

Intuitionistic Fuzzy Laplacian Twin Support Vector Machine for Semi-supervised Classification

Zhou, Jia Bin; Bai, Yan Qin; Guo, Yan Ru; Lin, Hai Xiang

DOI

[10.1007/s40305-021-00354-9](https://doi.org/10.1007/s40305-021-00354-9)

Publication date

2021

Document Version

Final published version

Published in

Journal of the Operations Research Society of China

Citation (APA)

Zhou, J. B., Bai, Y. Q., Guo, Y. R., & Lin, H. X. (2021). Intuitionistic Fuzzy Laplacian Twin Support Vector Machine for Semi-supervised Classification. *Journal of the Operations Research Society of China*, 10(1), 89-112. <https://doi.org/10.1007/s40305-021-00354-9>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Intuitionistic Fuzzy Laplacian Twin Support Vector Machine for Semi-supervised Classification

Jia-Bin Zhou¹ · Yan-Qin Bai¹ · Yan-Ru Guo¹ · Hai-Xiang Lin²

Received: 4 November 2020 / Revised: 16 April 2021 / Accepted: 19 May 2021
© The Author(s) 2021, corrected publication 2021

Abstract

In general, data contain noises which come from faulty instruments, flawed measurements or faulty communication. Learning with data in the context of classification or regression is inevitably affected by noises in the data. In order to remove or greatly reduce the impact of noises, we introduce the ideas of fuzzy membership functions and the Laplacian twin support vector machine (Lap-TSVM). A formulation of the linear intuitionistic fuzzy Laplacian twin support vector machine (IFLap-TSVM) is presented. Moreover, we extend the linear IFLap-TSVM to the nonlinear case by kernel function. The proposed IFLap-TSVM resolves the negative impact of noises and outliers by using fuzzy membership functions and is a more accurate reasonable classifier by using the geometric distribution information of labeled data and unlabeled data based on manifold regularization. Experiments with constructed artificial datasets, several UCI benchmark datasets and MNIST dataset show that the IFLap-TSVM has better classification accuracy than other state-of-the-art twin support vector machine (TSVM), intuitionistic fuzzy twin support vector machine (IFTSVM) and Lap-TSVM.

This work was supported by the National Natural Science Foundation of China (No.11771275). The second author thanks the partially support of Dutch Research Council (No.040.11.724).

✉ Yan-Qin Bai
yqbai@shu.edu.cn

Jia-Bin Zhou
jiab_zhou@163.com

Yan-Ru Guo
Guoyanru211@163.com

Hai-Xiang Lin
H.X.Lin@tudelft.nl

¹ Department of Mathematics, Shanghai University, Shanghai 200444, China

² Delft Institute of Applied Mathematics, Delft University of Technology, Delft 2600GA, The Netherlands

Keywords Twin support vector machine · Semi-supervised classification · Intuitionistic fuzzy · Manifold regularization · Noisy data

Mathematics Subject Classification 68T99 · 90C20

1 Introduction

Support vector machine (SVM) was proposed in details by Vapnik et al. [1]. The goal of SVM was to find an optimal hyperplane to separate the labeled data points into two classes. Because of its excellent performance in text classification tasks [2], it soon became the mainstream technology of machine learning. At present, SVM and its variants have been successfully applied in many fields such as face recognition [3], financial distree prediction [4], regression [5], traffic flow prediction [6], medical [7] and more. Proximal support vector machine (PSVM) [8,9] was derived from SVM; it aimed to find two parallel hyperplanes so that each plane was closer to one of two classes and as far away from the other as possible. Furthermore, in order to simplify the constraints, the generalized eigenvalue proximal support vector machine (GEPSVM) [10] was proposed. The main idea of GEPSVM was to replace two parallel hyperplanes with two nonparallel ones. According to this concept, Jayadeva et al. [11] proposed a well-known twin support vector machine (TSVM). Unlike the large quadratic programming problem (QPP) considered by traditional SVM, TSVM solves a pair of relatively smaller QPPs. The constraints of each QPP are only related to the data points of each of the two classes. Therefore, TSVM not only keeps the advantages of SVM, but also trains four times faster than SVM. Based on TSVM, Shao et al. [12] proposed an imbalanced weighted Lagrangian twin support vector machine (WLTSM) for the imbalanced data classification. Other extensions and applications of TSVM can be found in [13,14].

Recently, the research of semi-supervised learning (SSL) [15–17] has become a new hotspot in the field of machine learning. The main reason was that in many practical problems, labeled data are always scarce, but there are large amount of unlabeled data. SSL was to use these unlabeled data to assist a small number of labeled data for learning, so as to improve the performance of classifier. Manifold regularization (MR) [18,19] was one of the frameworks of SSL. In the MR framework, there are two regularization terms. One controls the complexity of classifier in the Reproducing Kernel Hilbert Spaces (RKHS), and the other controls the complexity as measured by the geometry of the distribution. Following the MR framework, Qi et al. [20] proposed a Laplacian twin support vector machine (Lap-TSVM), which was the first twin support vector machine applied in the SSL problem. Extensive experimental results show that Lap-TSVM has very good performance in semi-supervised classification. Other extensions and applications of semi-supervised twin support vector machine can be found in [21,22].

In general, data contain noises which come from faulty instruments, flawed measurements or faulty communication. Learning with data in the context of classification or regression is inevitably affected by noises in the data. If the training samples are mixed by noises, both SVM and its variants are often unable to find an optimal hyper-

plane and subsequently have difficulty to obtain satisfactory results. In order to solve such problem, fuzzy support vector machine (FSVM) [23] was proposed. The idea of FSVM was to use a membership function for each training sample. And the introduction of membership function can effectively reduce the effects of noises and outlier points and thus produce a robust classifier. Moreover, combining the TSVM with membership function can not only improve computational efficiency but also pursue robust performance. In recent years, intuitionistic fuzzy twin support vector machine (IFTSVM) [24] has been proposed which assigns a pair of membership and nonmembership functions to every training sample. These two functions help the IFTSVM to reduce the influence of noises and identify support vectors from noises.

The same difficulty was also encountered by the current semi-supervised twin support vector machine and its variants. When there are many noises in the data, the classification results are very poor and unsatisfactory. Ideally, we would like to determine which points are noisy, and then either remove them or greatly lower their weight. Therefore, inspired by the ideas of IFTSVM, we assign a pair of membership functions to each labeled point, which reduces the influence of noises on the classifier. And we introduce the ideas of fuzzy membership functions and the Lap-TSVM. In this paper, we proposed a novel intuitionistic fuzzy Laplacian twin support vector machine (IFLap-TSVM) for a semi-supervised classification problem. We use some constructed tests and several real datasets to evaluate the effectiveness of the IFLap-TSVM. The main advantages of our IFLap-TSVM are:

- (1) Membership and nonmembership functions are used for each training sample to indicate the contributions of different training samples to the learning of decision functions, which significantly reduces the negative impact of noises and outliers on classification accuracy.
- (2) Intuitionistic fuzzy number can reduce the influence of noises and outliers in labeled samples, and the semi-supervised framework of manifold regularization was introduced to deal with labeled and unlabeled samples in the primal space and the feature space. The combination of the two can further improve the classification accuracy.
- (3) IFLap-TSVM has better classification accuracy compared with other state-of-the-art TSVM, IFTSVM and Lap-TSVM on constructed tests and real-world datasets.

The remaining parts of this paper are organized as follows. In Sect. 2, we briefly introduce the background of SSL and Lap-TSVM. In Sect. 3, we describe the details of IFLap-TSVM. In Sect. 4, the numerical experiment results on the constructed test dataset, UCI dataset and MNIST dataset are reported. And Sect. 5 concludes the paper.

2 Background

In this section, we give a brief description of semi-supervised learning framework (SSL) and Lap-TSVM. The training data of the classification problem can be described as follows:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), x_{l+1}, \dots, x_{l+u}\}, \quad (1)$$

where $x_i \in \mathbb{R}^n$, $y_i = \{+1, -1\}$, $i = 1, 2, \dots, l$, are the labeled data, and x_i , $i = l+1, \dots, l+u$, are the unlabeled data. Denote the matrix $A \in \mathbb{R}^{l_1 \times n}$ as the labeled data belonging to class $+1$, where every row of matrix A represents a data point. Similarly, the matrix $B \in \mathbb{R}^{l_2 \times n}$ as the labeled data belonging to class -1 . Clearly, we have $l_1 + l_2 = l$.

2.1 Semi-supervised Learning Framework

SSL uses both labeled and unlabeled data to improve supervised learning. The goal is to build a more efficient classifier with large amounts of unlabeled data and relatively few labeled data. Regularization is a technique to prevent over fitting of training data, which is widely used in machine learning [25]. The MR framework takes advantage of the geometry of the probability distribution of the generated data and merges it as an additional regularization term. The decision function of the MR framework can be expressed as [18]

$$f^* = \arg \min_{f \in \mathcal{H}_K} \frac{1}{l} \sum_{i=1}^l V(x_i, y_i, f) + \gamma_A \|f\|_K^2 + \gamma_I \|f\|_I^2, \quad (2)$$

where f is an unknown decision function. The first part of the above expression is some loss function on the labeled data. The second part is a regularization term; γ_A is the weight of $\|f\|_K^2$ and controls the complexity of f in the Reproducing Kernel Hilbert Space. γ_I controls the complexity of the function in the intrinsic geometry of marginal distribution, and it is the weight of $\|f\|_I^2$, while $\|f\|_I^2$ is an appropriate penalty term that should reflect the intrinsic structure of marginal distribution.

