Weighted Intuitionistic Fuzzy Least Squares Twin SVM

M.Sc. Thesis

By

Avijit Bhattacharjee

DISCIPLINE OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

KHANDWA ROAD, SIMROL, INDORE-453552 INDIA

JUNE 2020

Weighted Intuitionistic Fuzzy Least Squares Twin SVM

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

Master of Science

by

Avijit Bhattacharjee

(Roll No. 1803141003)

Under the guidance of

Dr. M. Tanveer

DISCIPLINE OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2020

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Weighted Intuitionistic Fuzzy Least Squares Twin SVM** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DISCIPLINE OF MATHEMATICS**, **Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2019 to June 2020 under the supervision of **Dr. M. Tanveer**, Associate Professor and Ramanujan Fellow, Discipline of Mathematics, IIT Indore.

The matter presented in this thesis by me has not been submitted for the award of any other degree of this or any other institute.

Signature of the student with date

(**Avijit Bhattacharjee**)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of Thesis Supervisor with date

(**Dr. M. Tanveer**)

Avijit Bhattacharjee has successfully given his M.Sc. Oral Examination held on 2020.

Signature of Supervisor of M.Sc Thesis

Date:

Signature of Convener, DPGC

Date:

Signature of PSPC Member 1

Date:

Signature of PSPC Member 2

Date:

Acknowledgements

I wish to express my gratitude towards my thesis supervisor Dr. M. Tanveer for giving me an opportunity to do project under his supervision and being supportive from the first day, guiding, encouraging and also for sharing his pearls of wisdom with me in the whole journey without which it would have been impossible to do a good research in the given topic.

I am thankful to significant support and guidance of Research Scholars of our lab, Mudasir Ahmad Ganaie, Bharat Richhariya and Ashraf Haroon Rashid. Despite being busy and focused on their work, they always helped me in my project work.

I would like to thank PSPC members, Dr. Vijay Kumar Sohani and Dr. Mohd Farooq Azam for their valuable remarks, suggestions and questionnaires.

I would also like to thank Dr. Md. Aquil Khan, Head-Discipline of Mathematics, for facilitating our OPTIMAL research lab with the state-of-the-art computing facilities to complete the research with ease.

I would also like to thank my friends who have directly or indirectly helped me in completing my work.

Avijit Bhattacharjee

1803141003

Discipline of Mathematics, IIT Indore.

Abstract

Twin support vector machine is one of the widely used classifiers for real world applications. Many algorithms are proposed in the literature to deal with noisy data sets. One of the different approaches is the intuitionistic fuzzy twin support vector machine (IFTWSVM) technique. However, most of the real-world problems have data sets with some correlation between two points in the same class which is not addressed by IFTWSVM. Within this thesis, we propose a twin SVM based classifier for binary data sets termed as weighted intuitionistic fuzzy least squares twin support vector machines. The proposed model not only judges input data by considering the membership and nonmembership values but also uses the correlation between two data points in the same class for the betterment of classification performance. This weighting technique can give weights considering the relative significance of each data point. Retrospecting the time complexity for big data sets we have used the least square model, which can solve the classification problem without solving any Quadratic Programming Problem. We have checked the efficiency of the proposed model on a considerable amount of data sets. It can be seen that the proposed model can generalize brilliantly when compared to other baseline models in terms of accuracy and area under the curve (AUC).

Contents

List of Figures

List of Tables

Abbreviations

K.K.T.	**K**arush-**K**uhn **T**ucker
SVM	**S**upport **V**ector **M**achine
QPP	**Q**uadratic **P**rogramming **P**roblem
TWSVM	**T**win **S**upport **V**ector **M**achine
LSTWSVM	**L**east **S**quares **T**win **S**upport **V**ector **M**achine
IFTWSVM	**I**ntuitionistic **F**uzzy **T**win **S**upport **V**ector **M**achine
WIFLSTWSVM	**W**eighted **I**ntuitionistic **F**uzzy **L**east **S**quares **T**win **S**upport **V**ector **M**achine

Chapter 1

Introduction

One of the popular, interpretable machine learning algorithms, Support Vector Machine (SVM) was introduced by Vapnik [1]. This algorithm is effective in classification and regression problems [1], [2]. While current trending algorithms like the artificial neural network has no guarantee to achieve global minimum, SVM can give this guarantee to achieve global minimum. There are so many challenging areas where support vector machines can be useful like medical [3], face detection [4] etc. SVM was also found useful in speech recognition [5], cancer classification with micro array data [6], satellite radiation [7], text categorization [8].

SVM is the result of the deal between structural complexity and empirical risk. In SVM the hypothesis is to find a function f: $\mathbb{R}^n \rightarrow$ [-1, 1] so that positive and negative instances are separated by the hyperplane. Vapnik [9] showed that the hyperplane maximizing the margin of training dataset will have minimal VC dimension in the set of all consistent hyperplanes, which will thus be optimal.

If any dataset containing support vectors is mixed with noise then SVM finds it difficult to find the optimal hyperplane. To address this issue FSVMs [10]- [11] came with a degree membership function for each training data point. The effects of outliers and noises were reduced by FSVM partially as while calculating the membership value of each training datapoint, distance from the class center with this data point was used only. A new idea with the combination of dual membership and SVM came in [12]. Though this helped in the betterment of FSVM's performance but also had some complications. Points far away from the center of that class produced superior membership function values than points proximal to the class center. Balasundaram and Tanveer [13] have introduced proximal bilateral-weighted fuzzy support vector machine classifiers in the SVM family for a binary classification problem. They have considered an input data point can belong to both the classes having different fuzzy memberships.

Two parallel support hyperplanes are build by classical support vector machines and then it maximizes the margin, minimizes the structural risk. A new approach was introduced in the SVM family with the start of the use of unparalleled hyperplanes, TWSVM [14] and GEPSVM [15]. In TWSVM, GEPSVM the main focus is to generate two unparalleled hyperplanes in such a way that the distance of each plane is minimized from one class and maximized from the other. Classical SVM solves one big quadratic programming problem (QPP) but when TWSVM came it solved the problem by solving two smaller sized QPPs. TWSVM [14] has shown astonishing results while remaining four times faster. In [16], Tanveer has reformulated TWSVM and made it a strongly convex problem named robust and sparse linear programming twin support vector machines. Richhariya and Tanveer [17] has integrated universum learning with SVM. They have used a rectangular matrix of small size to do the tasks in low memory and with faster speed.

Recently, concepts of intuitionistic fuzzy number and SVM were combined to form Intuitionistic fuzzy support vector machines [18, 19]. In fuzzy support vector machines for the classification problem, every point had a degree of membership associated with it to cut down error rates caused by noise and outliers. The problem with this degree membership function is that the distance between the class center and the sample point was measured in the sample space. To address the above issue in [18, 19], each point is mapped to a higher dimensional feature space. Thenceforth an intuitionistic fuzzy number is assigned to it and then with the help of a precise score function, one can calculate the contribution of this point in the classification algorithm.
However, algorithms like TWSVM and IFTWSVM and their variants suffer from the

following deficiency: most of the datasets have a neighbourhood structure which they failed to preserve. In a time where most of us are getting recommendation's in NETFLIX, YOUTUBE and in every ecommerce websites depending on the neighbourhood structure of our choices, the above deficiency of the algorithms make them incompetent. Moreover, we will explain the proposed algorithm **" Weighted Intuitionistic Fuzzy Least Squares Twin SVM"** minutely from where we can wrap up prior deficiency.

Chapter 2

Related Works

In this chapter, we have discussed minutely the concepts and algorithms analogous to the proposed algorithm.

2.1 Twin Support Vector Machines (TWSVMs)

Two parallel support hyperplanes are build by classical support vector machines and then it maximizes the margin, minimizes the structural risk. A new approach was introduced in the SVM family with the start of the use of unparalleled hyperplanes, Twin Support Vector Machine [14]. In TWSVM [14] the main focus is to engender two unparalleled hyperplanes in such a way that the distance of each plane is minimized from one class and maximized from the other. Classical SVM solves one big quadratic programming problem (QPP) but when TWSVM came it solved the problem by solving two smaller sized QPPs. TWSVM has shown astonishing results while remaining four times faster.

In this section, we will discuss about the TWSVM with linear and nonlinear kernels.

2.1.1 Linear TWSVM

To solve a binary classification problem with linear kernel, the hyperplanes

$$w_1^T x + b_1 = 0 \quad and \quad w_2^T x + b_2 = 0, \tag{2.1}$$

are obtained keeping in mind that each hyperplane has to be closest to the one class and away from the other class. A test data point gets it's class label by considering the minimum distance from the two hyperplanes.

The primal problem of linear TWSVM [14] is defined as:

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}||X_1 w_1 + e_1 b_1||^2 + c_1 e_2^T \xi \\ \text{s.t.} \quad & -(X_2 w_1 + e_2 b_1) + \xi \geq e_2, \ \xi \geq 0 \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} \min_{w_2,b_2,\eta} \quad & \frac{1}{2}||X_2 w_2 + e_2 b_2||^2 + c_2 e_1^T \eta \\ \text{s.t.} \quad & (X_1 w_2 + e_1 b_2) + \eta \geq e_1, \ \eta \geq 0, \end{aligned} \tag{2.3}$$

where $c_i > 0, i = 1, 2$ are penalty parameters, ξ and η are the slack variables, e_1 and e_2 are vectors of ones of appropriate dimensions, X_1 and X_2 are matrices corresponding to class one and class two with m_1 and m_2 datapoints respectively.

