2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-6819-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/BIBM55620.2022.9994971

Fuzzy kernel-free support vector machine and applications in medicine

Ai-bing Ji [a], Bo-wen Wei [b], Ye Ji* [c]

[a] *College of Mathematics and Information Science, Hebei University, Baoding, 071002, China*

[b] *College of Management, Hebei University, Baoding, 071002, China*

[c] *Acuity knowledge services, Beijing, 10086, China*

Abstract. Classification is widely used in healthcare management. Support vector machines (SVMs), as an important classification algorithm, have been used in target disease classification and prediction. Considering that SVMs are highly dependent on kernel function or parameters of a kernel and medical data uncertainty. This paper extends the classical kernel-free quadratic surface SVM to the case where the dataset is fuzzy samples set. A concept of fuzzy quadratically separable is first introduced, and then kernel-free fuzzy quadratic surface support vector machine (FQSSVM) is proposed, and the algorithm of the FQSSVM is given. A two-stage FQSSVM algorithm is provided for the unbalanced fuzzy sample set. Finally, FQSSVM is used in the diagnosis of coronary and classification of Haberman's Survival Data. The actual applications verify the performance of our proposed model.

Keywords: Kernel-free quadratic surface, Support vector machine, Possibility measure, Fuzzy training examples, Diseases classification

1. Introduction

Classification is widely used in healthcare management. Disease diagnosis is essentially a classification problem [1]. SVM is a machine learning algorithm for training regression and classification rules from data. SVMs can be traced back to the classification model of Vapnik [2,3]. To classify the nonlinearly separable data sets, the kernel method is used in SVM. However, the performance of kernel-based SVM is highly dependent on the choice of kernel functions and kernel parameters [4,5], training kernel-based SVM often has high computational and time complexity. To take advantage of the idea of maximal margin while avoiding the troubles of using kernel trick, a kernel-free quadratic surface SVM (QSSVM) was introduced by Dagher [6]. Luo et al. extended QSSVM to soft margin QSSVM (SQSSVM) that considered outliers and noise [7]. Both models find a quadratic separating surface which maximizes an approximation of a relative geometric margin [6,7]. Bai et al. [8] proposed the quadratic kernel-free least squares SVM and used it in disease classification. For direct nonlinear semi-supervised classification, polyhedral separability [9] was first proposed for nonlinear classification based on the QSSVM. Kernel-free SVMs were also applied in multi-class classification [10] and quadratic surface support vector regression [11,12].

These kernel-free quadratic surface SVMs are easier to operate than kernel-based SVM models because the structure of the quadratic separating surface is explicit and clear [7].

In the QSSVM model, the training data is a real-valued input and the output is $y = \pm 1$. Considering the outliers and noise in the training data set, Ye Tian et al [13,14] introduced the fuzzy QSSVM, which utilized the membership function to represent the membership level of an example to a negative class or positive class. But essentially, it is still a common QSSVM of Dagher[6], the training examples are still real-valued data.

In practice, the training data is often uncertain or fuzzy because of the error of measurements and outliers. Therefore, the study of kernel-free quadratic surface SVM with fuzzy training data is very significant and necessary.

The rest of this paper is organized as follows. Section 2 gives some preliminary knowledge. The model and algorithm of soft kernel-free fuzzy quadratic surface SVM are introduced in Section 3. Simulations on actual data are given in Section 4. The conclusions are given in Section 5.

2. Preliminary

In this section, we give some related preliminaries.

2.1. Dagher's QSSVM Models

Definition 1 [5]. *For the training data set* $S = \{(x_i, y_i) \,|x_i \in R^p, y_i = \pm 1, i = 1, 2, ..., n\}$, *if there exists a symmetric matrix*

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1p} \\ w_{12} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ w_{1p} & w_{2p} & \cdots & w_{pp} \end{pmatrix} \in R^{p\times p}, b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} \in R^p, c \in R$$

such that $y_i(\frac{1}{2}x_i^T W x_i + b^T x_i + c) \geq 1, i = 1, 2, ..., n$, *then the training data set* S *is called quadratically separable.*

Let $g(x) = \frac{1}{2}x^T W x + bx + c$, $g(x) = \frac{1}{2}x^T W x + bx + c = 0$ *is called a quadratic classification surface. The vector* $\nabla g(x_i) = W x_i + b$ *is the gradient direction at* x_i *and the relative geometrical margin at* x_i *with respect to the hyper-surface* $g(x) = 0$ *can be approximately expressed as* $\gamma \approx \frac{1}{\|\nabla g(x_i)\|_2} = \frac{1}{\|W x_i + b\|_2}$.

