i\|^2 / R_i^2 > \|\varphi(x) - a_j\|^2 / R_j^2 \end{cases} \tag{35}$$

where

$$\begin{aligned} \|\varphi(x) - a_i\|^2 = & K(x, x) + 2v_i \left(\sum_{p=1}^{n_i} \alpha_p K(x_p^i, x) - \frac{v_1}{n_j} \sum_{p=1}^{n_j} K(x_p^j, x) - \sum_{p=1}^{n_{ij}} \beta_p K(x_p^{ij}, x) \right) \\ & + v_i^2 \left(\sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_i} \alpha_{p_1} \alpha_{p_2} K(x_{p_1}^i, x_{p_2}^i) + \left(\frac{v_1}{n_j} \right)^2 \sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_j} K(x_{p_1}^j, x_{p_2}^j) + \sum_{p_1=1}^{n_{ij}} \sum_{p_2=1}^{n_{ij}} \beta_{p_1} \beta_{p_2} K(x_{p_1}^{ij}, x_{p_2}^{ij}) \right) \\ & - \frac{2v_1}{n_j} \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_j} \alpha_{p_1} K(x_{p_1}^i, x_{p_2}^j) - 2 \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_{ij}} \alpha_{p_1} \beta_{p_2} K(x_{p_1}^i, x_{p_2}^{ij}) + \frac{2v_1}{n_j} \sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_{ij}} \beta_{p_2} K(x_{p_1}^j, x_{p_2}^{ij}) \end{aligned} \quad (36)$$

and

$$\begin{aligned} \|\varphi(x) - a_j\|^2 = & K(x, x) + 2v_j \left(\sum_{p=1}^{n_j} \theta_p K(x_p^j, x) - \frac{v_2}{n_i} \sum_{p=1}^{n_i} K(x_p^i, x) - \sum_{p=1}^{n_{ij}} \gamma_p K(x_p^{ij}, x) \right) \\ & + v_j^2 \left(\sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_j} \theta_{p_1} \theta_{p_2} K(x_{p_1}^j, x_{p_2}^j) + \left(\frac{v_2}{n_i} \right)^2 \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_i} K(x_{p_1}^i, x_{p_2}^i) + \sum_{p_1=1}^{n_{ij}} \sum_{p_2=1}^{n_{ij}} \gamma_{p_1} \gamma_{p_2} K(x_{p_1}^{ij}, x_{p_2}^{ij}) \right) \\ & - \frac{2v_2}{n_i} \sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_i} \theta_{p_1} K(x_{p_1}^j, x_{p_2}^i) - 2 \sum_{p_1=1}^{n_j} \sum_{p_2=1}^{n_{ij}} \theta_{p_1} \gamma_{p_2} K(x_{p_1}^j, x_{p_2}^{ij}) + \frac{2v_2}{n_i} \sum_{p_1=1}^{n_i} \sum_{p_2=1}^{n_{ij}} \gamma_{p_2} K(x_{p_1}^i, x_{p_2}^{ij}). \end{aligned} \quad (37)$$

For the testing sample x , the final label can be also determined by vote rule.

3.4. Analysis of Learning Complexity

Next, the learning complexity of the proposed Twin Hypersphere-KSVC will be discussed. We take the 4-class classification as an example, suppose samples of 4 classes are approximately equal, and present the learning complexity of K-SVCR, Twin-KSVC and Twin Hypersphere-KSVC in Table 1. The main calculating burden in Twin-KSVC includes solving QPPs and calculating inverse matrices. Therefore the learning complexities of linear and nonlinear Twin-KSVC are respectively $K(K-1)(O(d^3) + O((\frac{3}{4}n)^3))$ and $K(K-1)(O(n^3) + O((\frac{3}{4}n)^3))$. However, K-SVCR and Twin Hypersphere-KSVC avoid computing inverse matrices, the learning complexities of linear and nonlinear K-SVCR are both $\frac{K(K-1)}{2}O((\frac{3}{2}n)^3)$ while the learning complexities of linear and nonlinear Twin Hypersphere-KSVC are both $K(K-1)O((\frac{3}{4}n)^3)$. From the above analysis, we can see that our Twin Hypersphere-KSVC requires less learning time.