Lagrangian of (2.2) is given by:

$$L(w_1, b_1, \xi, \alpha, \beta) = \frac{1}{2}||X_1 w_1 + e_1 b_1||^2 + c_1 e_2^T \xi - \alpha^T(-(X_2 w_1 + e_2 b_1 + e_2) + \xi) - \beta^T \xi, \tag{2.4}$$

where Lagrange multipliers are given by $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{m2})^T \geq 0$ and
$\beta = (\beta_1, \beta_2, \ldots, \beta_{m1})^T \geq 0$.
The Karush-Kuhn Tucker (K.K.T.) [20] conditions are given as:

$$X_1^T(X_1 w_1 + e_1 b_1) + X_2^T \alpha = 0, \ e_1^T(X_1 w_1 + e_1 b_1) + e_2^T \alpha = 0 \text{ and } c_1 e_2 - \alpha - \beta = 0. \tag{2.5}$$

Using equation (2.5), we can obtain the solution as:

$$u_1 = -(R^T R)^{-1} P^T \alpha, \quad u_1 = [w_1, b_1]^T \tag{2.6}$$

where $R = [X_1 \quad e_1]$ and $P = [X_2 \quad e_2]$ are matrices augmented with ones. While computing inverse matrices in TWSVM we may face that some matrices are singular. Therefore to avoid this situation a small positive number, say ϵI, is added to the matrix, where I represents an identity matrix of appropriate dimension. Hence, equation (2.6) can be written as:

$$u_1 = -(R^T R + \epsilon I)^{-1} P^T \alpha, \quad u_1 = [w_1, b_1]^T. \tag{2.7}$$

Using equations (2.5) and (2.6) in equation (2.4), the dual formulation of QPP (2.2) is as follows:

$$\begin{aligned} &\max_{\alpha} \ -\frac{1}{2}\alpha^T P (R^T R)^{-1} P^T \alpha + e_2^T \alpha \\ &\text{s.t.} \quad 0 \leq \alpha \leq c_1. \end{aligned} \tag{2.8}$$

Similarly, the dual of QPP (2.3) is given by:

$$\begin{aligned} &\max_{\beta} \ -\frac{1}{2}\beta^T R (P^T P)^{-1} R^T \beta + e_1^T \beta \\ &\text{s.t.} \quad 0 \leq \beta \leq c_2. \end{aligned} \tag{2.9}$$

A test data point x gets it's class label from the decision function given by;

$$class = \arg\min_{i=1,2} |w_i^t x + b_i|. \tag{2.10}$$

2.1.2 Non-linear TWSVMs

TWSVM with non-linear kernel finds the hyperplanes given by:

$$K(x^T, A^T) w_1 + b_1 = 0 \quad and \quad K(x^T, A^T) w_2 + b_2 = 0, \tag{2.11}$$

where $A = [X_1; X_2]$, w_1 and $w_2 \in \mathbb{R}^n$ and K is denoting the Gaussian kernel. The primal problem of non-linear TWSVM is defined as :

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}||K(X_1, A^T)w_1 + e_1b_1||^2 + c_1e_2^T\xi \\ \text{s.t.} \quad & -(K(X_2, A^T)w_1 + e_2b_1) + \xi \geq e_2,\ \xi \geq 0 \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} \min_{w_2,b_2,\eta} \quad & \frac{1}{2}||K(X_2, A^T)w_2 + e_2b_2||^2 + c_2e_1^T\eta \\ \text{s.t.} \quad & (K(X_1, A^T)w_2 + e_1b_2) + \eta \geq e_1,\ \eta \geq 0. \end{aligned} \tag{2.13}$$

The corresponding duals are:

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2}\alpha^T P(R^T R)^{-1}P^T\alpha + e_2^T\alpha \\ \text{s.t.} \quad & 0 \leq \alpha \leq c_1 \end{aligned} \tag{2.14}$$

and

$$\begin{aligned} \max_{\beta} \quad & -\frac{1}{2}\beta^T R(P^T P)^{-1}R^T\beta + e_1^T\beta \\ \text{s.t.} \quad & 0 \leq \beta \leq c_2. \end{aligned} \tag{2.15}$$

Finally, one can obtain u_1, u_2 as:

$$u_1 = -(R^T R)^{-1}P^T\alpha, \quad u_1 = [w_1, b_1]^T \tag{2.16}$$

and

$$u_2 = (P^T P)^{-1}R^T\beta, \quad u_2 = [w_2, b_2]^T. \tag{2.17}$$

A new test data point $x \in \mathbb{R}^n$ gets it's class label from the decision function given by:

$$class = \arg\min_{i=1,2} |w_i^t K(x^T, A^T) + b_i|. \tag{2.18}$$

2.2 Least Squares Twin Support Vector Machines (LSTWSVMs)

Least squares TWSVM [21] was proposed to complement the performance of TWSVM [14]. LSTWSVM solves two systems of linear equations instead of two QPPs. This algorithm is awfully fast when compared with TWSVM. In the next parts of this section, we will discuss the linear as well as non-linear formulations of LSTWSVMs in a nutshell.

2.2.1 Linear LSTWSVMs

Objective function of linear LSTWSVM [21] is given as follows:

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}||X_1w_1 + e_1b_1||^2 + \frac{c_1}{2}||\xi||^2 \\ \text{s.t.} \quad & -(X_2w_1 + e_2b_1) + \xi = e_2 \end{aligned} \tag{2.19}$$

and

$$\begin{aligned} \min_{w_2,b_2,\eta} \quad & \frac{1}{2}||X_2w_2 + e_2b_2||^2 + \frac{c_2}{2}||\eta||^2 \\ \text{s.t.} \quad & (X_1w_2 + e_1b_2) + \eta = e_1. \end{aligned} \tag{2.20}$$

The above formulation is solved as simultaneous systems of linear equations. Now, substitute the equality constraint of (2.19) into the objective function (2.19) and obtain the following equation

$$\min_{w_1,b_1} \frac{1}{2}||X_1w_1 + e_1b_1||^2 + \frac{c_1}{2}||X_2w_1 + e_2b_1 + e_2||^2. \tag{2.21}$$

Using K.K.T. conditions, the solution of (2.19) is given as:

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -\Big(\frac{1}{c_1}R^TR + P^TP\Big)^{-1}P^Te_1. \tag{2.22}$$

Similarly, solution of (2.20) is given by:

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = \Big(\frac{1}{c_2}P^TP + R^TR\Big)^{-1} R^Te_2. \tag{2.23}$$

A new test data point $x \in \mathbb{R}^n$ gets it's class label from the decision function given by: (2.10).

2.2.2 Non-linear LSTWSVMs

TWSVM with non-linear kernel finds the hyperplanes given by (2.11). Primal formulation of non-linear LSTWSVM [21] is as follows:

$$\begin{aligned} \min_{w_1,b_1,\xi} \quad & \frac{1}{2}||K(X_1,A^T)w_1 + e_1b_1||^2 + \frac{c_1}{2}||\xi||^2 \\ \text{s.t.} \quad & -(K(X_2,A^T)w_1 + e_2b_1) + \xi = e_2 \end{aligned} \tag{2.24}$$

and

$$\begin{aligned} \min_{w_2,b_2,\eta} \quad & \frac{1}{2}||K(X_2,A^T)w_2 + e_2b_2||^2 + \frac{c_2}{2}||\eta||^2 \\ \text{s.t.} \quad & (K(X_1,A^T)w_2 + e_1b_2) + \eta = e_1. \end{aligned} \tag{2.25}$$

We can obtain the solutions of non-linear LSTWSVM similarly as we did in linear LSTWSVM as,

$$\begin{bmatrix} u_1 \\ b_1 \end{bmatrix} = -\Big(\frac{1}{c_1}S^TS + R^TR\Big)^{-1} R^Te_1 \tag{2.26}$$

and

$$\begin{bmatrix} u_2 \\ b_2 \end{bmatrix} = \Big(\frac{1}{c_2}R^TR + S^TS\Big)^{-1} S^Te_2. \tag{2.27}$$

A new test data point $x \in \mathbb{R}^n$ gets it's class label from the decision function given by (2.18).

2.3 Fuzzy Support Vector Machines

Though SVM is an omnipotent tool in classification jobs but this theory also has some boundary. While solving real-life problems by machine learning different point makes different effects in the training of the algorithm. From this problem the concept of fuzzy set, SVM was combined and Fuzzy SVM [10] came in the market.
Let A be a nonempty set. Fuzzy set F in the universe A is defined by
F=$\{(y, \mu_F(y)) : y \in A\}$
where $\mu_F : X \to [0, 1]$ and μ_F represents the membership degree of belongingness of y to A.
One can find the optimal hyperplane after solving

$$\min \frac{1}{2}||w||^2 + cS_2^t\eta$$
$$\text{subject to} \quad y_i(w.z_i + b) \geq 1 - \eta_i, \quad \eta \geq 0, \quad i = 1, 2, 3...m,$$

here c is a given constant and η_i is the error in classifying ith sample and z_i is the feature vector of ith sample. Lagrangian of the above QPP is given as:

$$L(w, b, \eta, \alpha, \beta) = \frac{1}{2}||w||^2 + cS_2^t\eta - \sum_1^m \alpha_i[y_i(w.z_i + b) - 1 + \eta_i] - \sum_1^m \beta_i\eta_i, \tag{2.28}$$

here α and β are the Lagrange multipliers.
To solve this problem we need the saddle point of the above Lagrangian to be found out. Following conditions should be satisfied by the parameters.