The MR framework [18] incorporates additional information about the geometric structure of the marginal distribution. The important assumption of this approach is that the probability distribution of data has the geometry structure of a Riemannian manifold \mathcal{M} . If two points are very close in the intrinsic geometry, then they should have the same or similar labels. The RKHS regularization term $\|f\|_K^2$ and the intrinsic regularizer $\|f\|_I^2$ are as follows:

$$\|f\|_K^2 = \|f\|_2^2, \quad (3)$$

$$\begin{aligned} \|f\|_I^2 &= \frac{1}{(l+u)^2} \sum_{i,j=1}^{l+u} (f(x_i) - f(x_j))^2 W_{ij} \\ &= \frac{1}{(l+u)^2} f(\mathbf{X})^\top L f(X), \end{aligned} \quad (4)$$

where $f(X) = [f(x_1), \dots, f(x_{l+u})]$ represents the decision function values over labeled and unlabeled points. W_{ij} are edge weights in the data adjacency graph, L is the graph Laplacian given by $L = D - W$, $W \in \mathbb{R}^{(l+u) \times (l+u)}$ is the weight matrix with entries W_{ij} , D is the diagonal matrix with its i -th diagonal $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$. More detailed discussion of manifold regularization can be found in [18].

2.2 Laplacian Twin Support Vector Machine

Based on the TSVM, the Lap-TSVM [20] model is deduced by introducing the semi-supervised learning framework. According to the semi-supervised learning framework, for the linear case, the primal problems of Linear Lap-TSVM can be written as

$$\begin{aligned} \min_{w_1, b_1, \xi} \quad & \frac{1}{2} \|Aw_1 + e_1 b_1\|_2^2 + c_1 e_2^\top \xi + c_2 (\|w_1\|_2^2 + b_1^2) \\ & + c_3 (w_1^\top M^\top + e^\top b_1) L(Mw_1 + eb_1) \\ \text{s.t.} \quad & -(Bw_1 + e_2 b_1) + \xi \geq e_2, \quad \xi \geq 0, \end{aligned} \quad (5)$$

and

$$\begin{aligned} \min_{w_2, b_2, \eta} \quad & \frac{1}{2} \|Bw_2 + e_2 b_2\|_2^2 + c_1 e_1^\top \eta + c_2 (\|w_2\|_2^2 + b_2^2) \\ & + c_3 (w_2^\top M^\top + e^\top b_2) L(Mw_2 + eb_2) \\ \text{s.t.} \quad & (Aw_2 + e_1 b_2) + \eta \geq e_1, \quad \eta \geq 0. \end{aligned} \quad (6)$$

And W_{ij} are the edge weights in the data adjacency graph and may be defined by k -nearest neighbors or graph kernel as follows:

$$W_{ij} = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / 2\sigma^2), & \text{if } x_i, x_j \text{ are neighbor;} \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

$f_1 = [f_1(x_1), \dots, f_1(x_{l+u})]^\top = Mw_1 + eb_1$, $f_2 = [f_2(x_1), \dots, f_2(x_{l+u})]^\top = Mw_2 + eb_2$, $M \in R^{(l+u) \times n}$ includes all the training data, and e is an appropriate ones vector.

By introducing the Lagrangian multipliers, the Wolfe dual of the problem (5) and (6) can be formulated as

$$\begin{aligned} \max_{\alpha} \quad & e_2^\top \alpha - \frac{1}{2} \alpha^\top G (H^\top H + c_2 I + c_3 J^\top L J)^{-1} G^\top \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 e_2, \end{aligned} \quad (8)$$

and

$$\begin{aligned} \max_{\beta} \quad & e_1^\top \beta - \frac{1}{2} \beta^\top H (G^\top G + c_2 I + c_3 J^\top L J)^{-1} H^\top \beta \\ \text{s.t.} \quad & 0 \leq \beta \leq c_2 e_1. \end{aligned} \quad (9)$$

Here $G = [B \ e_2]$, $H = [A \ e_1]$ and $J = [M \ e]$. I is an identity matrix of appropriate dimension. It can be proved that $H^\top H + c_2 I + c_3 J^\top L J$ and $G^\top G + c_2 I + c_3 J^\top L J$ are positive definite matrices [26]. And the augmented vector v_1 , v_2 are given by

$$\begin{aligned} v_1 &= -(H^\top H + c_2 I + c_3 J^\top L J)^{-1} G^\top \alpha, & \text{where } v_1 &= [w_1^\top \ b_1]^\top, \\ v_2 &= (G^\top G + c_2 I + c_3 J^\top L J)^{-1} H^\top \beta, & \text{where } v_2 &= [w_2^\top \ b_2]^\top. \end{aligned} \quad (10)$$

The same as TSVM, the decision function of Lap-TSVM is as follows:

$$f(x) = \arg \min_{i \in 1,2} |w_i^\top x + b_i|, \quad (11)$$

where $|\cdot|$ is the perpendicular distance of point x from the planes $w_i^\top x + b_i$. For the nonlinear case, we refer to [20].

3 Intuitionistic Fuzzy Laplacian Twin Support Vector Machine

In this section, we first describe the concept of intuitionistic fuzzy set and then propose IFLap-TSVM model. The structures of two kernel functions, i.e., linear and nonlinear, are discussed in detail.

3.1 Intuitionistic Fuzzy Set

The traditional fuzzy set was given by Zadeh [27]. Let X be a nonempty set, the fuzzy set A in a universe X can be defined as

$$A = \{(x, \mu_A(x)) | x \in X\}, \quad (12)$$

where $\mu_A : X \rightarrow [0, 1]$ and $\mu_A(x)$ is the degree of membership of x belonging to X .

As an extension of fuzzy set, an intuitionistic fuzzy set [28] is defined as

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x), \nu_{\tilde{A}}(x)) | x \in X\}, \quad (13)$$

where $\mu_{\tilde{A}}(x)$ and $\nu_{\tilde{A}}(x)$ are the degrees of membership and nonmembership of x belonging to X . Here $\mu_{\tilde{A}} : X \rightarrow [0, 1]$, $\nu_{\tilde{A}}(x) : X \rightarrow [0, 1]$ and $0 \leq \mu_{\tilde{A}}(x) + \nu_{\tilde{A}}(x) \leq 1$. Define $\pi_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) - \nu_{\tilde{A}}(x)$; it denotes the hesitation degree of x belonging to X .

It is important to select an appropriate membership function to reduce the effect of noises and outlier points. For example, as shown in Fig. 1, the training points A and B are located on the boundary of the positive class, the degrees of membership of these two training points belonging to positive class are the same, but it is obvious that there are many negative points around point B. Therefore, the classification contribution of point A and point B is different. It may lead to wrong predictions if we only consider the membership degree. In this case, we employ the intuitionistic fuzzy number (μ, ν) to each training point as proposed in [29]. μ is the degree of membership function related to the one class, and ν is the degree of nonmembership function related to the other class. It is obvious that the points A and B in positive class have different degrees of nonmembership.

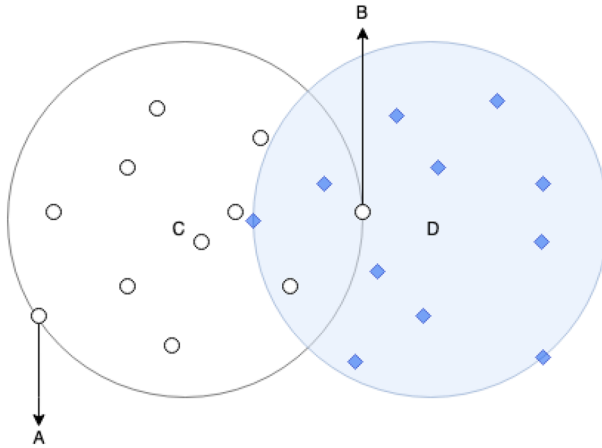


Fig. 1 Two training points with the same degree of membership

3.1.1 The Degree of Membership Function

In the high-dimensional feature space, the distance between training point and the class center is used as membership function. The distance between training points is expressed as

$$D(\phi(x_i), \phi(x_j)) = \|\phi(x_i) - \phi(x_j)\|, \quad (14)$$

where ϕ represents the mapping from the sample space to the high-dimensional feature space.

The class center of each class is given by

$$C^\pm = \frac{1}{l^\pm} \sum_{y_i = \pm 1} \phi(x_i), \quad (15)$$

where l_+ and l_- denote the total number of positive and negative points, respectively.