Table 1. The learning complexities of Twin-KSVC, K-SVCR and Twin Hypersphere-KSVC.

Kernel	Twin-KSVC	K-SVCR	Twin Hypersphere-KSVC
linear kernel	$K(K-1)O(d^3) + O((\frac{3}{4}n)^3)$	$\frac{K(K-1)}{2}O((\frac{3}{2}n)^3)$	$K(K-1)O((\frac{3}{4}n)^3)$
nonlinear kernel	$K(K-1)(O(n^3) + O((\frac{3}{4}n)^3))$	$\frac{K(K-1)}{2}O((\frac{3}{2}n)^3)$	$K(K-1)O((\frac{3}{4}n)^3)$

4. Experiments

In this section, we investigate classification performance of our Twin Hypersphere-KSVC, ITKSVC [31], THKSVM, Twin-KSVC and K-SVCR on a set of benchmark datasets from the UCI repository and several real engineering problems. The above algorithms are implemented in Matlab R2012a, and we use the “quadprog.m” function to solve the QPP and the “inv.m” function to calculate matrix inversion.

The parameter selection directly affects the classification performance of the above algorithms. We use the most popular exhaustive search to determine the parameters in this section. The K-SVCR

includes two penalty parameters $c_i (i = 1, 2)$ and bandwidth parameter ε . The Twin-KSVC has five parameters, including four penalty parameters $c_i (i = 1, 2, 3, 4)$ and bandwidth parameter ε . The THKSVM holds the penalty parameters c_1 and v_1 . The ITKSVC contains seven parameters which are six penalty parameters $c_i (i = 1, 2, 3, 4, 5, 6)$ and bandwidth parameter ε . There exist six penalty parameters $c_i (i = 1, 2, 3, 4)$ and $v_i (i = 1, 2)$ in our Twin Hypersphere-KSVC. The optimal values of penalty parameters c_i are searched from set $\{2^{-7}, \dots, 2^7\}$, penalty parameters v_i from $\{0.1, \dots, 0.9\}$ and bandwidth parameter ε from set $\{0, \dots, 0.5\}$.

4.1. Benchmark Datasets

We compare Twin Hypersphere-KSVC with ITKSVC, THKSVM, Twin-KSVC and K-SVCR on a set of benchmark datasets from the UCI repository in this subsection. The benchmark datasets are presented in Table 2. The 5-fold cross validation is used to estimate the testing accuracy and we use radial basis function $K(x, y) = e^{-\|x-y\|^2/\sigma^2}$ in this subsection, where the parameter σ are selected from $\{2^{-7}, \dots, 2^7\}$.

Table 2. The statistics of benchmark datasets.

Dataset	#Attributes	#Samples	#Classes
Wine	13	178	3
Iris	4	150	3
Ecoli	7	327	5
Soybean	35	47	4
Hayes-roth	5	132	3
Teaching-evaluation	5	151	3
Dermatology	34	358	6
Balance	4	625	3

The predicting accuracy and learning time of five algorithms on UCI benchmark data sets are respectively shown in Tables 3 and 4. By observing Tables 3 and 4, one can come to the following conclusions.

- (1) The Twin Hypersphere-KSVC, ITKSVC, Twin-KSVC and K-SVCR obtain better test accuracy than THKSVM. This is mainly because the sub-classifiers in Twin Hypersphere-KSVC, ITKSVC, Twin-KSVC and K-SVCR avoid the class imbalance problem appearing in THKSVM.
- (2) The Twin Hypersphere-KSVC works faster than Twin-KSVC, K-SVCR and ITKSVC. It is mainly because, compared with Twin-KSVC, the sub-classifiers in Twin Hypersphere-KSVC avoid calculating inverse matrices which appear in Twin-KSVC, while compared with K-SVCR and ITKSVC, each pair of sub-classifiers of our Twin Hypersphere-KSVC only needs to resolve two smaller QPPs.
- (3) For predicting accuracy of Twin-KSVC, K-SVCR, ITKSVC and Twin Hypersphere-KSVC, we can observe that not any method is superior to others for all data sets. We can apply Friedman test to analyze the test accuracy of these classifiers statistically [33,34]. The ranks of five algorithms for all data set in the light of test accuracy are presented in Table 5. The Friedman statistic χ_F^2 can be calculated by