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^m \alpha_i y_i z_i = 0, \tag{2.29}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^m \alpha_i y_i = 0, \tag{2.30}$$

$$\frac{\partial L}{\partial \eta_i} = cS_i - \alpha_i - \beta_i = 0. \tag{2.31}$$

On using the above three equations, we transform the primal problem into its wolfe dual form

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{i=1}^{m} \alpha_i \alpha_j y_i y_j K(z_i, z_j)$$

$$\text{subject to} \sum_{i=1}^{m} y_i \alpha_i = 0, \qquad 0 \leq \alpha \leq cS.$$

A sample point x_i is a support vector if it has corresponding α_i is positive. Fuzzy SVM has two possible types of support vectors. Support vectors for which $cS_i > \alpha_i > 0$ gets the position at the margin of the hyperplane and if $\alpha_i = cS_i$ then corresponding sample point gets misclassified.

2.3.1 Importance of Fuzzy Membership in Classification Performance

In traditional Support Vector Machines, there is only one penalty parameter to vary when we want to balance the margin width and number of misclassified samples. Whenever we want to make the margin wide we choose small value for c and selection of large value for c will imply that objective function will focus more to make the model accurate.

Fuzzy SVM behaves similar to traditional SVM when score values for all the points is 1. If a point gets bigger score value that means that point more important while forming the decision surface for the classification problem.

Therefore, from the above discussion we can say that number of variable parameters (only one in traditional SVM) increases from SVM to Fuzzy SVM (in Fuzzy SVM it is equal to the count of training samples).

2.4 Intuitionistic Fuzzy Twin Support Vector Machines

In fuzzy support vector machines for classification problem every point had a degree of membership associated with it to cutdown error rates caused by noise and outliers. The problem with this degree membership function is that the distance between class center and sample point was measured in the sample space. To address the above issue each point is mapped to a higher dimensional feature space then an intuitionistic fuzzy number is assigned to it and then with the help of a precise score function one can calculate the contribution of this point in the classification algorithm. In this section, we will discuss about the intuitionistic fuzzy set, intuitionistic fuzzy number and linear and non-linear formulations of intuitionistic fuzzy support vector machines.

2.4.1 Intuitionistic Fuzzy set

Let A be a nonempty set. Fuzzy set F in the universe A is defined by
F=$\{(y, \mu_F(y)) : y \in A\}$ where $\mu_F : X \rightarrow [0, 1]$ and μ_F represents the membership degree of belongingness of y to A.
One can define an intuitionistic fuzzy set as
IF=$\{(y, \mu_{IF}(y), \eta_{IF}(y)) : y \in A\}$,
where $\mu_{IF} : X \rightarrow [0, 1]$, $\eta_{IF} : X \rightarrow [0, 1]$, $0 \leq \mu_{IF}(y) + \eta_{IF}(y) \leq 1$, μ_{IF} & η_{IF} represents the membership degree and nonmembership degree of belongingness of y to A.
The degree of hesitation of y that it will belong to A is defined by
$h_{IF}(y) = 1 - \mu_{IF}(y) - \eta_{IF}(y)$.
An intuitionistic fuzzy number is defined by $IFN = (\mu_{IF}, \eta_{IF})$.
$\therefore$ Largest intuitionistic fuzzy number and smallest intuitionistic fuzzy numbers are respectively (1,0),(0,1) from the conditions put on μ_{IF} & η_{IF}.
An intuitionistic fuzzy number alone is not enough we need to measure the number by score value assigned to it. Therefore, the basic score function is
s(IFN)= $\mu_{IFN} - \eta_{IFN}$.

However, it is not possible for the above score function to measure all the IFNs. Addressing the problem a precise score function was defined by

p(IFN)= $\mu_{IFN} + \eta_{IFN}$,

$\therefore p(IFN) + h_{IF}(IFN) = 1$.

Given two IFNs we can say that $IFN_1 < IFN_2$ if

$s(IFN_1) = s(IFN_2)$ and $p(IFN_1) < p(IFN_2)$.

Many score function can be defined depending on the precise score function. One such is defined by

$$G(IFN) = \frac{1 - \eta_{IFN}}{2 - \mu_{IFN} - \eta_{IFN}}. \tag{2.32}$$

Thus we have the conjunction amidst membership and nonmembership values as,

$s(IFN_1) < s(IFN_2) \Rightarrow G(IFN_1) < G(IFN_2)$,

$s(IFN_1) = s(IFN_2),\, p(IFN_1) < p(IFN_2) \Rightarrow G(IFN_1) < G(IFN_2)$.

2.4.1.1 Process of assigning degree membership to every training data points

Here degree membership is calculated in the feature space unlike FSVM where degree membership is calculated in the sample space. We first project the training points in the high dimensional space. After that we find the center for each class, calculate the distance for each training point from its class center to determine the membership values. Given a training data point one can assign the degree of membership as

$$\mu(x_i) = \begin{cases} 1 - \frac{\|\psi(x_i) - C_+\|}{R_+ + \epsilon}, & if \quad y_i = +1, \\ 1 - \frac{\|\psi(x_i) - C_-\|}{R_- + \epsilon}, & if \quad y_i = -1. \end{cases} \tag{2.33}$$

Here $\epsilon > 0$ needed to be adjusted according to requirements, C_+, C_-, R_+, R_- are centers and radius of positive and negative classes respectively in the higher dimensional space, vector norm is represented by $\|.\|$.

Class centers can be found as

$C_+ = \frac{1}{m_+} \sum_{y_i=1} \psi(x_i) \quad and \quad C_- = \frac{1}{m_-} \sum_{y_i=-1} \psi(x_i)$,

here m_+ and m_- are the counts of positive and negative samples.

Following equation explains how radius of each class can be calculated,
$R_+ = max_{y_i=+1}\|\psi(x_i) - C_+\|$ and $R_- = max_{y_i=-1}\|\psi(x_i) - C_-\|$.

2.4.1.2 Process of assigning degree non-membership to every training data points

Non-membership values for each training data points in proportional to the ratio of total number of incongruous points and the cardinality of the set of all training points in its neighbourhood. Non-membership function can be defined as follows,
$\gamma(x_i) = (1 - \mu(x_i))\tau(x_i)$.
Here $1 \geq \gamma(x_i) + \mu(x_i) \geq 0$, $\tau(x_i)$ is defined by, ratio of
$|\{x_i : \|\psi(x_j) - \psi(x_i)\| \leq \beta, y_i \neq y_j|$ *over* $|\{x_i : \|\psi(x_j) - \psi(x_i)\| \leq \beta\}|$.
Here β>0 is an adjustable parameter and the cardinality is denoted by $|.|$

2.4.1.3 Calculating Score Values

Training data points can be transformed into IFN using membership and nonmembership values as follows: IFN=$\{x_1, y_1, \mu_1, \gamma_1\}, \{x_2, y_2, \mu_2, \gamma_2\}, \ldots, \{x_m, y_m, \mu_m, \gamma_m\}$. Here μ_i and γ_i are respectively membership and nonmembership values. Therefore score values can be found as

$$S_i = \begin{cases} \mu_i, & \gamma_i = 0, \\ 0, & \mu_i \leq \gamma_i, \\ \frac{1-\gamma_i}{2-\mu_i-\gamma_i} & others. \end{cases} \tag{2.34}$$

2.4.1.4 Relation between kernel function and inner product distance

All the distances used in calculating the membership and non-membership values are in feature space. These distances depends on the inner products of feature vectors in high dimensional space. Therefore there is a relationship between kernel function and inner product distance given by
$\|\psi(x_p) - \psi(x_q)\| = \sqrt{K(x_p, x_p) + K(x_q, x_q) - 2K(x_p, x_q)}$.

Above relationship between kernel function and inner product can be proved as

$\|\psi(x_p) - \psi(x_q)\|$

$=\sqrt{(\psi(x_p) - \psi(x_q)).(\psi(x_p) - \psi(x_q))},$

$=\sqrt{(\psi(x_p).\psi(x_p)) + (\psi(x_q).\psi(x_q)) - 2(\psi(x_p).\psi(x_q))},$

$=\sqrt{K(x_p, x_p) + K(x_q, x_q) - 2K(x_p, x_q)}.$

2.4.2 Intuitionistic Fuzzy Twin Support Vector Machines (IFTWSVM)

The main idea of Intuitionistic fuzzy twin support vector machines [19] has been triggered by the Intuitionistic fuzzy number and Fuzzy twin SVM concept. Obtaining two nonparallel hyperplanes from the solutions of two smaller QPPs instead of one large QPP IFTWSVM classifies a point. The performance of IFTWSVM has been elevated as it considers both membership and nonmembership values.