The radius of each class can be measured by

$$r^\pm = \max_{y_i = \pm 1} \|\phi(x_i) - C^\pm\|. \quad (16)$$

For each training point, the degree of membership can be defined as

$$\mu(x_i) = \begin{cases} 1 - \frac{\|\phi(x_i) - C^+\|}{r^+ + \delta}, & y_i = +1, \\ 1 - \frac{\|\phi(x_i) - C^-\|}{r^- + \delta}, & y_i = -1, \end{cases} \quad (17)$$

where $\delta > 0$ is an adjustable parameter.

3.1.2 The Degree of Nonmembership Function

The degree of nonmembership function is determined by the ratio between the number of all heterogeneous points and the total number of training points in its neighborhood. The degree of nonmembership function is as follows:

$$v(x_i) = (1 - \mu(x_i))\rho(x_i), \quad (18)$$

where $0 \leq \mu(x_i) + v(x_i) \leq 1$ and $\rho(x_i)$ is the proportion between all heterogeneous points and the total number of points in its neighborhood

$$\rho(x_i) = \frac{|\{x_j | |\phi(x_i) - \phi(x_j)| \leq \alpha, y_j \neq y_i\}|}{|\{x_j | |\phi(x_i) - \phi(x_j)| \leq \alpha\}|}, \quad (19)$$

where $|\cdot|$ denotes the cardinality and $\alpha > 0$ is an adjustable parameter.

The degree of membership and nonmembership of a training point is designed based on the inner product distance in the feature space. Therefore, the kernel functions are used to make the construction of intuitionistic fuzzy numbers.

3.1.3 The Score Function

Based on the above definitions, the training points can be converted into the intuitionistic fuzzy numbers as follows:

$$T = \{(x_1, y_1, \mu_1, v_1), (x_2, y_2, \mu_2, v_2), \dots, (x_l, y_l, \mu_l, v_l)\}, \quad (20)$$

where μ_i, v_i denote the degrees of membership functions and nonmembership functions of x_i , respectively. For each given intuitionistic fuzzy number, a score function can be used to measure the classification contribution of each training point. The score function can be defined as

$$s_i = \begin{cases} \mu_i, & v_i = 0, \\ 0, & \mu_i \leq v_i, \\ \frac{1 - v_i}{2 - \mu_i - v_i}, & \text{others.} \end{cases} \quad (21)$$

The score value s_i can easily distinguish the support vector from noises and outliers points. For example, when $v_i = 0$ (positive point A shown in Fig. 2), there are no negative points in the neighborhood of A; a correct degree of membership function can be easily defined. Obviously, the positive point A is far away from the class center so its classification contribution is small. When $\mu_i \leq v_i$ (negative point B shown in Fig. 2), the point B has no negative points in the neighborhood; the degree of nonmembership value is greater than the degree of membership value. Thus, B is a noise point with zero classification contribution. For the case of positive point C, we have $\mu_i > v_i$ and $v_i \neq 0$. C is far away from the class center, but there are some positive points in its

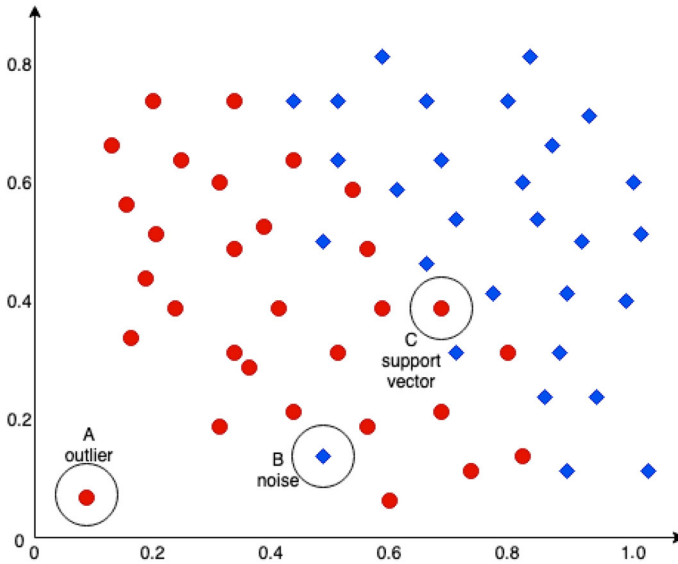


Fig. 2 Support vector, noise and outlier

neighborhood. Thus, it may be a support vector, instead of an outlier point. Hence, the classification contribution of point C is greater than that of outlier A.

3.2 Linear IFLap-TSVM

According to the semi-supervised learning framework, the square loss function and hinge loss function $V(x_i, y_i, f)$ can be expressed as

$$V_1(x_i, y_i, f_1) = ((A_{i,\cdot} \cdot w_1) + b_1)^2 + S_{2,i} \cdot \max(0, 1 - f_1(B_{i,\cdot})), \quad (22)$$

$$V_2(x_i, y_i, f_2) = ((B_{i,\cdot} \cdot w_2) + b_2)^2 + S_{1,i} \cdot \max(0, 1 - f_2(A_{i,\cdot})), \quad (23)$$

where $A_{i,\cdot}$ and $B_{i,\cdot}$ represent the i -th row of A and B , respectively. $S_{1,i}$ and $S_{2,i}$ denote the i -th element in the vector S_1 and S_2 , respectively. And $S_1 \in \mathbb{R}^{l+}$ and $S_2 \in \mathbb{R}^{l-}$ are the score values of positive and negative points, respectively.

The regularization terms $\|f_1\|_K^2$ and $\|f_2\|_K^2$ can be written as

$$\|f_1\|_K^2 = \frac{1}{2}(\|w_1\|_2^2 + b_1^2), \quad (24)$$

$$\|f_2\|_K^2 = \frac{1}{2}(\|w_2\|_2^2 + b_2^2). \quad (25)$$

And the manifold regularization terms $\|f_1\|_f^2$ and $\|f_2\|_f^2$ are defined by

$$\|f_1\|_f^2 = \frac{1}{(l+u)^2} f_1^\top L f_1, \quad (26)$$

$$\|f_2\|_l^2 = \frac{1}{(l+u)^2} f_2^\top L f_2. \quad (27)$$

In accordance with (2), the linear IFLap-TSVM can be written as

$$\begin{aligned} \min_{w_1, b_1, \xi} \quad & \frac{1}{2} \|Aw_1 + e_1 b_1\|_2^2 + c_1 S_2^\top \xi + \frac{1}{2} c_2 (\|w_1\|_2^2 + b_1^2) \\ & + \frac{1}{2} c_3 (Mw_1 + eb_1)^\top L (Mw_1 + eb_1) \\ \text{s.t.} \quad & -(Bw_1 + e_2 b_1) + \xi \geq e_2, \quad \xi \geq 0, \end{aligned} \quad (28)$$

and

$$\begin{aligned} \min_{w_2, b_2, \eta} \quad & \frac{1}{2} \|Bw_2 + e_2 b_2\|_2^2 + c_4 S_1^\top \eta + \frac{1}{2} c_5 (\|w_2\|_2^2 + b_2^2) \\ & + \frac{1}{2} c_6 (Mw_2 + eb_2)^\top L (Mw_2 + eb_2) \\ \text{s.t.} \quad & (Aw_2 + e_1 b_2) + \eta \geq e_1, \quad \eta \geq 0, \end{aligned} \quad (29)$$

where c_1, c_2, \dots, c_6 are pre-specified penalty factors, and ξ, η are slack variables, e_1, e_2, e are column vectors of ones of appropriate dimensions, L is the graph Laplacian.

The Lagrangian corresponding to the problem (28) is given by

$$\begin{aligned} L(w_1, b_1, \xi, \alpha, \beta) = & \frac{1}{2} \|Aw_1 + e_1 b_1\|_2^2 + c_1 S_2^\top \xi + \frac{1}{2} c_2 (\|w_1\|_2^2 + b_1^2) \\ & + \frac{1}{2} c_3 (Mw_1 + eb_1)^\top L (Mw_1 + eb_1) \\ & - \alpha^\top (-(Bw_1 + e_2 b_1) + \xi - e_2) - \beta^\top \xi, \end{aligned} \quad (30)$$

where $\alpha = (\alpha_1, \dots, \alpha_{l_2})^\top$ and $\beta = (\beta_1, \dots, \beta_{l_1})^\top$ are the Lagrangian multipliers.

With the KKT conditions, we get

$$\begin{aligned} \frac{\partial L}{\partial w_1} = & A^\top (Aw_1 + e_1 b_1) + c_2 w_1 \\ & + c_3 M^\top L (Mw_1 + eb_1) + B^\top \alpha = 0, \end{aligned} \quad (31)$$

$$\begin{aligned} \frac{\partial L}{\partial b_1} = & e_1^\top (Aw_1 + e_1 b_1) + c_2 b_1 \\ & + c_3 e^\top L (Mw_1 + eb_1) + e_2^\top \alpha = 0, \end{aligned} \quad (32)$$

$$\frac{\partial L}{\partial \xi} = c_1 S_2 - \alpha - \beta = 0. \quad (33)$$

Combining (31) and (32) leads to

$$\begin{aligned} & \begin{bmatrix} A^\top \\ e_1^\top \end{bmatrix} [A e_1] \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + c_2 \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} \\ & + c_3 \begin{bmatrix} M^\top \\ e^\top \end{bmatrix} L [M e] \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} B^\top \\ e_2^\top \end{bmatrix} \alpha = 0. \end{aligned} \quad (34)$$

Let $H = [A e_1]$, $J = [M e]$, $G = [B e_2]$, and the augmented vector $v_1 = [w_1 b_1]^\top$, then (34) can be rewritten as

$$\begin{aligned} & (H^\top H + c_2 I + c_3 J^\top L J) v_1 + G^\top \alpha = 0 \\ & \Rightarrow v_1 = -(H^\top H + c_2 I + c_3 J^\top L J)^{-1} (G^\top \alpha), \end{aligned} \quad (35)$$

where I is an identity matrix of appropriate dimensions. It can be proved that $H^\top H + c_2 I + c_3 J^\top L J$ is a positive definite matrix according to matrix theory [26].