$$\chi_F^2 = \frac{12m_2}{m_1(m_1 + 1)} \left[\sum_{k_1=1}^{m_1} rank_{k_1}^2 - \frac{m_1(m_1 + 1)^2}{4} \right], \quad (38)$$

where $rank_{k_1} = \frac{1}{m_2} \sum_{k_2=1}^{m_2} r_{k_1}^{k_2}$, $r_{k_1}^{k_2}$ denotes the rank of the k_1 -th of m_1 classifiers on the k_2 -th of m_2 data sets. Because χ_F^2 is undesirably conservative, we can use the other statistic

$$F_F = \frac{(m_2 - 1)\chi_F^2}{m_2(m_1 - 1) - \chi_F^2}, \quad (39)$$

which is distributed by the $F(m_1 - 1, (m_1 - 1)(m_2 - 1))$.

We can calculate the statistic $F_F = 10.41$, where $F_F \sim F(4, 28)$. For the level of significance $\alpha = 0.05$, $F(4, 28) = 2.95$ is smaller than F_F , which means there are significant differences among these classifiers. We can see from Table 5 that the average rank of Twin Hypersphere-KSVC is lower than ITKSVC, and is higher than Twin-KSVC and K-SVCR. It implies that classification accuracy of Twin Hypersphere-KSVC is slightly lower than ITKSVC, however, is much higher than Twin-KSVC and K-SVCR.

Table 3. Classification accuracy of THKSVM, ITKSVC, Twin-KSVC, K-SVCR and Twin Hypersphere-KSVC on UCI benchmark datasets.

Dataset	Twin-KSVC(%)	THKSVM(%)	K-SVCR(%)	ITKSVC(%)	Twin Hypersphere-KSVC(%)
Wine	97.29 ± 2.28	92.24 ± 10.22	97.96 ± 2.47	97.53 ± 2.20	98.30 ± 2.55
Iris	93.47 ± 3.67	91.20 ± 4.52	95.47 ± 3.07	95.53 ± 4.03	96.00 ± 3.06
Ecoli	84.33 ± 4.60	77.75 ± 2.80	87.34 ± 3.07	85.63 ± 3.70	83.04 ± 4.21
Soybean	98.00 ± 4.47	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
Hayes-roth	61.07 ± 7.90	39.40 ± 2.05	54.02 ± 8.31	71.28 ± 9.92	52.50 ± 8.21
Teaching-evaluation	56.43 ± 8.23	50.28 ± 11.71	51.91 ± 8.34	52.02 ± 6.27	59.44 ± 7.52
Dermatology	95.43 ± 2.19	70.18 ± 3.16	97.66 ± 1.82	94.19 ± 2.74	96.64 ± 2.39
Balance	96.92 ± 1.29	90.05 ± 2.94	94.37 ± 1.96	97.73 ± 1.19	90.91 ± 1.34

The bolded classification accuracy is the highest one in all multi-classifiers.

Table 4. Learning time of THKSVM, ITKSVC, Twin-KSVC, K-SVCR and Twin Hypersphere-KSVC on UCI benchmark datasets.