2.4.2.1 Linear IFTWSVM

The primal problem of linear IFTWSVM [19] is defined as

$$\begin{aligned}
&\min_{w_1,b_1,\xi} \frac{1}{2}||X_1w_1 + e_1b_1||^2+\frac{c_1}{2}||w_1||^2 + c_2S_2^t\xi \\
&\text{subject to} \\
&-(X_2w_1 + e_2b_1) + \xi \geq e_2, \quad \xi \geq 0
\end{aligned} \tag{2.35}$$

and

$$\begin{aligned}
&\min_{w_2,b_2,\eta} \frac{1}{2}||X_2w_2 + e_2b_2||^2 + \frac{c_3}{2}||w_2||^2 + c_4S_1^t\eta \\
&\text{subject to} \\
&(X_1w_2 + e_1b_2) + \eta \geq e_1, \quad \eta \geq 0.
\end{aligned} \tag{2.36}$$

Here $c_1, c_2, c_3, c_4 > 0$ are parameters; e_1, e_2 are vector of ones of appropriate dimensions; X_1, X_2 are features of training data sets corresponding to each class; ξ and η are slack variables.

Lagrangian of (2.35) is given as:

$$\begin{aligned} L(w_1, b_1, \xi, \alpha, \beta) = \frac{1}{2}||X_1 w_1 + e_1 b_1||^2 + \frac{c_1}{2}||w_1||^2 + c_2 S_2^t \xi \\ +\alpha[(X_2 w_1 + e_2 b_1) - \xi + e_2] - \beta\xi, \end{aligned} \tag{2.37}$$

here α and β are the Lagrange multipliers.

One can obtain the K.K.T. conditions as,

$$\frac{\partial L}{\partial w_1} = X_1^t(X_1 w_1 + e_1 b_1) + c_1 w_1 + \alpha X_2 = 0, \tag{2.38}$$

$$\frac{\partial L}{\partial b_1} = e_1^t(X_1 w_1 + e_1 b_1) + \alpha e_2 = 0, \tag{2.39}$$

$$\frac{\partial L}{\partial \xi} = c_2 S_2^t - \alpha - \beta = 0, \tag{2.40}$$

from (2.38)-(2.40) we can say that

$$\begin{bmatrix} X_1^t \\ e_1^t \end{bmatrix} \begin{bmatrix} X_1 & e_1 \end{bmatrix} \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \alpha \begin{bmatrix} X_2 \\ e_2 \end{bmatrix} = 0. \tag{2.41}$$

Equation (2.41) can be written as,

$$u_1 = -(R^t R + c_1 I)^{-1} P^t \alpha, \tag{2.42}$$

where R=$\begin{bmatrix} X_1 & e_1 \end{bmatrix}$, P=$\begin{bmatrix} X_2 & e_2 \end{bmatrix}$, $u_1 = \begin{bmatrix} w_1 \\ b_1 \end{bmatrix}$.

Using the K.K.T. conditions and Lagrangian method we can write the corresponding Wolfe dual as

$$\begin{aligned} &\max_{\alpha} e_2^t \alpha - \frac{1}{2}\alpha^t P(R^t R + c_1 I)^{-1} P^t \alpha \\ &\text{subject to} \qquad 0 \le \alpha \le c_2 S_2 \end{aligned}$$

and

$$\max_{\beta} e_1^t \beta - \frac{1}{2}\beta^t R(P^tP + c_3 I)^{-1}R^t\beta$$
$$\text{subject to} \quad 0 \leq \beta \leq c_4 S_1.$$

One can obtain u_1 and u_2 as

$$u_1 = -(R^tR + c_1 I)^{-1}P^t\alpha, \tag{2.43}$$

$$u_2 = (P^tP + c_3 I)^{-1}R^t\beta, \tag{2.44}$$

A new test data point $x \in \mathbb{R}^n$ gets it's class label from the decision function given by:

$$\arg\min_{i=1,2} \frac{|w_i^t x + b_i|}{||w_i||}. \tag{2.45}$$

2.4.2.2 Non-linear IFTWSVM

The primal problem of nonlinear IFTWSVM is defined as

$$\begin{aligned} &\min_{w_1,b_1,\xi} \frac{1}{2}||K(X_1,A^t)w_1 + e_1b_1||^2 + \frac{c_1}{2}||w_1||^2 \\ &\qquad\qquad + c_2 S_2^t \xi \\ &\text{subject to} \\ &-(K(X_2,A^t)w_1 + e_2b_1) + \xi \geq e_2, \quad \xi \geq 0 \end{aligned} \tag{2.46}$$

and

$$\begin{aligned} &\min_{w_2,b_2,\eta} \frac{1}{2}||K(X_2,A^t)w_2 + e_2b_2||^2 + \frac{c_3}{2}||w_2||^2 \\ &\qquad\qquad + c_4 S_1^t \eta \\ &\text{subject to} \\ &(K(X_1,A^t)w_2 + e_1b_2) + \eta \geq e_1, \quad \eta \geq 0. \end{aligned} \tag{2.47}$$

Here $c_1, c_2, c_3, c_4 > 0$ are parameters; e_1, e_2 are vectors of ones of appropriate dimensions; X_1, X_2 are features of training datasets corresponding to each class; $A^t = [X_1X_2]^t$; ξ and η are slack variables.

Lagrangian of (2.46) is given as:

$$\begin{aligned} L(w_1, b_1, \xi, \alpha, \beta) = \frac{1}{2}||K(X_1, A^t)w_1 + e_1 b_1||^2 + \frac{c_1}{2}||w_1||^2 \\ +c_2 S_2^t \xi + \alpha[(K(X_2, A^t)w_1 + e_2 b_1) - \xi + e_2] - \beta\xi, \end{aligned} \tag{2.48}$$

here α and β are the Lagrange multipliers.

One can obtain the K.K.T. conditions as,

$$\begin{aligned} \frac{\partial L}{\partial w_1} = K(X_1, A^t)^t (K(X_1, A^t)w_1 + e_1 b_1) + c_1 w_1 \\ +\alpha K(X_2, A^t) = 0, \end{aligned} \tag{2.49}$$

$$\frac{\partial L}{\partial b_1} = e_1^t (K(X_1, A^t)w_1 + e_1 b_1) + \alpha e_2 = 0, \tag{2.50}$$

$$\frac{\partial L}{\partial \xi} = c_2 S_2^t - \alpha - \beta = 0, \tag{2.51}$$

from (2.49)-(2.51) we can say that

$$\begin{bmatrix} K(X_1, A^t)^t \\ e_1^t \end{bmatrix} \begin{bmatrix} K(X_1, A^t) & e_1 \end{bmatrix} \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \alpha \begin{bmatrix} K(X_2, A^t) \\ e_2 \end{bmatrix} = 0. \tag{2.52}$$

Equation (2.52) can be written as,

$$u_1 = -(R^t R + c_1 I)^{-1} P^t \alpha, \tag{2.53}$$

where R=$\begin{bmatrix} K(X_1, A^t) & e_1 \end{bmatrix}$, P=$\begin{bmatrix} K(X_2, A^t) & e_2 \end{bmatrix}$, u$_1$ = $\begin{bmatrix} w_1 \\ b_1 \end{bmatrix}$.

Using the K.K.T. conditions and Lagrangian method we can write the corresponding Wolfe dual as

$$\begin{aligned} &\max_{\alpha} e_2^t \alpha - \frac{1}{2}\alpha^t P(R^t R + c_1 I)^{-1} P^t \alpha \\ &\text{subject to} \quad 0 \leq \alpha \leq c_2 S_2 \end{aligned}$$

and

$$\begin{aligned} &\max_{\beta} e_1^t \beta - \frac{1}{2}\beta^t R(P^t P + c_3 I)^{-1} R^t \beta \\ &\text{subject to} \quad 0 \leq \beta \leq c_4 S_1. \end{aligned}$$

One can obtain u_1 and u_2 as

$$u_1 = -(R^t R + c_1 I)^{-1} P^t \alpha, \tag{2.54}$$

$$u_2 = (P^t P + c_3 I)^{-1} R^t \beta. \tag{2.55}$$

A new test data point $x \in \mathbb{R}^n$ gets it's class label from the decision function given by:

$$\arg\min_{i=1,2} \frac{|w_i^t K(x, A^t) + b_i|}{\sqrt{w_i^t K(X_i, A^t) w_i}}. \tag{2.56}$$

Chapter 3

Proposed Algorithm

Within this chapter, we propose a new method for binary classification problems designated by Weighted Intuitionistic Fuzzy Least Squares Twin SVM (WIFLSTWSVM).

3.1 Weighted Intuitionistic Fuzzy Least Squares Twin SVM (WIFLSTWSVM)

To address intra-class local structure in a dataset, we used knn based weighting technique introduced in [22]. Given any pair of points (x_i, x_j) in the same class the weight matrix can be defined as

$$W_{ij} = \begin{cases} \exp \frac{-||x_i - x_j||^2}{t}, & \text{if } x_i \text{ is k nearest neighbour of } x_j, \\ 0, & otherwise. \end{cases} \tag{3.1}$$

Here t is a hyper-parameter. After this, the intra-class weights can be found as: $\rho_i^{(c)} = \sum_{j=1}^{k} W_{ij}$, where c varies over the number of class and i varies over the number of samples in that class.

The proposed WIFLSTWSVM acquires the subsequent alluring advantages:

- Paramountcy of the correlation between two points in the same class has been considered for the betterment of classification performance.