To find the discriminant function $g(x) = \frac{1}{2}x^T W x + bx + c$, *Dagher [6] proposed the following QSSVM model:*

$$\min Q(W, b, c) = \sum_{i=1}^{n} \|W x_i + b\|_2^2$$

$$s.t. \qquad (1)$$

$$\begin{cases} y_i\left(\frac{1}{2}x_i^T W x_i + b^T x_i + c\right) \geq 1, \\ W \in R^{p\times p}, b \in R^p, c \in R. \end{cases}$$

Taking account of the fact that some data points may be misclassified (or outliers of the data set), a vector of slack variables $\xi = (\xi_1, \dots, \xi_n)^T$ *was introduced to describe the amount of violation of the constraints. Jian Luo et al. [7] proposed the following soft quadratic surface support vector machine (SQSSVM):*

$$\min Q(W, b, c) = \sum_{i=1}^{n} \|W x_i + b\|_2^2 + \eta \sum_{i=1}^{n} \xi_i$$

$$s.t. \qquad (2)$$

$$\begin{cases} y_i\left(\frac{1}{2}x_i^T W x_i + b^T x_i + c\right) \geq 1 - \xi_i, \\ \xi_i \geq 0, i = 1, 2, ..., n, \\ W \in R^{p\times p}, b \in R^p, c \in R. \end{cases}$$

where η *is specified beforehand, which adjusts the proportion between maximizing the geometrical margin and minimizing the training error term.*

2.2. Possibility measure and fuzzy chance constrained programming

Definition 2 . *Let* X *be a nonempty set,* $P(X)$ *be the class of all subsets of* X*, a mapping* $Pos : P(X) \rightarrow [0, 1]$ *is called a possibility measure if it satisfies:*

(1) $Pos(\phi) = 0$

(2) $Pos(X) = 1$

(3) $Pos(\bigcup_{t\in T} A_t) = \underset{t\in T}{Sup}\, Pos(A_t)$

Definition 3 . *Let* $\tilde{a}$ *be a fuzzy number, its membership function is*

$$\mu_{\tilde{a}}(x) = \begin{cases} 1 - \frac{m\text{-}x}{\alpha}, m - \alpha \leq x < m, \\ \quad 1, \ x = m, \\ 1 + \frac{m-x}{\beta}, m < x \leq m + \beta. \end{cases}$$

then $\tilde{a}$ *is called a triangular fuzzy number, denoted by* $\tilde{a} = (m, \alpha, \beta)$*, where* m *is the center of* $\tilde{a}$ *and* $\alpha > 0, \beta > 0$ *are the left and right spreads, respectively, if* $\alpha = \beta$*, then* $\tilde{a} = (m, \alpha, \beta)$ *is called a symmetric triangular fuzzy number, denoted by* $\tilde{a} = (m, \alpha)$.

Definition 4 . *Let* $\tilde{a}$ *be a fuzzy number, then the possibility measure of fuzzy event* $\tilde{a} \leq b$ *is defined by* $Pos(\tilde{a} \leq b) = Sup\{\mu_{\tilde{a}}(x) \,|\, x \in R, x \leq b\}$*. Similarly,* $Pos(\tilde{a} \geq b) = Sup\{\mu_{\tilde{a}}(x) \,|\, x \in R, x \geq b\}$, $Pos(\tilde{a} = b) = \mu_{\tilde{a}}(b)$.

If $\tilde{x}_i (i = 1, 2, \cdots, n)$ *are all fuzzy numbers, then* $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n)$ *is called a fuzzy number vector, the set of all fuzzy number vectors is denoted by* $F^n(R)$*, especially when* $\tilde{x}_i (i = 1, 2, \cdots, n)$ *are all triangular fuzzy numbers, then* $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \cdots \tilde{x}_n)$ *is called a triangular fuzzy number vector, all the triangular fuzzy number vectors are denoted by* $T^n(R)$.

Following from the Zadeh extension principle, then for function $f : R^n \to R$, $\tilde{y} = f(\tilde{x}_1, \tilde{x}_2, \cdots \tilde{x}_n)$ is a fuzzy number, and its membership function is:

$$\mu_{\tilde{y}}(v) = \sup_{u_1, u_2, \cdots, u_n \in R} \left\{ \min_{1 \le i \le n} \mu_{\tilde{x}_i}(u_i) \,|\, v = f(u_1, u_2, \cdots, u_n) \right\}$$

Especially when $\tilde{a}, \tilde{b}$ are two fuzzy numbers, we can similarly define $\tilde{c} = f\left(\tilde{a}, \tilde{b}\right)$ and obtain the following results.