Since $\beta \geq 0$, from (33), we get

$$0 \leq \alpha \leq c_1 S_2. \quad (36)$$

Therefore, the Wolfe dual of the problem (28) can be written as

$$\begin{aligned} & \max_{\alpha} \quad e_2^\top \alpha - \frac{1}{2} \alpha^\top G (H^\top H + c_2 I + c_3 J^\top L J)^{-1} G^\top \alpha \\ & \text{s.t.} \quad 0 \leq \alpha \leq c_1 S_2. \end{aligned} \quad (37)$$

Likewise, the dual of (29) is

$$\begin{aligned} & \max_{\beta} \quad e_1^\top \beta - \frac{1}{2} \beta^\top P (Q^\top Q + c_5 I + c_6 F^\top L F)^{-1} P^\top \beta \\ & \text{s.t.} \quad 0 \leq \beta \leq c_4 S_1. \end{aligned} \quad (38)$$

where $P = [A e_1]$, $F = [M e]$, $Q = [B e_2]$ and the augmented vector $v_2 = [w_2 b_2]^\top$ is written as follows:

$$v_2 = (Q^\top Q + c_5 I + c_6 F^\top L F)^{-1} P^\top \beta. \quad (39)$$

Once optimal v_1^* , v_2^* are achieved, the two hyperplanes are known. A new input data point x can be classified as positive or negative class based on the decision function

$$f(x) = \arg \min_{i \in 1,2} \frac{|w_i^\top x + b_i|}{\|w_i\|}, \quad (40)$$

where $|\cdot|$ is the absolute value. And the whole procedure of linear IF Lap-TSVM is described in Algorithm 1.

Algorithm 1 The linear IFLap-TSVM algorithm**Input:**

l_1 data points of class +1, l_2 data points of class -1, u unlabeled data points,
Test data point x .

Output:

Class label of the test data point.

Step 1 Set the value of the penalty factors c_1, c_2, c_3 and the graph kernel parameter σ .

Step 2 Compute the score functions S_1 and S_2 of data points in class +1 and -1.

Step 3 Compute graph Laplacian $L = D - W$ where W is a weight matrix given by

$$W_{ij} = \exp(-\|x_i - x_j\|^2 / 2\sigma^2) \text{ and } D \text{ is a diagonal matrix given by } D_{ii} = \sum_{j=1}^{l_1+u} W_{ij}.$$

Step 4 Determine parameters of the two non-parallel hyperplanes using (35) and (39).

Step 5 Determine the class label of the test data point based on (40).

3.3 Nonlinear IFLap-TSVM

So far the above discussion is restricted to the linear case. Here, we extend the linear IFLap-TSVM to the nonlinear case. We consider the following two kernel-generated hyperplanes:

$$K(x^\top, M^\top)\lambda_1 + b_1 = 0, \quad K(x^\top, M^\top)\lambda_2 + b_2 = 0, \quad (41)$$

where $K(x_i, x_j) = (\phi(x_i), \phi(x_j))$ is a chosen kernel function. The nonlinear optimization problem can be written as

$$\begin{aligned} \min_{\lambda_1, b_1, \xi} \quad & \frac{1}{2} \|K(A, M^\top)\lambda_1 + e_1 b_1\|_2^2 + c_1 S_2^\top \xi + c_2 (\lambda_1^\top K \lambda_1 + b_1^2) \\ & + c_3 (K \lambda_1 + e b_1)^\top L (K \lambda_1 + e b_1) \\ \text{s.t.} \quad & -(K(B, M^\top)\lambda_1 + e_2 b_1) + \xi \geq e_2, \quad \xi \geq 0, \end{aligned} \quad (42)$$

and

$$\begin{aligned} \min_{\lambda_2, b_2, \eta} \quad & \frac{1}{2} \|K(B, M^\top)\lambda_2 + e_2 b_2\|_2^2 + c_4 S_1^\top \eta + c_5 (\lambda_2^\top K \lambda_2 + b_2^2) \\ & + c_6 (K \lambda_2 + e b_2)^\top L (K \lambda_2 + e b_2) \\ \text{s.t.} \quad & (K(A, M^\top)\lambda_2 + e_1 b_2) + \eta \geq e_1, \quad \eta \geq 0. \end{aligned} \quad (43)$$

The Lagrangian corresponding to the problem (42) is given by

$$\begin{aligned} L(\lambda_1, b_1, \xi, \alpha, \beta) = & \frac{1}{2} \|K(A, M^\top)\lambda_1 + e_1 b_1\|_2^2 + c_1 S_2^\top \xi \\ & + \frac{1}{2} c_2 (\lambda_1^\top K \lambda_1 + b_1^2) \\ & + \frac{1}{2} (K \lambda_1 + e b_1)^\top L (K \lambda_1 + e b_1) \\ & - \alpha^\top (-(K(B, M^\top)\lambda_1 + e_2 b_1) + \xi - e_2) - \beta^\top \xi. \end{aligned} \quad (44)$$

The KKT conditions are obtained as follows:

$$\begin{aligned} \frac{\partial L}{\partial \lambda_1} &= K(A, M^\top)^\top (K(A, M^\top) \lambda_1 + e_1 b_1) + c_2 K \lambda_1 \\ &\quad + c_3 K^\top L(K \lambda_1 + e b_1) + K(B, M^\top)^\top \alpha = 0, \end{aligned} \quad (45)$$

$$\begin{aligned} \frac{\partial L}{\partial b_1} &= e_1^\top (K(A, M^\top) \lambda_1 + e_1 b_1) + c_2 b_1 \\ &\quad + c_3 e^\top L(K \lambda_1 + e b_1) + e_2^\top \alpha = 0, \end{aligned} \quad (46)$$

$$\frac{\partial L}{\partial \xi} = c_1 S_2 - \alpha - \beta = 0. \quad (47)$$

Combining (45) and (46) leads to

$$\begin{aligned} &\begin{bmatrix} K(A, M^\top)^\top \\ e_1^\top \end{bmatrix} \begin{bmatrix} K(A, M^\top) & e_1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ b_1 \end{bmatrix} + c_2 \begin{bmatrix} K & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ b_1 \end{bmatrix} \\ &+ c_3 \begin{bmatrix} K^\top \\ e^\top \end{bmatrix} L \begin{bmatrix} K & e \end{bmatrix} \begin{bmatrix} \lambda_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} K(B, M^\top)^\top \\ e_2^\top \end{bmatrix} \alpha = 0. \end{aligned} \quad (48)$$

Let $H_{\text{non}} = [K(A, M^\top) \ e_1]$, $O_{\text{non}} = \begin{bmatrix} K & 0 \\ 0 & 1 \end{bmatrix}$, $J_{\text{non}} = [K \ e]$, $G_{\text{non}} = [K(B, M^\top) \ e_2]$, and the augmented vector $v_{\text{non}1} = [\lambda_1 \ b_1]^\top$, then (48) can be rewritten as

$$\begin{aligned} &(H_{\text{non}}^\top H_{\text{non}} + c_2 O_{\text{non}} + c_3 J_{\text{non}}^\top L J_{\text{non}}) v_{\text{non}1} + G_{\text{non}}^\top \alpha = 0 \\ &\Rightarrow v_{\text{non}1} = -(H_{\text{non}}^\top H_{\text{non}} + c_2 O_{\text{non}} + c_3 J_{\text{non}}^\top L J_{\text{non}})^{-1} G_{\text{non}}^\top \alpha. \end{aligned} \quad (49)$$

Therefore, the Wolfe dual of the problem (42) can be written as

$$\begin{aligned} \max_{\alpha} \quad & e_2^\top \alpha - \frac{1}{2} \alpha^\top G_{\text{non}} (H_{\text{non}}^\top H_{\text{non}} + c_2 O_{\text{non}} + c_3 J_{\text{non}}^\top L J_{\text{non}})^{-1} G_{\text{non}}^\top \alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 S_2. \end{aligned} \quad (50)$$

Likewise, the dual of (43) is

$$\begin{aligned} \max_{\beta} \quad & e_1^\top \beta - \frac{1}{2} \beta^\top P_{\text{non}} (Q_{\text{non}}^\top Q_{\text{non}} + c_5 O_{\text{non}} + c_6 F_{\text{non}}^\top L F_{\text{non}})^{-1} P_{\text{non}}^\top \beta \\ \text{s.t.} \quad & 0 \leq \beta \leq c_4 S_1, \end{aligned} \quad (51)$$

where $P_{\text{non}} = [K(A, M^\top) \ e_1]$, $F_{\text{non}} = [K \ e]$, $Q_{\text{non}} = [K(B, M^\top) \ e_2]$ and the augmented vector $v_{\text{non}2} = [\lambda_2 \ b_2]^\top$ is follows:

$$v_{\text{non}2} = (Q_{\text{non}}^\top Q_{\text{non}} + c_5 O_{\text{non}} + c_6 F_{\text{non}}^\top L F_{\text{non}})^{-1} P_{\text{non}}^\top \beta. \quad (52)$$

Once the optimal $v_{\text{non}1}^*$, $v_{\text{non}2}^*$ are obtained, the two hyperplanes are known. A new input data point x can be classified as positive or negative class based on the decision function

$$f(x) = \arg \min_{i \in 1,2} \frac{|K(x, M^\top) \lambda_i + b_i|}{\sqrt{\lambda_i^\top K \lambda_i}}. \quad (53)$$

The whole procedure of nonlinear IFLap-TSVM is described in Algorithm 2.