- We have augmented the objective function of the primal problem by adding the regularizing term which helps our algorithm to perform better.
- The solution of our algorithm is unique.

In the coming parts of this chapter, we will discuss formulations of the proposed WIFLSTWSVM for the linear and non-linear cases.

3.1.1 Linear WIFLSTWSVM

In the primal formulation of WIFLSTWSVM, we take linear equality constraints instead of inequality constraints. The primal formulation is given by:

$$\begin{aligned} \min_{w_1, b_1, \xi} \quad & \frac{1}{2}||\rho_1(X_1 w_1 + e_1 b_1)||^2 + \frac{c_1}{2}||S_2 \xi||^2 + \frac{c_2}{2}\left(||w_1||^2 + b_1^2\right) \\ \text{s.t.} \quad & -(X_2 w_1 + e_2 b_1) + \xi = e_2 \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} \min_{w_2, b_2, \eta} \quad & \frac{1}{2}||\rho_2(X_2 w_2 + e_2 b_2)||^2 + \frac{c_3}{2}||S_2 \eta||^2 + \frac{c_4}{2}\left(||w_2||^2 + b_2^2\right) \\ \text{s.t.} \quad & (X_1 w_2 + e_1 b_2) + \eta = e_1. \end{aligned} \tag{3.3}$$

Here $c_1, c_2, c_3, c_4 > 0$ are parameters; e_1, e_2 are vectors of ones of appropriate dimensions; X_1, X_2 are features of training data sets corresponding to each class; ρ_1, ρ_2 are intraclass weight matrices for respective class. ξ and η are slack variables.
Using the equality constraints (3.2), the objective function (3.2) can be rewritten as

$$\begin{aligned} \min_{w_1, b_1} \frac{1}{2}||\rho_1(X_1 w_1 + e_1 b_1)||^2 &+ \frac{c_1}{2}||S_2(X_2 w_1 + e_2 b_1 + e_2)||^2 \\ &+ \frac{c_2}{2}\left(||w_1||^2 + b_1^2\right). \end{aligned} \tag{3.4}$$

Taking the gradient of (3.4) with respect to w_1, b_1 and equating to zero we get respectively,

$$(\rho X_1)^t \rho_1(X_1 w_1 + e_1 b_1) + c_1(S_2 X_2)^t(S_2(X_2 w_1 + e_2 b_1 + e_2)) + c_2 w_1 = 0, \tag{3.5}$$

$$(\rho_1 e_1)^t \rho_1(X_1 w_1 + e_1 b_1) + c_1(S_2 e_2)^t(S_2(X_2 w_1 + e_2 b_1 + e_2)) + c_2 b_1 = 0. \tag{3.6}$$

Writing the above two equations in matrix form we get

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(c_1 T^t T + R^t R + c_2 I)^{-1} T^t S_2 e_2, \tag{3.7}$$

where R=$\begin{bmatrix} \rho_1 X_1 & \rho_1 e_1 \end{bmatrix}$, T=$\begin{bmatrix} S_2 X_2 & S_2 e_2 \end{bmatrix}$,
I = identity matrix of appropriate dimension.
In a similar way, we can compute the parameters for the other hyperplane by the following formula

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (c_3 R^t R + T^t T + c_4 I)^{-1} R^t S_1 e_1, \tag{3.8}$$

where R=$\begin{bmatrix} S_1 X_1 & S_1 e_1 \end{bmatrix}$, T=$\begin{bmatrix} \rho_2 X_2 & \rho_2 e_2 \end{bmatrix}$.

3.1.2 Nonlinear WIFLSTWSVM

The objective functions of nonlinear least squares weighted IFTWSVM (LSWIFTWSVM) are written as

$$\min_{w_1,b_1,\xi} \frac{1}{2}||\rho_1(K(X_1, A^t)w_1 + e_1 b_1)||^2 + \frac{c_1}{2}||S_2\xi||^2 + \frac{c_2}{2}\left(||w_1||^2 + b_1^2\right)$$

subject to

$$-(K(X_2, A^t)w_1 + e_2 b_1) + \xi = e_2 \tag{3.9}$$

and

$$\min_{w_2,b_2,\eta} \frac{1}{2}||\rho_2(K(X_2,A^t)w_2 + e_2b_2)||^2 + \frac{c_3}{2}||S_2\eta||^2 + \frac{c_4}{2}\big(||w_2||^2 + b_2^2\big)$$

subject to

$$(K(X_1,A^t)w_2 + e_1b_2) + \eta = e_1. \tag{3.10}$$

Here $c_1, c_2, c_3, c_4 > 0$ are parameters; e_1, e_2 are vectors of ones of appropriate dimensions; X_1, X_2 are features of training data sets corresponding to each class; $A^t = [X_1X_2]^t$; ξ and η are slack variables.
Using the equality constraints (3.9), the objective function (3.9) can be rewritten as

$$\begin{aligned}\min_{w_1,b_1} \frac{1}{2}||\rho_1(K(X_1,A^t)w_1 + e_1b_1)||^2 \\ + \frac{c_1}{2}||S_2(K(\mathrm{X}_2,\mathrm{A}^t)\mathrm{w}_1 + \mathrm{e}_2b_1 + e_2)||^2 + \tfrac{c_2}{2}\big(||w_1||^2 + b_1^2\big).\end{aligned} \tag{3.11}$$

Taking the gradient of (3.11) with respect to w_1, b_1 and equating to zero

$$\begin{aligned}(\rho K(X_1,A^t))^t\rho_1(K(X_1,A^t)w_1 + e_1b_1) + \\ c_1(S_2K(X_2,A^t))^t(S_2(K(X_2,A^t)w_1 + e_2b_1 + e_2)) + c_2w_1 = 0,\end{aligned} \tag{3.12}$$

$$(\rho_1e_1)^t\rho_1(\mathrm{K}(X_1,\mathrm{A}^t)\mathrm{w}_1 + e_1b_1) + c_1(S_2e_2)^t(\mathrm{S}_2(\mathrm{K}(X_2,\mathrm{A}^t)\mathrm{w}_1 + e_2b_1 + e_2) + c_2b_1 = 0. \tag{3.13}$$

Writing the above two equations in matrix form we get

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -(c_1T^tT + R^tR + c_2I)^{-1}T^tS_2e_2,$$

where $R = \begin{bmatrix}\rho_1K(X_1,A^t) & \rho_1e_1\end{bmatrix}$, $T = \begin{bmatrix}S_2K(X_2,A^t) & S_2e_2\end{bmatrix}$,
I = identity matrix of appropriate dimension, A is the feature matrix of all training samples.
In a similar way, we can compute the parameters for the other hyperplane by the

following formula

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = (c_3 R^t R + T^t T + c_4 I)^{-1} R^t S_1 e_1,$$

where R=$\begin{bmatrix} S_1 K(X_1, A^t) & S_1 e_1 \end{bmatrix}$, T=$\begin{bmatrix} \rho_2 K(X_2, A^t) & \rho_2 e_2 \end{bmatrix}$.

3.2 Theoretical Justification and Computational Complexity of the Proposed Algorithm

3.2.1 Theorem 1: Finding Hyperplanes of WIFLSTWSVM is a Convex Optimization Problem

Proof: Constraints of an optimization problem forms a convex set, ξ is linear and convex function is known to us. Therefore we need to check the convexity of the function

$$\frac{1}{2}||\rho(Xw_1 + eb_1)||^2 + \frac{1}{2}c_2(||w_1||^2 + b_1^2), \tag{3.14}$$

for ρ, X, e, c_2 and $t\in$ [0,1].
That is, we need to check,

$$\frac{t}{2}||\rho(Xw_1 + eb_1)||^2 + \frac{t}{2}c_2(||w_1||^2 + b_1^2) + \frac{1-t}{2}||\rho(Xw_2 + eb_2)||^2 + \frac{1-t}{2}c_2(||w_2||^2 + b_2^2)$$

$\geq$

$\frac{1}{2}||\rho X(t\mathrm{w}_1 + (1\text{-t})\ \mathrm{w}_2) + e(t\mathrm{b}_1 + (1\text{-t})\ \mathrm{b}_2)||^2 + \frac{c_2}{2}(||t\mathrm{w}_1 + (1\text{-t})\ \mathrm{w}_2||^2 + (t\mathrm{b}_1 + (1\text{-t})\ b_2)^2)$.
When we subtract the expression in the left hand side of the inequality from the expression on the right hand side we will get the expression,
$\rho^2 X^2 \mathrm{w}_1{}^2(\mathrm{t}^2 - t) + \rho^2 X^2 \mathrm{w}_2{}^2(\mathrm{t}^2 - t) + 2\rho^2 X^2 \mathrm{w}_1 w_2(\text{t-}t^2) + \mathrm{e}^2 \mathrm{b}_1{}^2(\mathrm{t}^2 - t) + e^2 \mathrm{b}_2{}^2(\mathrm{t}^2 - t) + 2e^2 \mathrm{b}_1 b_2(\text{t-}t^2) + 2\rho X \mathrm{w}_1 \mathrm{e} b_1(t^2 - t) + 2\rho X \mathrm{w}_1 \mathrm{e} b_2(\text{t-}t^2) + 2\rho X \mathrm{w}_2 \mathrm{e} \mathrm{b}_1(\text{t-}t^2) + 2\rho X \mathrm{w}_2 \mathrm{e} b_2(t^2 - t) + \mathrm{c}_2(w_1{}^2(\mathrm{t}^2 - t) + \mathrm{w}_2{}^2(\mathrm{t}^2 - t) + 2\mathrm{w}_1 w_2(\text{t-}t^2) + b_1{}^2(\mathrm{t}^2 - t) + \mathrm{b}_2{}^2(\mathrm{t}^2 - t) + 2\mathrm{b}_1 b_2(\text{t-}t^2))$
after simplifying the above expression we get
$(t^2 - t)[(\rho X(\mathrm{w}_1\text{-}w_2)^2 + (e(\mathrm{b}_1\text{-}b_2))^2 + 2\rho X e(\mathrm{w}_1\text{-}w_2)(b_1\text{-}b_2) + c_2[(w_1\text{-}w_2)^2 + (\mathrm{b}_1\text{-}b_2))^2]]$