Theorem 1[16]. *If $\tilde{a}_1 = (a_1, \alpha_1, \beta_1)$ and $\tilde{a}_2 = (a_2, \alpha_2, \beta_2)$ are two triangular fuzzy numbers, ρ is a real number, then*

- $\tilde{a}_1 + \tilde{a}_2 = (a_1 + a_2, \alpha_1 + \alpha_2, \beta_1 + \beta_2)$.
- $\rho \tilde{a}_1 = \begin{cases} (\rho a_1, \rho \alpha_1, \rho \beta_1), \rho \geqslant 0, \\ (\rho a_1, -\rho \beta_1, -\rho \alpha_1), \rho < 0. \end{cases}$
- *For* $\tilde{a}_1 > 0, \tilde{a}_2 > 0$,
  $\tilde{a}_1 \cdot \tilde{a}_2 \cong (m_1 m_2, m_1 \alpha_2 + m_2 \alpha_1, m_1 \beta_2 + m_2 \beta_1)$

Especially for symmetric triangular fuzzy number $\tilde{a} = (m, \alpha)$, $\lambda \cdot \tilde{a} = \lambda \cdot (m, \alpha) = (\lambda m, |\lambda| \alpha), \lambda \in R$

Theorem 2[14]. *Let $\tilde{a} = (a, \alpha, \beta)$ be a triangular fuzzy number, then*

$$Pos\{\tilde{a} \le 0\} = \begin{cases} 1, \ a \le 0, \\ 1 - \frac{a}{\alpha}, a - \alpha \le 0, a > 0, \\ 0, \ a - \alpha > 0. \end{cases}$$

3. Kernel-free fuzzy quadratic surface SVM for fuzzy training data (FQSSVM)

Consider the fuzzy training sample set $\tilde{S} = (\tilde{X}_1, y_1), (\tilde{X}_2, y_2), \cdots, (\tilde{X}_n, y_n)$, where $\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, ..., \tilde{x}_{ip})^T \in T^p(R)$, $\tilde{x}_{ik} = (x_{ik}, \alpha_{ik}, \beta_{ik})$, $y_i \in \{-1, 1\}$, $k = 1, 2, \cdots, p$, $i = 1, 2, \cdots, n$, $y_i = \pm 1$ is the class label.

The classification based on the fuzzy training set $\tilde{S} = (\tilde{X}_1, y_1), (\tilde{X}_2, y_2), \cdots, (\tilde{X}_n, y_n)$ is to find a decision function $g(\tilde{X})$ such that the positive class and the negative class can be separated with good generalization performance and low classification error.

In this study, a fuzzy quadratic surface is utilized to separate the fuzzy training data set $\tilde{S}$ into two classes. The proposed fuzzy quadratic surface SVM (FQSSVM) intends to find the parameters $W \in R^{p \times p}, b \in R^p, c \in R$ of a fuzzy quadratic surface $g(\tilde{X}) = \frac{1}{2}\tilde{X}^T W \tilde{X} + b\tilde{X} + c$ that separates the fuzzy training set $\tilde{S}$ into two classes with maximum relative geometric margin.

Definition 5. *For the fuzzy training sample set $\tilde{S} = \left\{\left(\tilde{X}_1, y_1\right), \left(\tilde{X}_2, y_2\right), \cdots, \left(\tilde{X}_n, y_n\right)\right\}$, if for a given level $\lambda\,(0 < \lambda \le 1)$, there exists a symmetric matrix $W = W^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{12} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{pp} \end{pmatrix} \in R^{p \times p}$, $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} \in R^p, c \in R$ such that, for $i = 1, 2, \cdots, n$*

$$Pos\left\{ y_i \left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c \right) \ge 1 \right\} \ge \lambda \quad (3)$$

then the fuzzy training sample set $\tilde{S}$ is called fuzzy quadratically separable with respect to the level λ. The fuzzy quadratic surface $\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c = \tilde{0}$ is a quadratic fuzzy classification surface.

Theorem 3. *If the fuzzy training example set $\tilde{S}$ is fuzzy quadratically separable with respect to the level $\lambda\,(0 < \lambda \le 1)$, where $\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{ip})^T$, $\tilde{x}_{jk} = (x_{jk}, \alpha_{jk}, \beta_{jk})$ is a triangular fuzzy number, then for the level λ, the inequality (3) is equivalent to:*

$$\begin{cases} y_i \left(\frac{1}{2}(t_{i1}, t_{i2}, \cdots, t_{ip}) W (t_{i1}, t_{i2}, \cdots, t_{ip})^T + b^T (t_{i1}, t_{i2}, \cdots, t_{ip})^T + c \right) \ge 1, \\ x_{ij} - \alpha_{ij}(1 - \lambda) \le t_{ij} \le x_{ij} + \beta_{ij}(1 - \lambda), i = 1, 2, \cdots, n; j = 1, 2, \cdots, p. \end{cases} \quad (4)$$

Proof. Since

$$Pos\left\{ y_i \left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c \right) \ge 1 \right\}$$

$$= Pos\{y_i(\frac{1}{2}(\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{ip}) W (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{ip})^T$$

$$+ b^T (\tilde{x}_{i1}, \tilde{x}_{i2}, \cdots, \tilde{x}_{ip})^T + c) \ge 1\}$$

$$= \sup_{t_{i1}, t_{i2}, \cdots, t_{ip} \in R} \left\{ \begin{array}{l} \min_{1 \le k \le p} \mu_{\tilde{x}_{ik}}(t_{ik}) | y_i(\frac{1}{2}(t_{i1}, t_{i2}, \\ \cdots, t_{ip}) W (t_{i1}, t_{i2}, \cdots, t_{ip})^T \\ + b^T (t_{i1}, t_{i2}, \cdots, t_{ip})^T + c) \ge 1 \end{array} \right\} \ge \lambda$$