Dataset	Twin-KSVC(s)	THKSVM(s)	K-SVCR(s)	ITKSVC(s)	Twin Hypersphere-KSVC(s)
Wine	2.1985	0.2749	2.5362	1.8673	1.1457
Iris	1.7260	0.2533	2.0569	1.5347	0.7429
Ecoli	38.7993	1.0329	62.2968	27.2987	15.1017
Soybean	1.1527	0.1787	0.4666	0.6355	0.5374
Hayes-roth	2.4358	0.1848	1.7696	1.1364	0.5671
Teaching-evaluation	3.6512	0.1849	1.7180	1.2330	1.0899
Dermatology	114.424	1.0811	89.2057	41.0467	19.6405
Balance	45.2907	3.7804	68.7381	43.7113	13.5046

Table 5. The rank of five algorithms on UCI benchmark datasets in the light of classification accuracy.

Dataset	Twin-KSVC	THKSVM	K-SVCR	ITKSVC	Twin Hypersphere-KSVC
Wine	4	5	2	3	1
Iris	4	5	3	2	1
Ecoli	3	5	1	2	4
Soybean	5	2.5	2.5	2.5	2.5
Hayes-roth	2	5	3	1	4
Teaching-evaluation	2	5	4	3	1
Dermatology	3	5	1	4	2
Balance	2	5	3	1	4
Average	3.13	4.69	2.81	2.31	2.44

4.2. Handwritten Digits Recognition

We use Twin Hypersphere-KSVC to recognize handwritten digits. The USPS database is used to compare our Twin Hypersphere-KSVC with Twin-KSVC, THKSVM, ITKSVC and K-SVCR. The USPS dataset consists of 8-bit grayscale images of handwritten digits from 0 to 9, as presented in Figure 2. We choose 55 images for each handwritten digit in the USPS database, 550 images in total. We only consider linear kernel function $K(x, y) = x^T y$, and also use 5-fold cross validation to estimate the testing accuracy.



Figure 2. Illustration of 10 digits in the USPS dataset.

The handwritten digits recognition results of five algorithms are presented in Table 6. From Table 6, we can observe that, the proposed Twin Hypersphere-KSVC obtains better accuracy among all algorithms. In terms of learning time, our Twin Hypersphere-KSVC costs shorter learning time, compared with ITKSVC, Twin-KSVC and K-SVCR.

Table 6. Classification results of Twin Hypersphere-KSVC, Twin-KSVC, THKSVM, K-SVCR and ITKSVC for handwritten digits recognition on the USPS dataset.

Algorithms	Twin-KSVC	THKSVM	K-SVCR	ITKSVC	Twin Hypersphere-KSVC
Accuracy (%)	53.13 \pm 5.38	72.55 \pm 3.55	85.09 \pm 2.99	77.18 \pm 4.24	85.75 \pm 2.15
Learning time (s)	377.5463	2.1101	641.946	348.6982	342.983

4.3. Text Classification

We evaluate our Twin Hypersphere-KSVC to text classification in this subsection and compare it with the other algorithms on the Reuters21578 dataset. We choose 6 classes from the Reuters21578 dataset, 708 documents in total, which are presented in Table 7. We also consider linear kernel function in this subsection.

Table 7. The statistics of text classification dataset.

Label	Cocoa	Coffee	Corn	Cice	Cubber	Soybean
Training dataset	52	101	176	46	35	83
Test dataset	23	44	76	21	15	36

The experimental results of five algorithms for text classification are presented in Table 8. By observing Table 8, we can notice that, our Twin Hypersphere-KSVC can get better accuracy among all multi-classifiers. The proposed Twin Hypersphere-KSVC runs faster in comparison with ITKSVC, Twin-KSVC and K-SVCR.

Table 8. Classification results of Twin Hypersphere-KSVC, Twin-KSVC, THKSVM, K-SVCR and ITKSVC for text classification on the Reuters21578 dataset.

Algorithms	Twin-KSVC	THKSVM	K-SVCR	ITKSVC	Twin Hypersphere-KSVC
F1	0.4474	0.5932	0.7076	0.7078	0.7414
Learning time (s)	36.9486	0.4758	68.7947	26.4216	11.9575

5. Conclusions

In this paper, we propose a novel multi-class classification algorithm, named Twin Hypersphere-KSVC. The Twin Hypersphere-KSVC evaluates each training sample into 1-vs-1-vs-rest structure, as in Twin-KSVC and K-SVCR, and constructs two hyperspheres in each pair of sub-classifiers, instead of two nonparallel hyperplanes as in Twin-KSVC. Compared with Twin-KSVC, the sub-classifiers in Twin Hypersphere-KSVC avoid computing inverse matrices, and for nonlinear problems, can apply the kernel trick to linear case directly. The classification results on a set of

benchmark datasets from UCI repository, handwritten digits recognition and text classification, show that the Twin Hypersphere-KSVC gets better classification performance in comparison with the other classical multi-classifiers.