Algorithm 2 The nonlinear IFLap-TSVM algorithm

Input:

l_1 data points of class +1, l_2 data points of class -1, u unlabeled data points,
Test data point x .

Output:

Class label of the test data point.

Step 1 Choose a kernel function $K(x, y)$. And set the value of the penalty factors c_1, c_2, c_3 and the kernel parameter σ .

Step 2 Compute the score functions S_1 and S_2 of data points in class +1 and -1.

Step 3 Compute graph Laplacian $L = D - W$, where W is a weight matrix given by $W_{ij} = \exp(-||x_i - x_j||^2 / 2\sigma^2)$ and D is a diagonal matrix given by $D_{ii} = \sum_{j=1}^{l_1+l_2} W_{ij}$.

Step 4 Determine parameters of the two non-parallel hyperplanes using (49) and (52).

Step 5 Determine the class label of the test data point based on (53).

4 Experiment

In this section, we investigate the effectiveness and generalization capability of the proposed method on artificial and UCI datasets, and we compare IFLap-TSVM with Lap-TSVM [20], IFTSVM [24] and TSVM [11].

The testing accuracies of all experiments are computed using standard 10-fold cross-validation [30]. The pre-specified penalty factors c_i ($i = 1, \dots, 6$) and the RBF kernel parameter σ are selected from the set $\{2^i | i = -5, \dots, 5\}$, and we set $c_1 = c_4, c_2 = c_5, c_3 = c_6$. In addition, Gaussian kernel is applied to deal with the nonlinear case, i.e., $K(x_1, x_2) = \exp(-||x_1 - x_2||^2 / \sigma^2)$. Each experiment is repeated 10 times. All the methods are implemented in MATLAB R2017b environment on a PC with Intel Core i5 processor with 8 GB RAM.

4.1 Artificial Datasets

In order to verify the validity of the model, two artificial datasets are constructed to evaluate IFLap-TSVM. We use two lines and half-moons containing 200 points as tests. And for the two lines dataset, we select a linear kernel, for the half-moons dataset, we select an RBF kernel. We choose 10 labeled points of each class as training set. And we inject different proportion of noise, i.e., 10% and 20%, into the training points. For example, 10% of the training points are randomly selected and their class are changed to another class.

4.1.1 The Impact of the Parameters

In this subsection, the effects of different setting of the parameters c_1 and c_3 are analyzed using the half-moons dataset. In the first experiment, we compare the performance of IFLap-TSVM and Lap-TSVM with different c_1 . For these two classifiers, we consider the half-moons dataset with 10% noise, and we fix the regularization parameters $c_2 = c_3 = 1$ and the RBF kernel parameter $\sigma = 0.5$, and then, let c_1 vary from 2^{-5} to 2^5 . Figure 3(a) shows the accuracy rates of the IFLap-TSVM and Lap-TSVM. IFLap-TSVM achieves the optimal accuracy when $c_1 = 2^2$, while Lap-TSVM obtains the best result when $c_1 = 2^0$. The value of c_1 corresponding to the best performance for IFLap-TSVM is larger than that of Lap-TSVM, because the score s_i of each training sample point in IFLap-TSVM is less than or equal to 1. In IFLap-TSVM, to achieve different levels of penalty, training samples are given different score values. The smaller the score value, the smaller the effect of training sample.

In the second experiment, in order to reflect the effectiveness of the manifold regularization, we fix $c_1 = c_2 = 1$ and the RBF kernel parameter $\sigma = 0.5$, and let c_3 vary from 2^{-5} to 2^{10} . It is easy to see from the results in Fig. 3(b) that with the increase in c_3 , the accuracy of IFLap-TSVM is also improved. However, the value of parameter c_3 should not be too large. When the value of c_3 exceeds 2^3 , the accuracy begins to decline drastically, and finally it will drop to 50%. The reason for this is that when the value of c_3 exceeds a certain limit, the manifold regularization will be penalized too much, which makes it lose its original function and make the model degenerate into a supervised model.

4.1.2 Comparison with Other Methods

In this subsection, we compare the effectiveness of our IFLap-TSVM with Lap-TSVM, IFTSVM and TSVM on the two lines and half-moons datasets. Figure 4 shows the one-run results of each classifier on the two lines. And Figs. 4(b)–4(d) show the

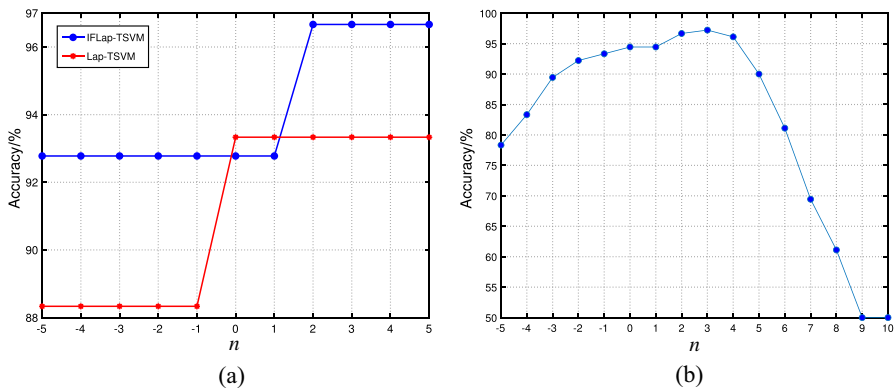


Fig. 3 (a) Comparison of IFLap-TSVM and Lap-TSVM on half-moons dataset with different values of parameter c_1 ($c_1 = 2^n$); (b) Accuracy of IFLap-TSVM on half-moons dataset with different values of parameter c_3 ($c_3 = 2^n$)

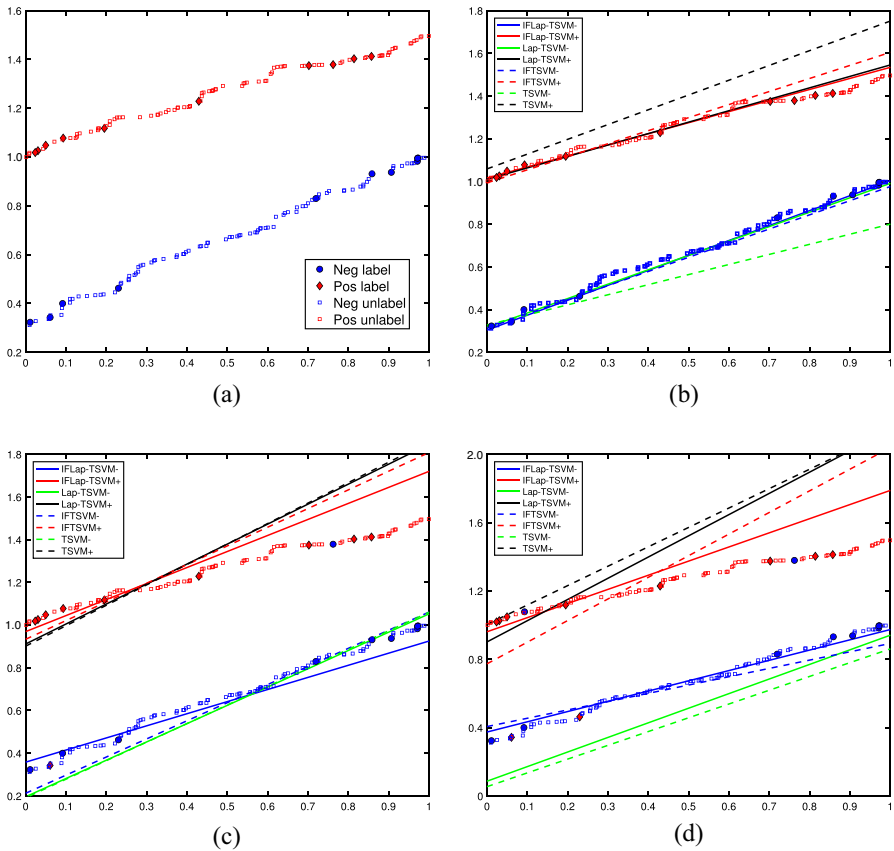


Fig. 4 (a) Original data points of the two lines without noise distortion. The classification results of TFSVM, IFTSVM, Lap-TSVM and IFLap-TSVM with different level of noise: (b) 0% noise; (c) 10% noise; (d) 20% noise

results of each classifier with a noise level of 0%, 10% and 20%, respectively. It can be seen from the results that with the increase in noise, our IFLap-TSVM can produce more accurate hyperplanes than other models.