□

Therefore, there exists $T_i = (t_{i1}, t_{i2}, \cdots, t_{ip})^T \in R^p$, such that for $1 \le k \le p$, $\mu_{\tilde{x}_{ik}}(t_{ik}) \ge \lambda$ and

$$y_i(\frac{1}{2}(t_{i1}, t_{i2}, \cdots, t_{ip}) W (t_{i1}, t_{i2}, \cdots, t_{ip})^T$$

$$+ b^T (t_{i1}, t_{i2}, \cdots, t_{ip})^T + c)$$

$$= y_i(\frac{1}{2}T_i^T W T_i + b^T T_i + c) \ge 1, i = 1, 2, \cdots, n.$$

By $\mu_{\tilde{x}_{ik}}(t_{ik}) \geq \lambda$, we have $x_{ik} - \alpha_{ik}(1-\lambda) \leq t_{ik} \leq x_{ik} + \beta_{ik}(1-\lambda)$, then

$$\begin{cases} y_i(\frac{1}{2}(t_{i1}, t_{i2}, \cdots, t_{ip})W(t_{i1}, t_{i2}, \cdots, t_{ip})^T \\ +b^T(t_{i1}, t_{i2}, \cdots, t_{ip})^T + c) \geq 1, \\ x_{ik} - \alpha_{ik}(1-\lambda) \leq t_{ij} \leq x_{ik} + \beta_{ik}(1-\lambda), \\ i = 1, 2, \cdots, n; k = 1, 2, \cdots, p. \end{cases}$$

According to the algorithm of Dagher's QSSVM and Theorem 3, the fuzzy quadratic surface SVM (FQSSVM) $g(\tilde{X}) = \frac{1}{2}\tilde{X}^T W \tilde{X} + b\tilde{X} + c$ can be derived by the following models.

FQSSVM is built based on the principle that maximizes the sum of the approximated relative geometrical margins at the center of the training fuzzy data with respect to $g(\tilde{X}) = \tilde{0}$, constrained by the conditions that the fuzzy training example set $\tilde{S}$ is fuzzy quadratically separable with respect to level λ. FQSSVM finds the fuzzy quadratic discriminant surface $g(\tilde{X}) = \frac{1}{2}\tilde{X}^T W \tilde{X} + b\tilde{X} + c$ by solving the following fuzzy chance constrained programming: for the given level $\lambda\,(0 < \lambda \leq 1)$,

$$\min Q(W, b, c) = \sum_{i=1}^{n} \|WX_i + b\|_2^2$$

s.t (5)

$$\begin{cases} Pos\left\{y_i\left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c\right) \geq 1\right\} \geq \lambda, \\ i = 1, 2, \cdots, n. \end{cases}$$

where $W = W^T \in R^{p\times p}, b \in R^p, c \in R$, $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$.

Considering the conditions that some fuzzy data points may be not fuzzy quadratically separable with respect to the level λ, a vector of slack variables $\xi = (\xi_1, \ldots, \xi_n)^T$ is introduced to measure the amount of violation of the constraints, then the soft fuzzy quadratic surface SVM (SFQSSVM) is presented as follows:

$$\min Q(W, b, c) = \sum_{i=1}^{n} \|WX_i + b\|_2^2 + \eta \sum_{i=1}^{n} \xi_i$$

s.t. (6)

$$\begin{cases} Pos\left\{y_i\left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c\right) \geq 1 - \xi_i\right\} \geq \lambda, \\ \xi_i \geq 0, (i = 1, 2, \cdots, n). \end{cases}$$

where $W = W^T \in R^{p\times p}$, $b \in R^p$, $c \in R$, $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$. $\eta > 0$ is specified adjustment parameter.

The fuzzy chance constrained programming (6) can be solved using the hybrid intelligent algorithm [17]. By Theorem 3, the fuzzy chance-constrained programming (6) is equivalent to the following classical convex quadratic programming problem (SFQSSVM):

$$\min Q(W, b, c) = \sum_{i=1}^{n} \|WX_i + b\|_2^2 + \eta \sum_{i=1}^{n} \xi_i$$

$s.t.$ (7)

$$\begin{cases} y_i\{(\frac{1}{2}(t_{i1}, t_{i2}, \cdots, t_{ip})W(t_{i1}, t_{i2}, \cdots, t_{ip})^T \\ +b^T(t_{i1}, t_{i2}, \cdots, t_{ip})^T + c)\} \geq 1 - \xi_i; \\ x_{ik} - \alpha_{ik}(1-\lambda) \leq t_{ik} \leq x_{ik} + \beta_{ik}(1-\lambda); \\ \xi_i \geq 0, ; (i = 1, 2, \cdots, n, k = 1, 2, \cdots, p) \\ W = W^T \in R^{p\times p}, b \in R^p, c \in R. \end{cases}$$