Author Contributions: Investigation, Q.A., Y.W.; Methodology, Q.A., A.W.; Software, Q.A.; Validation, Q.A., W.W.; Supervision, A.W., A.Z.; Funding acquisition, Q.A., A.W.; Writing-original draft, Q.A.; Writing—review and editing, Q.A.

Funding: This research was funded by Natural Science Foundation of Liaoning province in China (20180551048 and 201601291) and Talent Cultivation Project of University of Science and Technology Liaoning in China (2018RC05).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2000; pp. 17–34.
2. Zhang, X. Introduction to statistical learning theory and support vector machines. *Acta Automatica Sinica* **2000**, *26*, 32–42.
3. Long, W.; Song, L.; Tian, Y. A new graphic kernel method of stock price trend prediction based on financial news semantic and structural similarity. *Expert Syst. Appl.* **2019**, *118*, 411–424. [[CrossRef](#)]
4. Lei, C.; Deng, J.; Cao, K.; Xiao, Y.; Ma, L.; Wang, W.; Ma, T.; Shu, C. A comparison of random forest and support vector machine approaches to predict coal spontaneous combustion in gob. *Fuel* **2019**, *239*, 297–311. [[CrossRef](#)]
5. Zhao, X.; Zang, W.; Lv, R.; Cui, W. Effective information filtering mining of internet of brain things based on support vector machine. *IEEE Access* **2019**, *7*, 191–202. [[CrossRef](#)]
6. Xie, F.; Li, F.; Lei, C.; Yang, J.; Zhang, Y. Unsupervised band selection based on artificial bee colony algorithm for hyperspectral image classification. *Appl. Soft Comput.* **2019**, *75*, 428–440. [[CrossRef](#)]
7. Qiao, X.; Bao, J.; Zhang, H.; Wan, F.; Li, D. fvUnderwater sea cucumber identification based on Principal Component Analysis and Support Vector Machine. *Measurement* **2019**, *133*, 444–455. [[CrossRef](#)]
8. Maltarollo, V.G.; Kronenberger, T.; Espinoza, G.Z.; Oliveira, P.R.; Honorio, K.M. Advances with support vector machines for novel drug discovery. *Expert Opin. Drug Discov.* **2019**, *14*, 23–33. [[CrossRef](#)]
9. Jayadeva.; Khemchandani, R.; Chandra, S. Twin support vector machines for pattern classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 905–910. [[CrossRef](#)]
10. Mir, A.; Nasiri, J.A. KNN-based least squares twin support vector machine for pattern classification. *Appl. Intell.* **2018**, *48*, 4551–4564. [[CrossRef](#)]
11. Wang, H.; Zhou, Z.; Xu, Y. An improved ν -twin bounded support vector machine. *Appl. Intell.* **2018**, *48*, 1041–1053. [[CrossRef](#)]
12. Shao, Y.; Zhang, C.; Wang, X.; Deng, N. Improvements on twin support vector machines. *IEEE Trans. Neural Netw.* **2011**, *22*, 962–968. [[CrossRef](#)] [[PubMed](#)]
13. Qi, Z.; Tian, Y.; Shi, Y. Structural twin support vector machine for classification. *Knowl. Based Syst.* **2013**, *43*, 74–81. [[CrossRef](#)]
14. Tian, Y.; Qi, Z.; Ju, X.; Shi, Y.; Liu, X. Nonparallel support vector machines for pattern classification. *IEEE Trans. Syst. Man Cybern.* **2014**, *44*, 1067–1079. [[CrossRef](#)]
15. Wang, Z.; Shao, Y.; Bai, L.; Deng, N. Twin support vector machine for clustering. *IEEE Trans. Neural Netw.* **2015**, *26*, 2583–2588. [[CrossRef](#)]
16. Ye, Q.; Zhao, C.; Gao, S.; Zheng, H. Weighted twin support vector machines with local information and its application. *Neural Netw.* **2012**, *35*, 31–39. [[CrossRef](#)]
17. Peng, X.; Xu, D. Bi-density twin support vector machines for pattern recognition. *Neurocomputing* **2013**, *99*, 134–143. [[CrossRef](#)]
18. Chen, S.; Wu, X. A new fuzzy twin support vector machine for pattern classification. *Int. J. Mach. Learn. Cybern.* **2018**, *9*, 1553–1564. [[CrossRef](#)]
19. Xu, Y.; Yang, Z.; Pan, X. A novel twin support-vector machine with pinball loss. *IEEE Trans. Neural Netw.* **2017**, *28*, 359–370. [[CrossRef](#)]