$= (\mathrm{t}^2 - t)[[\rho X(\mathrm{w}_1\text{-}w_2)]+\mathrm{e}(b_1\text{-}b_2)]^2+\mathrm{c}_2[(w_1\text{-}w_2)^2 + (\mathrm{b}_1\text{-}b_2)^2]]\leq 0.$

$As(\mathrm{t}^2 - t) \leq 0$ for t $\in$ [0,1] (3.15)

Thus we have proved the required inequality by proving that the expression on the right hand side of inequality is smaller than the expression on the left side of the inequality. On the similar lines one can check the convexity of the other hyperplane. Thus we can conclude that WIFLSTWSVM model is convex.

3.2.2 Theorem 2: Uniqueness of the solutions for any ρ, X, e, S, c_1>0, c_2>0

Suppose Z is the objective function

$$Z = \frac{1}{2}||\rho(Xw + eb)||^2 + \frac{c_1}{2}||S\xi||^2 + \frac{c_2}{2}\big(||w||^2 + b^2\big), \tag{3.16}$$

let if possible m_1 and m_2 be the two solutions of the objective function. Since the objective function is convex. Therefore, a family of solutions is possible $m_t = (1-t)m_1+tm_2$, t$\in$ [0,1]. Also $Z(m_t) = Z(m_1) = Z(m_2)$. Therefore Z(m_t)-Z(m_1)=0.

From the above equation, we get

$\frac{1}{2}||\rho(X[(1-t)\mathrm{w}_1+\mathrm{t}w_2]\ +\mathrm{e}[(1\text{-}\mathrm{t})b_1\ +\mathrm{t}b_2])||^2 + \frac{c_1}{2}||S[(1-t)\xi_1 + t\xi_2]||^2 + \frac{c_2}{2}(||(1-t)w_1 + tw_2||^2 + [(1-t)b_1 + tb_2]^2)$ - $\frac{1}{2}||\rho(Xw_1 + eb_1)||^2 - \frac{c_1}{2}||S\xi_1||^2 - \frac{c_2}{2}\big(||w_1||^2 + b_1^2\big) = 0$

On differentiating the above equation w.r.t. t we get

$(\rho X)^2[(t-1)||w_1||^2 + t||w_2||^2 + (1-2t)<w_1, w_2>] + e^2[(t-1)||b_1||^2 + t||b_2||^2 + (1-2t)<b_1, b_2>]+\rho Xe[2(t-1)<w_1, b_1>+(1-2t)<w_1, b_2>+(1-2t)<w_2, b_1>+2t<w_2, b_2>]+ c_1S^2[(t-1)||\xi_1||^2 + t||\xi_2||^2 + (1-2t)<\xi_1, \xi_2>] + c_2[(t-1)||w_1||^2 + t||w_2||^2 + (1-2t)<w_1, w_2> + (t-1)||b_1||^2 + t||b_2||^2 + (1-2t)<b_1, b_2>] = 0$

On differentiating the above equation w.r.t. t we get

$(\rho X)^2[||w_1||^2 + ||w_2||^2 - 2<w_1, w_2>] + e^2[||b_1||^2 + ||b_2||^2 - 2<b_1, b_2>] + \rho Xe[2<w_1, b_1> - 2<w_1, b_2> - 2<w_2, b_1> + 2<w_2, b_2>] + c_1S^2[||\xi_1||^2 + ||\xi_2||^2 - 2<\xi_1, \xi_2>] + c_2[||w_1||^2 + ||w_2||^2 - 2<w_1, w_2> + ||b_1||^2 + ||b_2||^2 - 2<b_1, b_2>] = 0$

$\Rightarrow (\rho X)^2||w_1 - w_2||^2 + e^2||b_1 - b_2||^2 + 2\rho Xe<w_1 - w_2, b_1 - b_2> + c_1S^2||\xi_1 - \xi_2||^2 + c_2[||w_1 - w_2||^2 + ||b_1 - b_2||^2] = 0$

$\Rightarrow ||(\rho X)(w_1 - w_2) + e(b_1 - b_2)||^2 + c_1 S^2 ||\xi_1 - \xi_2||^2 + c_2[||w_1 - w_2||^2 + ||b_1 - b_2||^2] = 0$

$\Rightarrow w_1 = w_2; b_1 = b_2; \xi_1 = \xi_2$

Thus we can conclude that the solution is unique.

3.2.3 Complexity of the Proposed WIFLSTWSVM

We will use big-O notation to analyse the time complexity of the proposed model. Here n, m denotes the total number of training samples, number of samples in a single class respectively. The important contributions in computation cost is coming from:

1: To generate score values for each sample, we need O(m) computations [19].

2: Calculating the k-nearest neighbour we can compute the intra-class weights in $O(2m^2log(m))$ [22].

Thus, optimization of the proposed model requires the same time complexity as in RFLSTSVM by Richhariya and Tanveer [23] as we are using the same sized matrices while minimizing our objective function.

3.3 Experimental Results

In this section, we embrace the achievement of the proposed WIFLSTWSVM. To authenticate the preeminence of the proposed WIFLSTWSVM, we have compared the proposed algorithm with TWSVM [14], IFTWSVM [19]. Moreover, we have also given six number of tables containing experimental results of accuracy rates, AUC rates, ranks of three algorithms with linear as well as non linear kernels. We have done all these experiments using MATLAB R2017a on PC having configuration 2x intel Xeon Processor, 128 GB of RAM with 4 TB of secondary storage.

3.3.1 Parameter selection

While doing these experiments to bring out the best accuracy from the algorithms we have used grid search method to tune the hyper-parameters with five-fold cross validation method. We have set the parameters $c_1 = c_3$ and $c_2 = c_4$ for IFTWSVM and our proposed algorithm and $c_1 = c_2$ for TWSVM. For all three nonlinear models Gaussian ($\kappa(x_p, x_q) = exp(-||x_p - x_q||^2/\sigma^2)$)kernel has been used with kernel parameter $\sigma =[2^{-1}, 2, 2^1, 2^2, 2^3, 2^4, 2^5]$. In IFTWSVM and proposed WIFLSTWSVM's linear models Gaussian ($\kappa(x_p, x_q) = exp(-||x_p - x_q||^2/\sigma^2)$) kernel has been used with kernel parameter $\sigma =[2,2^1, 2^2, 2^3, 2^4, 2^5]$ to calculate the intuitionistic fuzzy score values in the feature space. Optimal values of c_1 and c_2 were searched in $[10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10,$ $10^2, 10^3, 10^4, 10^5]$ for all three models. For the proposed WIFLSTWSVM we have searched for the optimal value of the hyperparameter k in [1,2,3,4,5,6,7,8,9]. Before training features were normalised so that mean is zero and standard deviation is one as for some data set may have different range which in return may hamper the models performance.

3.3.2 Results and discussion

Here we will discuss the experimental results to authenticate the preeminence of the proposed WIFLSTWSVM in contrast with TWSVM and IFTWSVM.

3.3.2.1 Comparison on different benchmark datasets

While comparing the proposed WIFLSTWSVM with TWSVM and IFTWSVM we have taken the datasets from [24–26]. We have encapsulated the experimental results in Table 3.1, 3.2, 3.4, 3.3 along with the optimal value of the hyper-parameter k for the proposed WIFLSTWSVM.

In Table 3.1 we have presented accuracy rates for 34 datasets using nonlinear kernel. We can see from the table that in 14 datasets our model performs better than other two algorithms and in 3 datasets it has tied with one of the other two algorithms having best performance for that dataset.

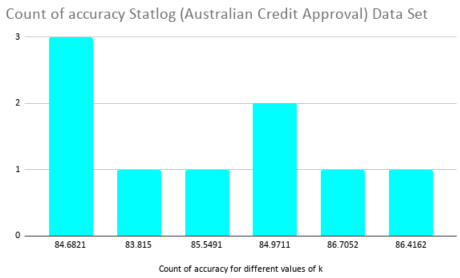
In Table 3.2, we presented accuracy rates for 32 datasets using linearkernel. We can see from the table that in 19 datasets our model out performs other two algorithms and in 1 dataset it is tied with one of the other two algorithms having best performance for that dataset.

Also in Table 3.4 and Table 3.3 we have shown that the proposed WIFLSTWSVM is not only capable of handling balanced dataset but it can also handle imbalanced dataset.