When all $\tilde{X}_i (i = 1, 2, ..., n)$ are real-valued data, then SFQSSVM degrades to classical QSSVM. To simplify the SFQSSVM model, the symmetric matrix W can be expressed in the following vector form $U = (a_{11}, a_{12}, \cdots, a_{1p}, a_{22}, ..., a_{2p}, \cdots, a_{pp})^T \in R^{\frac{p^2+p}{2}}$. The M_i is a $\frac{p^2+p}{2} \times p$ matrix formed as follows:

If the data $X_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$, then in each j-th row of M_i, check the elements of vector U one by one, if the k-th element of U is a_{jk} (where d is any number$1 \leq k \leq p$), then assign the k-th element of the j-th row of M_i to be x_{ik}, otherwise, assign it to be 0.

Let $H_i = (M_i, I) \in R^{p\times(\frac{p^2+p}{2}+p)}$, I is the identity matrix with p-dimensional,

$$Z = \begin{pmatrix} U \\ b \end{pmatrix} \in R^{\frac{p^2+3p}{2}}$$

and

$$S_i = \begin{pmatrix} \frac{1}{2}t_{i1}t_{i1}, ..., t_{i1}t_{ip}, \frac{1}{2}t_{i2}t_{i2} \cdots, t_{i2}t_{ip}, ..., \\ \frac{1}{2}t_{i(p-1)}t_{i(p-1)}, t_{i(p-1)}t_{ip}, \frac{1}{2}t_{ip}t_{ip}, t_{i1}, t_{i2}, ..., t_{ip} \end{pmatrix} \in R^{\frac{p^2+p}{2}+p}.$$

Then the objective of the SFQSSVM is

$$\sum_{i=1}^{n} \|WX_i + b\|_2^2 = \sum_{i=1}^{n} \|H_i Z\|_2^2$$

$$= \sum_{i=1}^{n} (H_i Z)^T (H_i Z)$$

$$= \sum_{i=1}^{n} Z^T H_i^T H_i Z$$

$$= Z^T (\sum_{i=1}^{n} H_i^T H_i) Z.$$

Let $G = \sum_{i=1}^{n} H_i^T H_i \in R^{\left(\frac{p^2+3p}{2}\right) \times \left(\frac{p^2+3p}{2}\right)}$.

Then SFQSSVM model (7) becomes the following quadratic optimization:

$$\min Q(W, b, c) = Z^T G Z + \eta \sum_{i=1}^{n} \xi_i$$

$$s.t. \tag{8}$$

$$\begin{cases} y_i \left(S_i^T Z + c\right) \geq 1 - \xi_i, \\ x_{ik} - \alpha_{ik}(1-\lambda) \leq t_{ik} \leq x_{ik} + \beta_{ik}(1-\lambda), \\ \xi_i \geq 0 \quad (i = 1, 2, \cdots, n; k = 1, 2, \cdots, p). \end{cases}$$

Obviously, optimization problem (8) is a quadratic convex optimization, there exists optimal solutions $Z = \begin{pmatrix} U \\ b \end{pmatrix}$ and c, therefore, we can obtain $W = W^T \in R^{p \times p}, b \in R^p, c \in R$.

For fuzzy example with unknown class $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n)$, the decision rule is: for the given confidence level $\lambda\,(0 < \lambda \leq 1)$, if $Pos\left\{\left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c\right) \geq 0\right\} \geq \lambda$, then $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n)$ is a positive example; if $Pos\left\{(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c) < 0\right\} \geq \lambda$, then $\tilde{X} = (\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n)$ is a negative example.

4. Simulations on actual data and disturbed databases from the UCI repository

To better demonstrate the performance of our proposed model, a performance measure is given as follows: TP denotes the true positive cases, that is, the number of samples whose target category and prediction category are all positive classes. Similarly, TN is the true negative case, FP represents the false positive case, and FN is the false negative case.

The accuracy (ACC) is the percentage of examples that are correctly classified, and its calculation formula is as follows:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}.$$

4.1. Simulations on actual data

In the following, SFQSSVM will be applied to the diagnosis of coronary. The datasets [16] in Table 1 are the diastolic pressure $(\tilde{x}_{i1})$ and plasma cholesterol $(\tilde{x}_{i2})$ of twenty-four persons. $y_i = 1$ is a healthy person, $y_i = -1$ is a coronary patient. $\tilde{x}_{i1}$ and $\tilde{x}_{i2}$ are symmetric triangular fuzzy numbers. Let $S_1 = \left\{\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})\,|i = 1, 2, ..., 10, 13, 14, ..., 22\right\}$ be training fuzzy data set, $S_2 = \left\{\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})\,|i = 11, 12, 23, 24\right\}$ be test fuzzy data set.