And the one-run results of each classifier on the half-moons dataset with different level of noise are shown in Figs. 5, 6 and 7. It can be seen that compared with other methods, our IFLap-TSVM is more robust to noise, and the decision boundary is more accurate. In addition, for the half-moons dataset, we have done 10 experiments to further evaluate the classification results and the training time of the classifier, as shown in Table 1. The results show that with the increase in noise level, the accuracy of each method decreases. However, the effect of noise on the accuracy of IFLap-TSVM is the smallest. The training time of IFLap-TSVM is the longest, because compared with the supervised model, the objective function of dual QPP of semi-supervised model needs twice matrix inversion and the size of these matrices is $(n + 1) \times (n + 1)$. And IFLap-TSVM has more steps to calculate the score value of each training sample than Lap-TSVM.

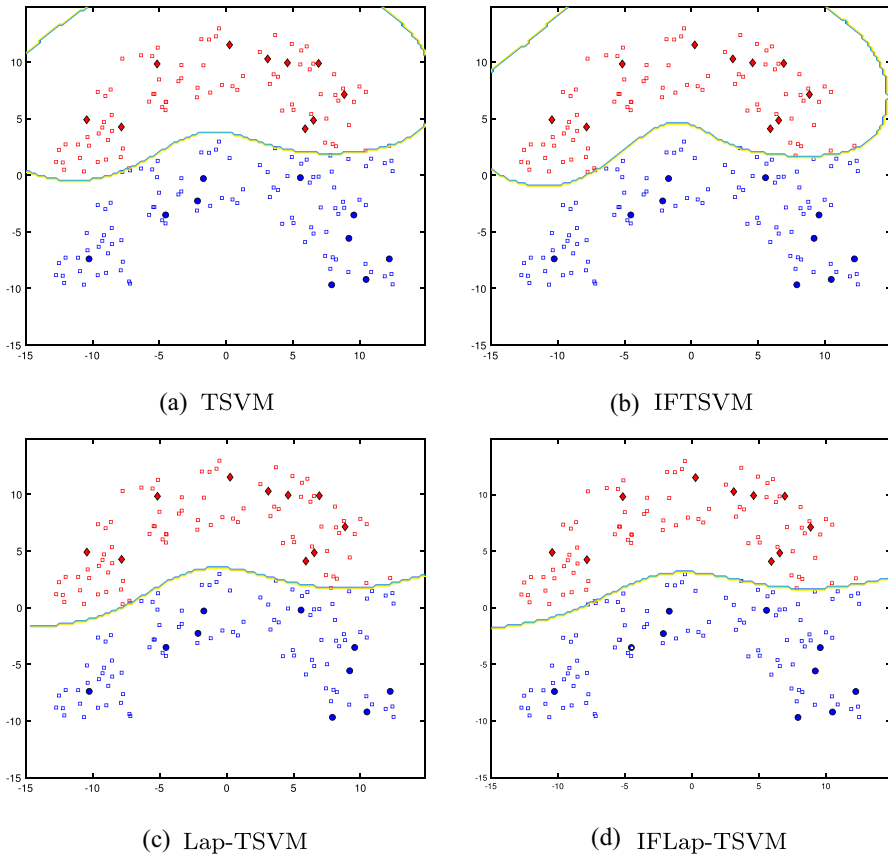


Fig. 5 Classification results of T SVM, IFT SVM, Lap-T SVM and IFLap-T SVM with 0% noise on half-moons dataset. (a) T SVM; (b) IFT SVM; (c) Lap-T SVM; (d) IFLap-T SVM

In general, compared with Lap-T SVM, IFT SVM and T SVM, IFLap-T SVM provides higher accuracy for both noiseless and noisy datasets. This is because IFLap-T SVM can use the information of unlabeled data to improve accuracy and use intuitionistic fuzzy number to reduce the effect of noises and outliers.

4.2 UCI Datasets

In this section, we investigate the performance of the IFLap-T SVM model on the UCI dataset [31]. And the results are compared with T SVM, Lap-T SVM and IFT SVM. Before training, all data are scaled such that all features locate in $[0, 1]$. First, each dataset is divided into two subsets: 65% for training and 35% for testing. Then, for each dataset, we randomly label m ($m = 10\%, 20\%, 30\%$) as labeled data and the remainder as unlabeled data. Table 2 shows the detailed information of the UCI dataset.

The classification accuracy and standard deviation of the IFLap-T SVM and other models are shown in Tables 3, 4 and 5. Tables 3, 4 and 5 show that with the increase

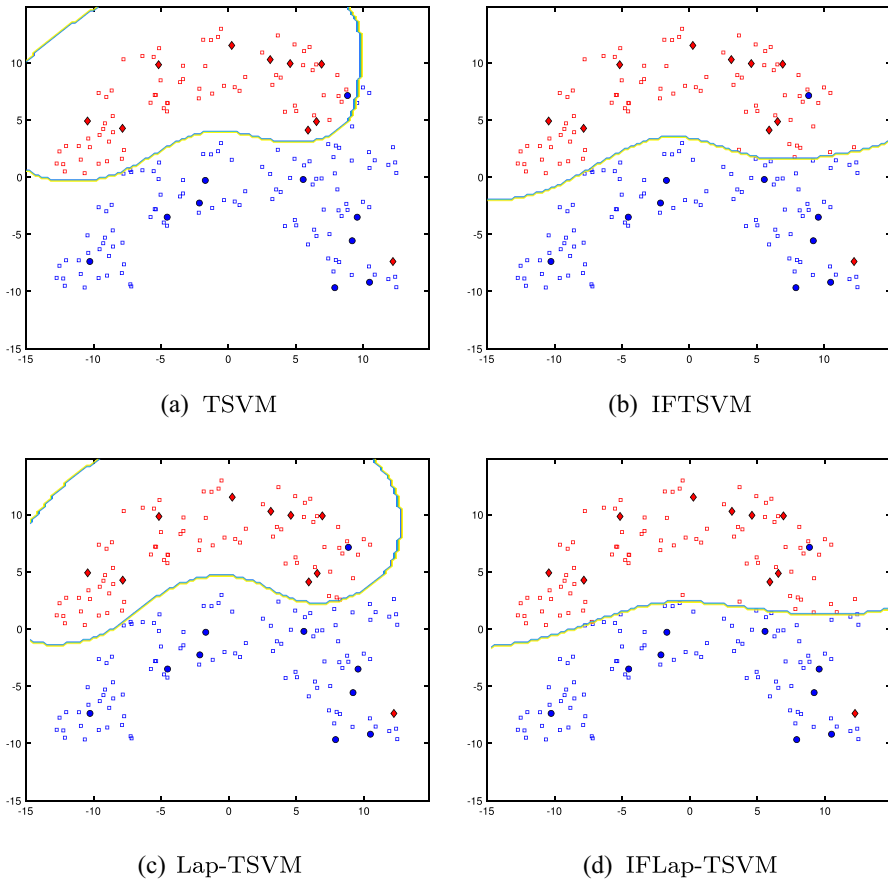


Fig. 6 Classification results of TSVM, IFTSVM, Lap-TSVM and IFLap-TSVM with 10% noise on half-moons dataset. **(a)** TSVM; **(b)** IFTSVM; **(c)** Lap-TSVM; **(d)** IFLap-TSVM

in the proportion of labeled data, the classification performance of all classifiers also increases. And from the average “mean” accuracy given in Tables 3, 4 and 5, the performance of IFLap-TSVM is better than other methods under the same labeled data.

Furthermore, IFLap-TSVM and Lap-TSVM have higher classification accuracy than IFTSVM and TSVM, which also shows that manifold regularization can help the classification model by using the geometric distribution information of labeled data and unlabeled data. More importantly, the accuracy of IFLap-TSVM is higher than that of Lap-TSVM, indicating that the intuitionistic fuzzy function is effective in reducing the effect of noise and outlier points.