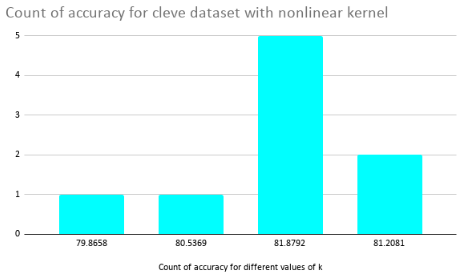
Moreover, we have given the count plot of accuracy for different number of neighbours used in intraclass weighting in Fig. 3.1 for four datasets using non-linear kernel.

In Fig. 3.1 (a), for Statlog (Australian Credit Approval) data set we got top accuracy for the proposed algorithm with nonlinear kernel as 86.7052 % for k value 8. From the count plot of accuracy for different values of k, it is very clear that if we can not choose k value properly the neighbourhood structure is not preserved properly and algorithm gets inferior results for this data set. Similarly, it can be recognized for other datasets in Fig. 3.1 (a), Fig. 3.1 (b) and Fig. 3.1 (c). Therefore the choice of the values of k is a significant job when we scrutinize the performance of our proposed algorithm.

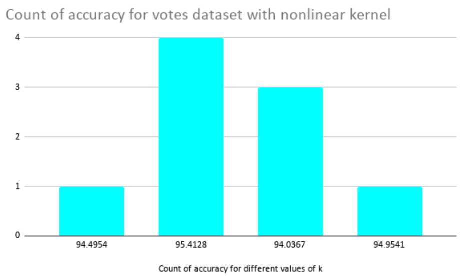
FIGURE 3.1: Effects of the hyper-parameter k on different datasets



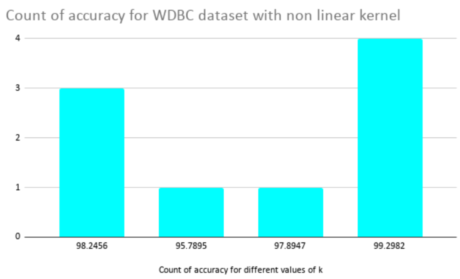
(a) Statlog (Australian Credit Approval) Dataset



(b) Heart Disease Dataset



(c) Congressional Voting Records Dataset



(d)Breast Cancer Wisconsin (Diagnostic) Dataset

Table 3.1: Accuracy rate with nonlinear kernel

Dataset	TWSVM	IFTWSVM	Proposed WIFLSTWSVM	k
iono	92.6136	**97.1591**	**97.1591**	4
wdbc	96.1404	96.4912	**99.2982**	6
wpbc	77.551	77.551	**78.5714**	1
heart-stat	77.9412	**85.2941**	84.5588	1
pima	78.4416	76.8831	**78.7013**	6
aus	86.9942	**88.7283**	86.7052	8
sonar	74.2857	**80**	74.2857	1
crossplane130	**100**	**100**	**100**	1
crossplane150	**100**	**100**	**100**	1
brwisconsin	96.7836	97.6608	**98.2456**	4
vehicle 1	**85.1415**	79.9528	82.0755	5
vehicle2	**99.0566**	**99.0566**	**99.0566**	5
cleve	78.5235	80.5369	**81.8792**	3
haberman	**74.6753**	59.7403	72.7273	3
votes	94.4954	94.0367	**95.4128**	2
transfusion	81.6	**84.8**	84	1
checkerboard_Data	86.9942	**88.7283**	86.7052	8
monk2	**87.3754**	82.7243	81.3953	5
monk3	**97.1223**	96.0432	93.8849	2
ripley	**91.3738**	91.2141	90.8946	4
acute-inflammation_R	**100**	**100**	91.8033	7
acute-nephritis_R	**100**	85.2459	**100**	1
breast-cancer_R	73.6111	67.3611	**79.8611**	3
breast-cancer-wisc_R	95.4286	97.4286	**97.7143**	2
breast-cancer-wisc-prog_R	75	74	**76**	1
credit-approval_R	87.8613	**89.0173**	87.2832	1
echocardiogram_R	84.8485	**87.8788**	**87.8788**	3
haberman-survival_R	74.026	**75.974**	72.0779	1
heart-hungarian_R	79.0541	**83.1081**	79.7297	5
hepatitis_R	82.0513	79.4872	**84.6154**	4
mammographic_R	81.4969	80.6653	**84.4075**	5
planning_R	69.5652	65.2174	**72.8261**	6
parkinsons_R	87.7551	89.7959	**91.8367**	1
molec-biol-promoter_R	75.9259	74.0741	**83.3333**	9

TABLE 3.2: Accuracy rate with linear kernel

Dataset	TWSVM	IFTWSVM	Proposed WIFLSTWSVM	k
iono	86.3636	**88.6364**	**88.6364**	2
wdbc	97.5439	**98.9474**	98.2456	1
wpbc	79.5918	78.5714	**81.6327**	1
heart-stat	**86.0294**	**86.0294**	**86.0294**	9
pima	76.6234	77.6623	**78.4416**	3
aus	86.9942	86.4162	**87.2832**	9
sonar	56.1905	72.381	**78.0952**	2
crossplane130	**100**	**100**	**100**	4
crossplane150	**100**	**100**	**100**	5
brwisconsin	96.1988	**98.538**	97.6608	1
vehicle 1	77.8302	77.1226	**78.7736**	4
vehicle2	94.5755	96.934	**97.8774**	7
cleve	**83.8926**	80.5369	83.2215	2
votes	**94.4954**	**94.4954**	**94.4954**	8
transfusion	83.7333	82.1333	**85.3333**	9
checkerboard_Data	86.9942	86.4162	**87.2832**	9
monk3	**83.8129**	82.0144	80.5755	1
ripley	88.9776	89.4569	**89.9361**	4
acute-inflammation_R	**100**	**100**	**100**	1
acute-nephritis_R	85.2459	**100**	**100**	1
breast-cancer_R	**72.9167**	70.1389	72.2222	3
breast-cancer-wisc_R	96.8571	96.8571	**97.1429**	1
breast-cancer-wisc-prog_R	75	66	**77**	2
credit-approval_R	**87.2832**	**87.2832**	85.8382	1
echocardiogram_R	84.8485	84.8485	**86.3636**	9
haberman-survival_R	76.6234	76.6234	**79.8701**	6
heart-hungarian_R	73.6486	83.1081	**86.4865**	4
hepatitis_R	**79.4872**	75.641	**79.4872**	1
mammographic_R	81.9127	80.8732	**83.9917**	1
planning_R	57.6087	67.3913	**70.6522**	1
parkinsons_R	82.6531	77.551	**85.7143**	1
molec-biol-promoter_R	64.8148	68.5185	**70.3704**	1

Table 3.3: AUC rate in test sets with nonlinear kernel

Dataset	TWSVM	IFTWSVM	Proposed WIFLSTWSVM	k
shuttle-6_vs_2-3	**100**	99.0991	79.0991	1
glass4	**91.1765**	**91.1765**	66.6667	1
ecoli-0-6-7_vs_3-5	85.7143	88.2653	**93.8776**	2
glass2	**64.898**	64.1837	55	1
ecoli-0-4-6_vs_5	85.8142	86.3636	**94.3556**	1
ecoli4	77.2727	**92.6064**	91.9735	2
ecoli-0-1_vs_5	82.1429	89.2857	**90.0534**	2
abalone9-18	76.5351	80.8114	**87.2442**	2
ecoli-0-3-4-6_vs_5	82.8014	88.8889	**94.4444**	1
ecoli-0-6-7_vs_5	**87.4183**	84.4771	83.4967	2
yeast2vs8	72.6496	73.7179	**75**	1
ecoli-0-1-4-7_vs_2-3-5-6	85.7143	**91.5668**	83.4562	1
ecoli2	80.4113	**89.1991**	86.2771	1
ecoli3	80.4113	**89.1991**	86.2771	1
yeast1vs7	64.3411	**69.5349**	62.5581	2
ecoli0137vs26	84.1518	92.6897	**97.2656**	1
yeast5	79.3503	96.6118	**97.9749**	1
ecoli-0-1_vs_2-3-5	75	74.0991	**78.7162**	1
ecoli-0-2-3-4_vs_5	91.8045	**96.8421**	**96.8421**	2
ecoli-0-2-6-7_vs_3-5	77.297	83.3333	**87.4466**	2
ecoli-0-1-4-7_vs_5-6	86.5323	**88.7634**	85.5376	2
yeast3	78.9388	**90.6366**	87.8057	2
yeast1	71.8094	**72.873**	70.8078	2