Based on the training data set S_1 provided in Table 1, when parameters $\eta = 0.1$, $\lambda = 0.85$, using Lingo 19 to solve the programming (7) or (8), we can obtain

$$W = \begin{pmatrix} 0.005522, \text{-}0.003817 \\ \text{-}0.003817\ \ 0.004394 \end{pmatrix},$$

$$b = \begin{pmatrix} \text{-}0.01299 \\ 0.05111 \end{pmatrix}, c = -0.2996.$$

Then the decision rule is as follows: for the given confidence level $\lambda = 0.85$, if

$$Pos\left\{\left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c\right) \geq 0\right\} \geq 0.85,$$

then $\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})$ is a positive example(healthy person); if

$$Pos\left\{\left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c\right) < 0\right\} \geq 0.85,$$

then $\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})$ is a negative example(coronary patient).

Using this decision rule to classify the data in Table 1. And meanwhile, our proposed model is compared with the fuzzy support vector machine [16]. The training accuracy and test accuracy are given in Table 2.

The results show that our proposed model has good training accuracy and test accuracy for this small fuzzy dataset, and compared with the existing fuzzy data classifier [16], it also has better performance and a high confidence level.

4.2. Simulations on UCI data set

The dataset S uses Haberman's Survival data from UCI, which contains cases from 1958 to 1970 at Billings Hospital of the University of Chicago on the survival of patients who had undergone surgery for breast cancer. There are 306 observations on four variables:

$y_i = 1$ if patient i survived 5 years or longer, $y_i = -1$ otherwise;

x_{i1} is the age of patient i at time of operation;

x_{i2} is the year of operation for patient i (minus 1900);

x_{i3} is the number of positive axillary nodes detected in patient i.

In this UCI data set, the independent variables (x_{i1}, x_{i2}, x_{i3}) are all described by positive integer. In fact, some of them are approximate values, such as x_{i1}, x_{i2}, and should be represented by a fuzzy number. According to the characteristics of the year data, x_{i1}, x_{i2} can be expressed as a special triangular fuzzy number of the following form:

$$\mu_{\tilde{x}}(t) = \begin{cases} 1, t = x, \\ 1 - \frac{1}{r}(t - x), x < t < x + r, 0 \leq r < 1. \\ 0, otherwise. \end{cases}$$

$\tilde{x}$ can be viewed as a fuzzy number generated by fuzzy right disturbance to the right of x, it is denoted by $\tilde{x} = (x, 0, r)$, its λ-cut $(\tilde{x})_\lambda = [x, x+r(1-\lambda)], 0 < \lambda \leq 1$.

Considering the uncertainty of x_{i1} and x_{i2}, we randomly give $x_{i1}(x_{i2})$ a fuzzy disturbance to the right, $r_{ik} \sim U(0,1), k = 1, 2$, then we have the fuzzy data set $F(S) = \{(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{i3})\,|i = 1, 2, ..., 306\}$, where $\tilde{x}_{ik}$=$(x_{ik}, 0, r_{ik}), i = 1, 2, \cdots, 306$, $i = 1, 2, \cdots, 306$; $p = 1, 2,$. In this simulation study, the first 204 sample sets $F(S_1)$ are selected as the fuzzy training data sets, the rest $F(S_2)$ as the fuzzy test data sets.

The SFQSSVM (7) for Haberman's Survival data with fuzzy right disturbance is as follows:

$$\min Q\,(W, b, c) = \sum_{i=1}^{204} \|WX_i + b\|_2^2 + \eta \sum_{i=1}^{204} \xi_i$$

s.t. (9)

$$\begin{cases} y_i \left(\frac{1}{2} \left(t_{i1}, t_{i2}, t_{i3}\right) W \left(t_{i1}, t_{i2}, t_{i3}\right)^T + b^T \left(t_{i1}, t_{i2}, t_{i3}\right)^T + c \right) \geq 1 - \xi_i, \\ x_{ij} \leq t_{ij} \leq x_{ij} + r_{ij}\,(1-\lambda)\,, i = 1, 2, \cdots, 204, k = 1, 2, \\ \xi_i \geq 0, i = 1, 2, \cdots, 204, \\ W = \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{12} & w_{22} & w_{23} \\ w_{13} & w_{23} & w_{33} \end{pmatrix} \in R^{p\times p}, b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in R^p, c \in R. \end{cases}$$

The optimization problem (9), as before, can be easily converted into a quadratic optimization form like that of (8).

Based on the training data given in $F(S_1)$, when parameters $\eta = 0.1$, $\lambda = 0.95$, using Lingo 19 to solve the programming (9), we can obtain

$$W = \begin{pmatrix} 0.5797E-04 & 0.1675E-03 & -0.1194E-02 \\ 0.1675E-03 & 0.3565E-03 & 0.1041E-02 \\ -0.1194E-02 & 0.1041E-02 & -0.1166E-02 \end{pmatrix},$$

$$b = \begin{pmatrix} -0.01280 \\ -0.03255 \\ -0.01759 \end{pmatrix}, c = 2.393934,$$

then the decision rule is obtained as follows:

For the given confidence level $\lambda = 0.95$, if $Pos\left\{\left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c\right) \geq 0\right\} \geq 0.95$, then $\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, x_{i3})$ is a positive example; if $Pos\left\{\left(\frac{1}{2}\tilde{X}_i^T W \tilde{X}_i + b^T \tilde{X}_i + c\right) < 0\right\} \geq 0.95$, then $\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, x_{i3})$ is a negative example. Using this decision rule to classify the Haberman's Survival data, the training accuracy and test accuracy are given in Table 3.