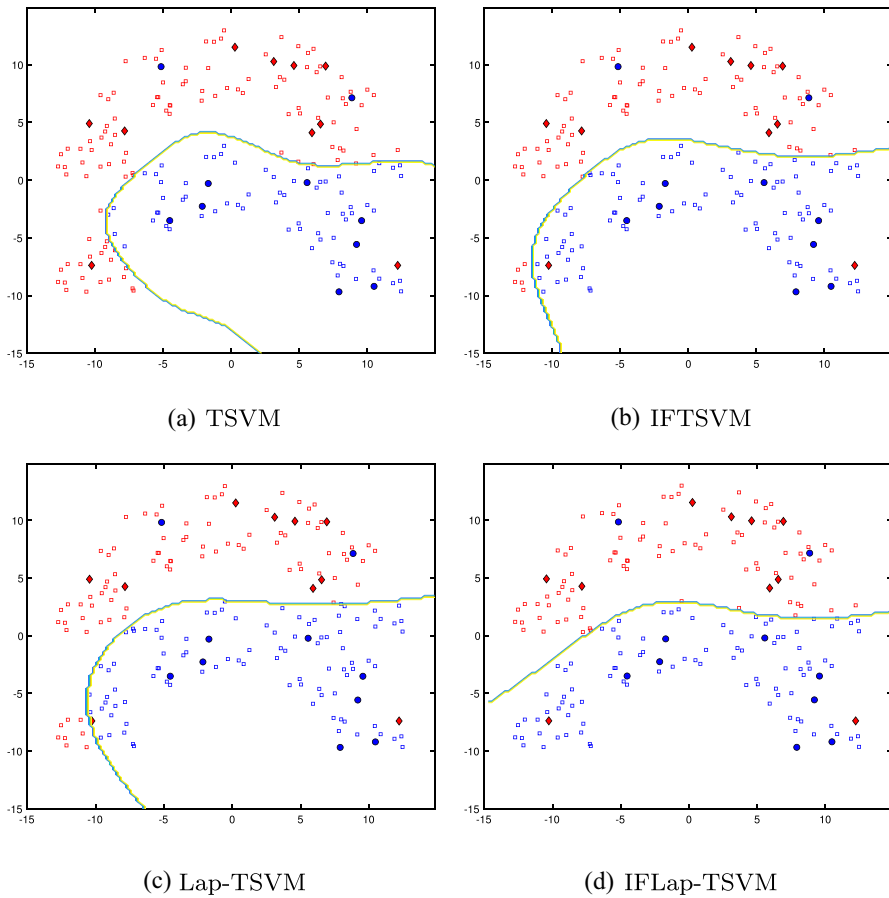


Fig. 7 Classification results of TSV, IFTSV, Lap-TSV and IFLap-TSV with 20% noise on half-moons dataset. (a) TSV; (b) IFTSV; (c) Lap-TSV; (d) IFLap-TSV

Table 1 Classification accuracy and the training time on the half-moons test with different levels of noise

Noise/%	0	10	20	Time/s
IFLap-TSV	97.06 ±1.79	96.28 ±1.97	94.96 ±1.92	0.090 4
Lap-TSV	96.50±3.06	93.11±4.96	90.83±5.33	0.082 4
IFTSV	95.89±1.55	94.77±1.62	91.88±1.53	0.081 7
TSV	95.28±3.68	91.44±4.50	87.06±5.25	0.078 5

Note: Bold represents the best accuracy of the four models

Table 2 Detailed information of UCI datasets

Dataset	Number	Attribute	Class
Hepatitis	155	19	2
Australian	690	14	2
BUPA	345	6	2
CMC	1 473	9	2
Ionosphere	351	34	2
Diabetes	768	8	2
Heart-Statlog	270	14	2
Credit	690	19	2
Spect	267	44	2
German	1 000	20	2

Table 3 Mean and standard deviation (%) of accuracy at 10% of labeled data points on UCI dataset

Dataset	IFLap-TSVM	Lap-TSVM	IFTSVM	TSVM
Hepatitis	72.81 ±2.56	71.46±4.86	71.64±1.75	70.18±4.79▲
Australian	67.40 ±1.36	65.83±4.62▲	65.25±1.88▲	63.97±4.02▲
BUPA	65.13±1.98	65.29 ±6.27	64.38±1.48	64.31±4.81
CMC	62.50 ±0.91	61.87±4.55	61.64±0.78	59.98±5.23▲
Ionosphere	74.39 ±1.28	73.09±4.39	70.73±0.87▲	68.86±3.96▲
Diabetes	60.97 ±1.26	60.34±3.43	59.56±1.26	59.77±4.03
Heart-Statlog	73.99 ±1.32	72.14±4.49▲	71.16±1.66▲	69.69±4.98▲
Credit	73.18 ±1.69	71.69±3.62	70.33±1.39▲	69.96±3.57▲
Spect	68.30±1.99	68.40 ±3.33	67.33±1.42	66.92±4.33
German	62.08 ±1.69	60.68±3.71	60.34±0.97▲	58.34±3.63▲
W/T/L	—	2/8/0	5/5/0	7/3/0
Ave.mean	68.08	67.08	66.24	65.20
Ave.std	1.60	4.33	1.35	4.34

Notes: (1) W/T/L (Win/Tie/Loss) denotes the number of datasets where IFLap-TSVM is significance Superior/Equal/ Inferior to the compared classifier; (2) Bold represents the best accuracy of the four models. ▲/▽ indicates whether IFLap-TSVM is statistically superior/inferior to the compared classifier

4.3 MNIST Dataset

In this section, we apply the IFLap-TSVM to handwritten symbol recognition. The MNIST dataset is shown in Fig. 8, which is a handwritten digital dataset composed of handwritten images from ‘0’ to ‘9’. The size of each image is 28×28 pixels with 256

Table 4 Mean and standard deviation (%) of accuracy at 20% of labeled data points on UCI dataset

Dataset	IFLap-TSVM	Lap-TSVM	IFTSVM	TSVM
Hepatitis	76.00 ±1.72	74.73±4.73	74.00±0.72▲	73.28±5.15▲
Australian	68.06 ±0.89	67.11±3.92	67.60±0.81	65.58±4.08▲
BUPA	66.61±1.66	67.93 ±5.93▽	65.70±0.79	65.12±4.77
CMC	64.84 ±1.34	64.32±3.91	63.01±0.67▲	61.32±3.92▲
Ionosphere	75.29 ±1.22	74.07±4.10	73.90±0.98	70.09±4.58▲
Diabetes	62.71 ±1.25	62.57±4.03	60.93±1.18▲	59.63±2.98▲
Heart-Statlog	74.42±1.22	74.53 ±4.46	72.53±1.04▲	70.42±4.22▲
Credit	75.99 ±2.04	74.30±4.56▲	73.14±1.39▲	72.31±4.16▲
Spect	71.91 ±1.82	71.81±4.74	69.79±1.52▲	68.72±3.92▲
German	65.37 ±1.26	62.34±3.82▲	62.69±1.05▲	60.09±3.91▲
W/T/L	–	2/7/1	7/3/0	9/1/0
Ave.mean	70.12	69.37	68.33	66.66
Ave.std	1.44	4.42	1.02	4.17

Notes: (1) W/T/L (Win/Tie/Loss) denotes the number of datasets where IFLap-TSVM is significance Superior/Equal/ Inferior to the compared classifier; (2) Bold represents the best accuracy of the four models. ▲/▽ indicates whether IFLap-TSVM is statistically superior/inferior to the compared classifier

Fig. 8 Samples from MINIST dataset

gray levels. Similar to [32], we select four pairwise digits on raw pixel features for comparison. And each set of pairwise digits contains 450 images, 300 images of which are used for training and another 150 images for testing. Furthermore, we randomly label 50 images for the training set, and m ($m = 50, 100, 150, 200, 250$) unlabeled images are selected from the remaining training images. In addition, we only consider these classifiers in the case of the RBF kernel.

Figure 9 shows the results of experiments. The results show that with the increase in unlabeled data, the test accuracies of semi-supervised learning classifiers are grad-

Table 5 Mean and standard deviation (%) of accuracy at 30% of labeled data points on UCI dataset

Dataset	IFLap-TSVM	Lap-TSVM	IFTSVM	TSVM
Hepatitis	77.82 ±1.43	76.18±5.31▲	74.91±0.98▲	73.82±4.55▲
Australian	69.01 ±1.20	67.52±4.45	68.35±1.25	66.28±4.71▲
BUPA	70.08 ±1.77	68.84±5.30	68.02±1.41	64.96±5.69▲
CMC	67.29 ±0.86	65.74±3.47▲	63.76±0.83▲	62.48±4.54▲
Ionosphere	76.42 ±1.48	74.96±4.33	75.61±1.08▲	72.36±4.29▲
Diabetes	64.54 ±1.07	63.75±4.73	61.90±1.04▲	60.48±3.69▲
Heart-Statlog	76.10±1.57	76.21 ±4.69	73.05±1.23▲	71.26±4.86▲
Credit	78.14 ±1.51	76.32±3.75▲	74.42±0.94▲	72.60±3.44▲
Spect	74.36±1.77	74.47 ±3.37	72.77±1.67▲	71.28±5.19▲
German	67.63 ±1.20	63.69±3.27▲	65.31±0.81▲	62.03±4.88▲
W/T/L	–	4/6/0	8/2/0	10/0/0
Ave.mean	72.14	70.77	69.81	67.76
Ave.std	1.38	4.27	1.12	4.58

Notes: (1) W/T/L (Win/Tie/Loss) denotes the number of datasets where IFLap-TSVM is significance Superior/Equal/ Inferior to the compared classifier; (2) Bold represents the best accuracy of the four models. ▲/▽ indicates whether IFLap-TSVM is statistically superior/inferior to the compared classifier

ually improved, because manifold regularization can use the geometric distribution information of labeled data and unlabeled data to find a more accurate classifier. In most cases, the classification results of IFLap-TSVM are better than other models, and the standard deviations of IFLap-TSVM and IFTSVM are smaller than that of Lap-TSVM and TSVM. Therefore, intuitionistic fuzzy can effectively reduce the impact of noise and outliers on classification accuracy.