Table 3.4: AUC rate in test sets with linear kernel

Datasset	TWSVM	IFTWSVM	WIFLSTWSVM	k
shuttle-6_vs_2-3	60	98.1982	**100**	1
glass4	66.6667	65.6863	**80.3922**	2
ecoli-0-6-7_vs_3-5	77.551	91.8367	**92.8571**	1
glass2	64.3878	**74.3878**	70.3061	2
ecoli-0-4-6_vs_5	**85.8142**	85.015	85.5644	1
ecoli4	89.9597	**93.8723**	**93.8723**	3
ecoli-0-1_vs_5	87.8838	84.7463	**89.4192**	2
abalone9-18	72.3319	**87.6827**	87.2807	2
ecoli-0-3-4-6_vs_5	82.8014	88.357	**93.3806**	1
ecoli-0-6-7_vs_5	**88.3987**	86.4379	85.9477	1
yeast2vs8	**75**	**75**	**75**	1
ecoli-0-1-4-7_vs_2-3-5-6	**81.4977**	78.5945	73.0645	1
ecoli2	85.8658	87.4892	**88.5498**	2
ecoli3	85.8658	87.4892	**88.5498**	2
yeast1vs7	60.0775	59.9225	**66.8217**	2
ecoli0137vs26	91.0714	94.0848	**94.308**	1
yeast5	94.76	**96.9972**	96.648	1
ecoli-0-1_vs_2-3-5	**81.0811**	78.2658	80.1802	1
ecoli-0-2-3-4_vs_5	**96.8421**	**96.8421**	85.2632	1
ecoli-0-2-6-7_vs_3-5	82.8526	87.4466	**91.0791**	1
ecoli-0-1-4-7_vs_5-6	**90.6989**	89.4086	89.7312	1
yeast3	85.4828	91.0534	**91.5861**	2
yeast1	68.597	**71.1101**	70.7969	2

Table 3.5: Rank table for TWSVM, IFTWSVM and WIFLSTWSVM with linear kernel on the basis of accuracy

Dataset	TWSVM	IFTWSVM	WIFLSTWSVM
iono	3	1.5	1.5
wdbc	3	1	2
wpbc	2	3	1
heart-stat	2	2	2
pima	3	2	1
aus	2	3	1
sonar	3	2	1
crossplane130	2	2	2
crossplane150	2	2	2
brwisconsin	3	1	2
vehicle 1	2	3	1
vehicle2	3	2	1
cleve	1	3	2
votes	2	2	2
transfusion	2	3	1
checkerboard_Data	2	3	1
monk3	1	2	3
ripley	3	2	1
acute-inflammation_R	1	1	1
acute-nephritis_R	2	2	2
breast-cancer_R	1	3	2
breast-cancer-wisc_R	1.5	1.5	1
breast-cancer-wisc-prog_R	2	3	1
credit-approval_R	1.5	1.5	2
echocardiogram_R	1.5	1.5	1
haberman-survival_R	1.5	1.5	1
heart-hungarian_R	3	2	1
hepatitis_R	1.5	2	1.5
mammographic_R	2	3	1
planning_R	3	2	1
parkinsons_R	2	3	1
molec-biol-promoter_R	3	2	1
average rank	2.1093	2.1406	1.4062

TABLE 3.6: Rank table for TWSVM, IFTWSVM and WIFLSTWSVM with nonlinear kernel on the basis of accuracy

Dataset	TWSVM	IFTWSVM	WIFLSTWSVM
iono	3	1.5	1.5
wdbc	3	2	1
wpbc	2.5	2.5	1
heart-stat	3	1	2
pima	2	3	1
aus	2	1	3
sonar	2.5	1	2.5
crossplane130	2	2	2
crossplane150	2	2	2
brwisconsin	3	2	1
vehicle 1	1	3	2
vehicle2	2	2	2
cleve	3	2	1
haberman	1	3	2
votes	2	3	1
transfusion	3	1	2
checkerboard_Data	2	1	3
monk2	1	2	3
monk3	1	2	3
ripley	1	2	3
acute-inflammation_R	1.5	1.5	2
acute-nephritis_R	1.5	2	1.5
breast-cancer_R	2	3	1
breast-cancer-wisc_R	3	2	1
breast-cancer-wisc-prog_R	2	3	1
credit-approval_R	2	1	3
echocardiogram_R	2	1.5	1.5
haberman-survival_R	2	1	3
heart-hungarian_R	3	1	2
hepatitis_R	2	3	1
mammographic_R	2	3	1
planning_R	2	3	1
parkinsons_R	3	2	1
molec-biol-promoter_R	2	3	1
average rank	2.1176	2.0294	1.7647

Chapter 4

Conclusions and Future Directions

In this work, we have proposed a novel WIFLSTWSVM for classification problem. The proposed WIFLSTWSVM finds a hyperplane by solving a convex optimization problem having a unique solution. Moreover, we have taken motivation from IFN and IFTWSVM, Locality preserving projections(LPP) [27]. Using this LPP to generate intra class weights we have given importance to the community structure of a dataset. Benefits of using the proposed algorithm in binary classification problems can be easily seen from the comparison of experimental results. Thus the proposed algorithm is capable of producing surpassing results. Moreover, the proposed algorithm is cognizant of the values of k. If one does not choose it correctly then this algorithm may produce subsidiary results. As we have seen that the proposed algorithm is not only able to handle balanced data sets but it is also able to handle imbalanced datasets. In future one can use the proposed algorithm with specified weights for different classes in class imbalance problems for binary classification.

Bibliography

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[2] P. Bradley and O. Mangasarian, "Massive data discrimination via linear support vector machines," *Optimization methods and software*, vol. 13, no. 1, pp. 1–10, 2000.

[3] H. Zhu, X. Liu, R. Lu, and H. Li, "Efficient and privacy-preserving online medical prediagnosis framework using nonlinear svm," *IEEE journal of biomedical and health informatics*, vol. 21, no. 3, pp. 838–850, 2016.

[4] E. Osuna, R. Freund, and F. Girosit, "Training support vector machines: an application to face detection," in *Proceedings of IEEE computer society conference on computer vision and pattern recognition.* IEEE, 1997, pp. 130–136.

[5] A. Ganapathiraju, J. E. Hamaker, and J. Picone, "Applications of support vector machines to speech recognition," *IEEE transactions on signal processing*, vol. 52, no. 8, pp. 2348–2355, 2004.

[6] F. Chu and L. Wang, "Applications of support vector machines to cancer classification with microarray data," *International journal of neural systems*, vol. 15, no. 06, pp. 475–484, 2005.

[7] Y. Lee, Y. Lin, and G. Wahba, "Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 67–81, 2004.

[8] E. Leopold and J. Kindermann, "Text categorization with support vector machines. how to represent texts in input space?" *Machine Learning*, vol. 46, no. 1-3, pp. 423–444, 2002.

[9] V. Vapnik, E. Levin, and Y. L. Cun, "Measuring the vc-dimension of a learning machine," *Neural computation*, vol. 6, no. 5, pp. 851–876, 1994.

[10] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE transactions on neural networks*, vol. 13, no. 2, pp. 464–471, 2002.

[11] R. K. Sevakula and N. K. Verma, "Compounding general purpose membership functions for fuzzy support vector machine under noisy environment," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1446–1459, 2017.

[12] M.-M. Zhou, L. Li, and Y.-L. Lu, "Fuzzy support vector machine based on density with dual membership," in *2009 International Conference on Machine Learning and Cybernetics*, vol. 2. IEEE, 2009, pp. 674–678.

[13] S. Balasundaram and M. Tanveer, "On proximal bilateral-weighted fuzzy support vector machine classifiers," *International Journal of Advanced Intelligence Paradigms*, vol. 4, no. 3-4, pp. 199–210, 2012.

[14] Jayadeva, R. Khemchandani, and S. Chandra, "Twin support vector machines for pattern classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 905–910, 2007.

[15] O. L. Mangasarian and E. W. Wild, "Multisurface proximal support vector machine classification via generalized eigenvalues," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 1, pp. 69–74, 2005.

[16] M. Tanveer, "Robust and sparse linear programming twin support vector machines," *Cognitive Computation*, vol. 7, no. 1, pp. 137–149, 2015.

[17] B. Richhariya and M. Tanveer, "A reduced universum twin support vector machine for class imbalance learning," *Pattern Recognition*, p. 107150, 2020.

[18] M. Ha, S. Huang, C. Wang, and X. Wang, "Intuitionistic fuzzy support vector machine," *Journal of Hebei University (Natural Science Edition)*, vol. 3, pp. 225–229, 2011.

[19] S. Rezvani, X. Wang, and F. Pourpanah, "Intuitionistic fuzzy twin support vector machines," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 11, pp. 2140–2151, 2019.

[20] H. W. Kuhn and A. W. Tucker, "Nonlinear programming, in (j. neyman, ed.) proceedings of the second berkeley symposium on mathematical statistics and probability," *University of California Press, Berkeley*, 1951.

[21] M. A. Kumar and M. Gopal, "Least squares twin support vector machines for pattern classification," *Expert systems with applications*, vol. 36, no. 4, pp. 7535–7543, 2009.

[22] X. Hua and S. Ding, "Weighted least squares projection twin support vector machines with local information," *Neurocomputing*, vol. 160, pp. 228–237, 2015.

[23] B. Richhariya and M. Tanveer, "A robust fuzzy least squares twin support vector machine for class imbalance learning," *Applied Soft Computing*, vol. 71, pp. 418–432, 2018.

[24] A. Asuncion and D. Newman, "Uci machine learning repository," 2007.

[25] K. Napierała, J. Stefanowski, and S. Wilk, "Learning from imbalanced data in presence of noisy and borderline examples," in *International Conference on Rough Sets and Current Trends in Computing*. Springer, 2010, pp. 158–167.

[26] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework." *Journal of Multiple-Valued Logic & Soft Computing*, vol. 17, 2011.

[27] X. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, 2004, pp. 153–160.