To show the advantages of our algorithm, we compare our proposed model with other classification models. There are several classification models to classify Haberman's Survival data, such as Decision Tree [18], Support Vector Machine [18], Probabilistic Modeling Approach [19], and Multi-Branch Ferns-based Naive Bayesian classifier [20]. Their training accuracy is listed in Table 4.

Table 4 summarizes the accuracy comparison with other classification model. It is evident that our proposed model has better training accuracy and test accuracy, although its training data is subjected to certain fuzzy disturbance. Table 3 shows that positive samples have high training accuracy and testing accuracy, but negative samples have very low training accuracy and testing accuracy. This may be due to the imbalance between the two classes of samples. Positive samples

Table 1

the data of diastolic pressure and plasma cholesterol of the patient of Coronary and healthy people

i	$\tilde{x}_{i1}$(KPa)	$\tilde{x}_{i2}$(mmol/L)	y_i	i	$\tilde{x}_{i1}$	$\tilde{x}_{i2}$	y_i
1	$(9.86, 0.02)$	$(5.18, 0.01)$	1	13	$(10.66, 0.04)$	$(2.07, 0.01)$	-1
2	$(13.33, 0.02)$	$(3.73, 0.01)$	1	14	$(12.53, 0.02)$	$(4.45, 0.01)$	-1
3	$(14.66, 0.03)$	$(3.89, 0.02)$	1	15	$(13.33, 0.03)$	$(3.06, 0.02)$	-1
4	$(9.33, 0.01)$	$(7.10, 0.02)$	1	16	$(9.33, 0.01)$	$(3.94, 0.04)$	-1
5	$(12.80, 0.03)$	$(5.49, 0.02)$	1	17	$(10.66, 0.02)$	$(4.45, 0.02)$	-1
6	$(10.66, 0.02)$	$(4.09, 0.03)$	1	18	$(10.66, 0.02)$	$(4.92, 0.03)$	-1
7	$(10.66, 0.01)$	$(4.45, 0.02)$	1	19	$(9.33, 0.02)$	$(3.68, 0.02)$	-1
8	$(13.33, 0.02)$	$(3.63, 0.03)$	1	20	$(10.66, 0.02)$	$(3.21, 0.01)$	-1
9	$(13.33, 0.01)$	$(5.70, 0.02)$	1	21	$(10.40, 0.03)$	$(3.94, 0.02)$	-1
10	$(12.00, 0.03)$	$(6.19, 0.02)$	1	22	$(9.33, 0.02)$	$(4.92, 0.02)$	-1
11	$(14.66, 0.02)$	$(4.01, 0.01)$	1	23	$(11.20, 0.01)$	$(3.42, 0.02)$	-1
12	$(13.33, 0.02)$	$(4.01, 0.02)$	1	24	$(9.33, 0.02)$	$(3.63, 0.01)$	-1

Table 2

Performance evaluation and the accuracy comparison

Model	Parameters	Training accuracy ACC	Test accuracy ACC
SFQSSVM	$C = 0.1, \lambda = 0.85$	85%	100%
SVM based on fuzzy data[16]	Linear kernel,$C = 0.1, \lambda = 0.65$	80%	100%

Table 3

Performance evaluation of SFQSSVM

Training accuracy ACC		Test accuracy ACC	
Positive examples	99.3%	Positive examples	98.7%
Negative examples	8.93%	Negative examples	4%
Total	74.5%	Total	75.4%

Table 4

The accuracy comparison with other classification models

Model	Training accuracy ACC	Test accuracy ACC
SFQSSVM	74.5%	75.4%
Decision tree[18]	67.41%	Not provided
SVM with radial basis function kernel[18]	73.9%	Not provided
probabilistic modeling approach[19]	69%	Not provided
ulti-Branch Ferns-based Naive Bayesian classifier[20]	75.8%	Not provided

accounted for 73.53%, much higher than the negative samples.