5 Conclusion

In this paper, we have proposed an intuitionistic fuzzy Laplacian twin support vector machine for a semi-supervised classification problem, which is inspired by the intuitionistic fuzzy number and Lap-TSVM. Not only can it reduce the effect of noises and outliers through the membership and non-membership functions, but also use the geometric distribution information of labeled data and unlabeled data to construct a more accurate classifier. Experimental results indicated that our IFLap-TSVM performs well on both constructed test data and several real-world datasets. Compared with Lap-TSVM, IFTSVM and TSVM, IFLap-TSVM has the best performance. In the future, we will pay attention to improve intuitionistic fuzzy number to further reduce the effect of noises and outliers on the model. Moreover, there may be noises in unlabeled samples and that how to deal with this should also be considered. And another possible work is to extend IFLap-TSVM to multi-class classification.

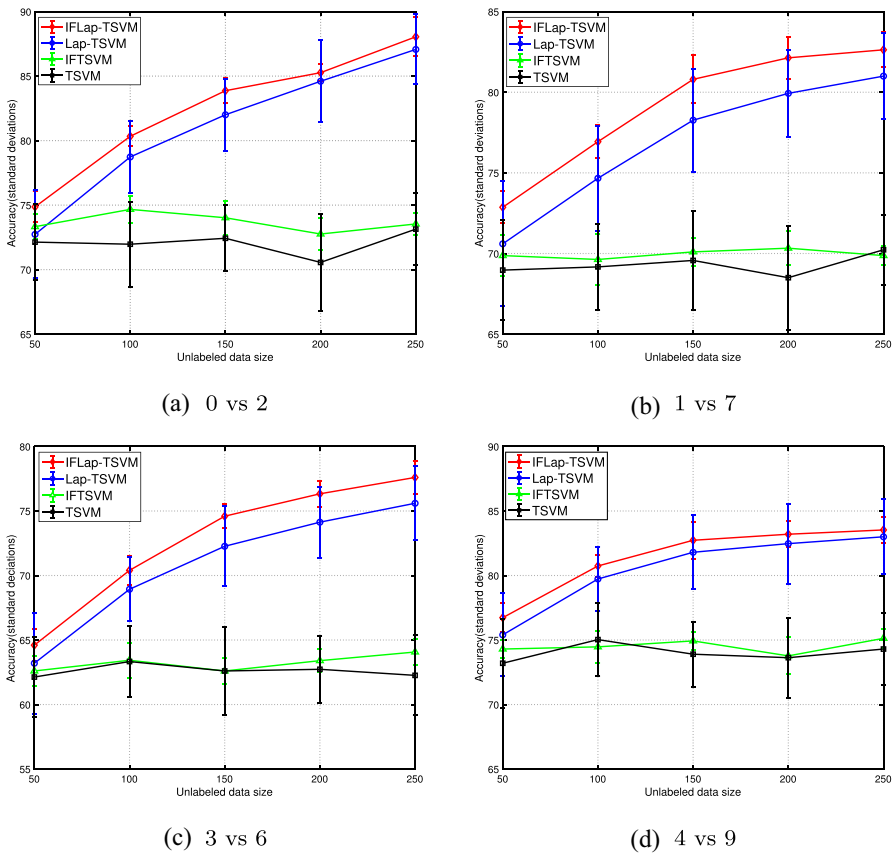


Fig. 9 Test accuracy and standard deviations of IFLap-TSVM, Lap-TSVM, IFTSVM and TSVM on MNIST dataset. (a) 0 versus 2; (b) 1 versus 7; (c) 3 versus 6; (d) 4 versus 9

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- [1] Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
- [2] Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: *Proceedings of Conference on Machine Learning* (1998) <https://doi.org/10.1007/BFb0026683>
- [3] Li, X., Chen, G.: Face recognition based on PCA and SVM. In: *IEEE Photonics and Optoelectronics*, pp. 1–4 (2012), <https://doi.org/10.1109/SOPO.2012.6270973>

- [4] Sun, J., Shang, Z., Li, H.: Imbalance-oriented SVM methods for financial distress prediction: a comparative study among the new SB-SVM-ensemble method and traditional methods. *J. Oper. Res. Soc.* **65**(12), 1905–1919 (2014)
- [5] Chen, L., Zhou, M., Wu, M., et al.: Three-layer weighted fuzzy support vector regression for emotional intention understanding in human-robot interaction. *IEEE Trans. Fuzzy Syst.* **26**(5), 2524–2538 (2018)
- [6] Zhang, M., Zhen, Y., Hui, G., Chen, G.: Accurate multisteps traffic flow prediction based on SVM. *Math. Probl. Eng.* 1–8 (2013)
- [7] Bai, Y., Han, X., Chen, T., Yu, H.: Quadratic kernel-free least squares support vector machine for target diseases classification. *J. Combin. Optim.* **30**(4), 850–870 (2015)
- [8] Fung, G., Mangasarian, O.: Proximal support vector machine classifiers. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 77–86. Assoc. Comput. Mach., New York (2001)
- [9] Bai, Y., Zhu, Z., Yan, W.: Sparse proximal support vector machine with a specialized interior-point method. *J. Oper. Res. Soc. China* **3**, 1–15 (2015)
- [10] Mangasarian, O.L., Wild, E.W.: Multisurface proximal support vector machine classification via generalized eigenvalues. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(1), 69–74 (2006)
- [11] Jayadeva, R.K., Khemchandani, R., Chandra, S.: Twin support vector machines for pattern classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(5), 905–910 (2007)
- [12] Shao, Y., Chen, W., Zhang, J., et al.: An efficient weighted Lagrangian twin support vector machine for imbalanced data classification. *Pattern Recognit.* **47**(9), 3158–3167 (2014)
- [13] Tian, Y., Ju, X.: Nonparallel support vector machine based on one optimization problem for pattern recognition. *J. Oper. Res. Soc. China* **3**, 499–519 (2015)
- [14] Gao, Q., Bai, Y., Zhan, Y.: Quadratic kernel-free least square twin support vector machine for binary classification problems. *J. Oper. Res. Soc. China* **7**, 539–559 (2019)
- [15] Zhu, X., Goldber, A.: *Introduction to semi-supervised learning*. Morgan & Claypool (2009)
- [16] Yan, X., Bai, Y., Fang, S., Luo, J.: A kernel-free quadratic surface support vector machine for semi-supervised learning. *J. Oper. Res. Soc.* **67**(7), 1001–1011 (2016)
- [17] Zhan, Y., Bai, Y., Zhang, W., Ying, S.: A P-ADMM for sparse quadratic kernel-free least squares semi-supervised support vector machine. *Neurocomputing*. **306**, 37–50 (2018)
- [18] Belkin, M., Niyogi, P., Sindhvani, V.: Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **7**, 2399–2434 (2006)
- [19] Chen, W., Shao, Y., Xu, D., Fu, Y.: Manifold proximal support vector machine for semi-supervised classification. *Appl. Intell.* **40**(4), 623–638 (2014)
- [20] Qi, Z., Tian, Y., Shi, Y.: Laplacian twin support vector machine for semi-supervised classification. *Neural Netw.* **35**, 46–53 (2012)
- [21] Rastogi, R., Pal, A.: Fuzzy semi-supervised weighted linear loss twin support vector clustering. *Knowl. Based Syst.* **165**, 132–148 (2019)
- [22] Dong, H., Yang, L., Wang, X.: Robust semi-supervised support vector machines with Laplace kernel-induced correntropy loss functions. *Appl. Intell.* **51**(21), 1–15 (2021)
- [23] Lin, C., Wang, S.: Fuzzy support vector machines. *IEEE Trans. Neural Netw.* **13**(2), 464–471 (2002)
- [24] Revani, S., Wang, X., Pourpanah, F.: Intuitionistic fuzzy twin support vector machines. *IEEE Trans. Fuzzy Syst.* **27**(11), 2140–2151 (2019)
- [25] Tikhonov, A.: Regularization of incorrectly posed problems. *Sov. Math. Doklady*. **4**, 1624–1627 (1963)
- [26] Gantmacher, F.R.: *Matrix Theory*. Chelsea, New York (1990)
- [27] Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
- [28] Atanssov, K.: Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **20**, 87–96 (1986)
- [29] Ha, M., Wang, C., Chen, J.: The support vector machine based on intuitionistic fuzzy number and kernel function. *Soft Comput.* **17**(4), 635–641 (2013)
- [30] Deng, N., Tian, Y., Zhang, C.: *Support Vector Machines: Theory, Algorithms, and Extensions*. CRC Press, Philadelphia (2013)
- [31] Asuncion, A., Newman, D.: UCI machine learning repository (2007) <https://archive.ics.uci.edu/ml/index.php>
- [32] Chen, W., Shao, Y., Deng, N., Feng, Z.: Laplacian least squares twin support vector machine for semi-supervised classification. *Neurocomputing* **145**, 465–476 (2014)