20. Chen, W.; Shao, Y.; Li, C.; Deng, N. MLTSVM: A novel twin support vector machine to multi-label learning. *Pattern Recognit.* **2016**, *52*, 61–74. [[CrossRef](#)]
21. Tang, J.; Li, D.; Tian, Y.; Liu, D. Multi-view learning based on nonparallel support vector machine. *Knowl. Based Syst.* **2018**, *158*, 94–108. [[CrossRef](#)]
22. Tang, L.; Tian, Y.; Yang, C. Nonparallel support vector regression model and its SMO-type solver. *Neural Netw.* **2018**, *105*, 431–446. [[CrossRef](#)] [[PubMed](#)]
23. Xie, X. Improvement on projection twin support vector machine. *Neural Comput. Appl.* **2018**, *30*, 371–387. [[CrossRef](#)]
24. Tang, L.; Tian, Y.; Yang, C.; Pardalos, P.M. Ramp-loss nonparallel support vector regression: Robust, sparse and scalable approximation. *Knowl. Based Syst.* **2018**, *147*, 55–67. [[CrossRef](#)]
25. Peng, X.; Xu, D. A twin-hypersphere support vector machine classifier and the fast learning algorithm. *Inf. Sci.* **2013**, *221*, 12–27. [[CrossRef](#)]
26. Xu, Y.; Wang, Q.; Pang, X.; Tian, Y. Maximum margin of twin spheres machine with pinball loss for imbalanced data classification. *Appl. Intell.* **2018**, *48*, 23–34. [[CrossRef](#)]
27. Peng, X.; Shen, J. A twin-hyperspheres support vector machine with automatic variable weights for data classification. *Inf. Sci.* **2017**, *417*, 216–235. [[CrossRef](#)]
28. Ai, Q.; Wang, A.; Wang, Y.; Sun, H. Improvements on twin-hypersphere support vector machine using local density information. *Prog. Artif. Intell.* **2018**, *7*, 167–175. [[CrossRef](#)]
29. Angulo, C.; Parra, X.; Catala, A. K-SVCR. A support vector machine for multi-class classification. *Neurocomputing* **2003**, *55*, 57–77. [[CrossRef](#)]
30. Xu, Y.; Guo, R.; Wang, L. A twin multi-class classification support vector machine. *Cogn. Comput.* **2013**, *5*, 580–588. [[CrossRef](#)]
31. Ai, Q.; Wang, A.; Wang, Y.; Sun, H. An improved Twin-KSVC with its applications. *Neural Comput. Appl.* **2018**. [[CrossRef](#)]
32. Xu, Y.; Guo, R. A twin hyper-sphere multi-class classification support vector machine. *J. Intell. Fuzzy Syst.* **2014**, *27*, 1783–1790.
33. Ar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
34. García, S.; Fernández, A.; Luengo, J.; Herrera, F. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Inf. Sci.* **2010**, *180*, 2044–2064. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).