In order to improve the prediction ability of negative samples, we need to give a further classification method for negative samples. The composition of the new training dataset S* is as follows: partially correctly classified positive samples (28 examples) and all misclassified negative samples (76 examples). Using the new dataset S* to train the optimization (7) or (8). The following fuzzy quadratic discriminant function:

$$g(\tilde{X}) = \frac{1}{2}\left(\tilde{x}_{i1}, \tilde{x}_{i2}, x_{i3}\right) \begin{pmatrix} 0.005173 & 0.001822 & 0.002868 \\ 0.001822 & -0.002914 & 0.00227 \\ 0.002868 & 0.00227 & 0.001151 \end{pmatrix} \begin{pmatrix} \tilde{x}_{i1} \\ \tilde{x}_{i2} \\ x_{i3} \end{pmatrix}$$

$$+ \left(-0.4488\ 0.05937\ -0.3003\right) \begin{pmatrix} \tilde{x}_{i1} \\ \tilde{x}_{i2} \\ x_{i3} \end{pmatrix} + 11.2401.$$

Using the above decision rule to classify the samples in the new dataset S*, the total ACC is 97.13%, only two negative examples are misclassified, and the classification accuracy for negative examples is 97.37%.

The following conclusions can be drawn from the two applications. Our proposed model has better training accuracy and test performance. A two-stage classification algorithm can be utilized for the imbalanced dataset.

5. Conclusions

The classification model is widely used in healthcare management. SVM is a widely used statistical tool for classification and is a training algorithm for learning classification rules from data. But the performance of SVM is closely related to the selection of an appropriate kernel function or the kernel parameters, existing kernel-free SVM are all based on real-value training data. This paper proposes a kernel-free fuzzy quadratic surface SVM (FQSSVM), in which the training dataset is fuzzy training data. The main innovations are as follows:

- Give the concept of fuzzy quadratically separable and its equivalent expression.
- Introduce the model of soft kernel-free fuzzy quadratic surface SVM.
- Present the algorithm of soft kernel-free fuzzy quadratic surface SVM.
- Apply our proposed model in medical diagnosis to verify the validity of the method.

In further study, we are to extend our proposed model to a semi-supervised model and give intelligent algorithms.

References

[1] Ai-bing Ji, Yanhua Qiao(2019)Fuzzy DEA-based classifier and its applications in healthcare management, Health Care Management Science, vol 22, No.3:560–568.

[2] V.N. Vapnik(1995), The Nature of Statistical Learning Theory, Springer-Verlag, New York.

[3] V.N. Vapnik(1998), Statistical Learning Theory, Wiley, New York.

[4] Bernhard Scholkopf and Alexander J Smola(2001), Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

[5] Valluru D, Jeya IJS(2020), IoT with cloud based lung cancer diagnosis model using optimal support vector machine, Health Care Management Science, Vol.23,4: 670-679.

[6] Issam Dagher(2008), Quadratic kernel-free non-linear support vector machine, Journal of Global Optimization, 41(1): 15–30.

[7] Jian Luo, Shu-Cherng Fang, Zhibin Deng, and Xiaoling Guo(2016), Soft quadratic surface support vector machine for binary classification. Asia-Pacific Journal of Operational Research, 33(06): 1650046.

[8] Bai Y, Han X, Chen T, Yu H(2015), Quadratic kernel-free least squares support vector machine for target diseases classification. J Comb Optim 30(4): 850–870.

[9] Astorino A, Fuduli A(2015b), Support vector machine polyhedral separability in semisupervised learning, J Optim Theory Application 164(3): 1039–1050.

[10] Gao, Zheming et al(2021), A novel kernel-free least squares twin support vector machine for fast and accurate multi-class classification, Knowledge-Based Systems, European J. Oper. Res., Vol.290,1: 248-262.

[11] Junyou, Ye et al(2022), Kernel-free soft quadratic surface support vector regression, Information Sciences, Vol.594: 177-199.

[12] Ye, Junyou,Yang Zhixia, Li, Zhilin(2021), Quadratic hypersurface kernel-free least squares support vector regression, Intelligent Data Analysis, Vol.25, 2: 265-281.

[13] Ye Tian, Ziyang Yong, Jian Luo(2018), A new approach for reject inference in credit scoring using kernel-free fuzzy quadratic surface support vector machines, Applied Soft Computing, Vol.73, 1: 96-105.

[14] Jian Luo et al.(2016), Fuzzy quadratic suraface support vector machine based on fisher discriminant analysis,Journal of industratial and management optimization, Vol 12, 1: 357–373.

[15] D. Dubois, H. Prade(1980), Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York.

[16] Ai-bing Ji, Jia-hong Pang, Hong-jie Qiu(2010), Support vector machine for classification based on fuzzy training data. Expert Syst Appl 37(4): 3495–3498.

[17] Liu B(2002), Theory and practice of uncertain programming, Heidelberg: Physica.

[18] Cheng Tang et al.(2020), A novel machine learning technique for computer-aided diagnosis, Engineering Applications of Artificial Intelligence, Vol 92: 103627.

[19] Shuaiyu Yaoa, Jian-Bo Yang, Dong-Ling Xu(2021), A probabilistic modeling approach for interpretable data inference and classification, Journal of Intelligent, Fuzzy Systems 40: 5101–5117.

[20] U. B. Angadi1, Anil Rai1, G. Uma, MBFerns(2021), Classification and extraction of actionable knowledge using Multi-Branch Ferns-based Naive Bayesian classifier, Soft Computing, 25: 8357–8369.