CORPORATE AND PERSONAL CREDIT SCORING VIA FUZZY NON-KERNEL SVM WITH FUZZY WITHIN-CLASS SCATTER

JIAN Luo
School of Management Science and Engineering
Dongbei University of Finance and Economics
Dalian 116025, China

XUEQI YANG
School of Management Science and Engineering
Dongbei University of Finance and Economics
Dalian 116025, China

YE TIAN*
School of Business Administration and
Collaborative Innovation Center of Financial Security
Southwestern University of Finance and Economics
Chengdu 611130, China

WENWEN YU
School of Management Science and Engineering
Dongbei University of Finance and Economics
Dalian 116025, China

(Communicated by Wenxun Xing)

Abstract. Nowadays, the effective credit scoring becomes a very crucial factor for gaining competitive advantages in credit market for both customers and corporations. In this paper, we propose a credit scoring method which combines the non-kernel fuzzy 2-norm quadratic surface SVM model, T-test feature weighting strategy and fuzzy within-class scatter together. It is worth pointing out that this new method not only saves computational time by avoiding choosing a kernel and corresponding parameters in the classical SVM models, but also addresses the “curse of dimensionality” issue and improves the robustness. Besides, we develop an efficient way to calculate the fuzzy membership of each training point by solving a linear programming problem. Finally, we conduct several numerical tests on two benchmark data sets of personal credit and one real-world data set of corporation credit. The numerical results strongly demonstrate that the proposed method outperforms eight state-of-the-art and commonly-used credit scoring methods in terms of accuracy and robustness.

2020 Mathematics Subject Classification. Primary: 62H30; Secondary: 90C20.
Key words and phrases. Credit scoring, non-kernel SVM, fuzzy membership function, feature weighting, fuzzy within-class scatter.

The first author is supported by NNSFC grant # 71701035 and # 71831003.
* Corresponding author.
1. **Introduction.** Many financial institutions have lost much from a steady growth in customers’ defaults on loan in the past several financial crises. However, in order to gain high profits, those credit granting institutions cannot simply refuse growing credit applications from both customers and corporations. To control the credit risk, most financial institutions use the credit scoring models to make credit granting decisions. Therefore, an effective credit risk assessment model plays a significant role in gaining competitive advantages in the growing credit market nowadays.

In the past few decades, many quantitative methods have been introduced to improve the credit scoring accuracy. One primary stream of credit scoring models is using statistical methods. These methods are relatively easy to implement and capable of generating straightforward results. In 1936, Fisher discriminant analysis (FDA) was proposed as the first statistical method for credit scoring [7]. Then logistic regression method was introduced to the “bank crisis early warning classification” in 1977 [19]. Compared to Fisher discriminant analysis, logistic regression model achieves great improvements in the classifying accuracy [19]. Besides, since logistic regression is simple and easy to be illustrated, it became a primary approach for real-world credit scoring [23]. The main statistical methods for credit scoring are summarized in [9]. However, these statistical methods have a few limitations. The first fatal issue is the “curse of dimensionality”, which indicates that the results may become misleading due to the multi-collinearity between variables. Therefore, lots of data preprocessing efforts need to be done on choosing the variables. Additionally, all reported statistical models in [9] rely on one or more hypotheses. In practical applications, those hypotheses such as “the dependent variable should follow a logic normal distribution”, are difficult to be held. Moreover, there exist difficulties for these statistical models to automate the modeling process and remain robust.

In response to the credit scoring accuracy, researchers found out that support vector machine (SVM) is quite an effective tool in credit scoring. A benchmarking study of seventeen different classifying methods on eight various real-world credit data sets was presented in [8]. Note that, least square SVM with linear and Gaussian kernel were utilized in this study. From the numerical results, the least square SVM model overall outperforms all other state-of-the-art methods. Combing the resampling techniques proposed in [18], the SVM model performs better for credit scoring. Then the clustered SVM model is proposed in [11] to greatly improve its efficiency while achieving good classifying accuracy performance. Moreover, for corporate credit ratings, [26] showed that SVM beats neural networks and other state-of-the-art learning algorithms.

Moreover, since irrelevant input features may mess with the underlying structure of a data set and reduce the classification accuracy, many novel feature selecting and weighting strategies have been proposed to handle this problem. Note that, the feature selection and weighting process can also address the issue of “curse of dimensionality” for some large-scaled data sets at the same time. A hybrid strategy which combines the genetic algorithm and SVM models, then simultaneously does the feature selection task and tunes model parameters was proposed in [12]. As a classical method, orthogonal dimension reduction was proposed for selecting the features in [10]. The costs of feature selection become obvious. To fully utilize the information in all input features and reduce the costs of feature selection, a weighted 2-norm SVM (W2NSVM) model incorporating the feature weighting strategy was proposed for credit scoring in [27].
Notice that, the classifying performance of general SVM classifiers could be greatly damaged by outliers and noise points. To reduce the impact of erroneous points and handle the unbalanced data better, the fuzzy (or weighted) SVM (FSVM) is introduced for credit scoring [27] and bankruptcy prediction [22] by assigning each point a fuzzy membership. Many fuzzy membership functions have been proposed for enhancing the noise resistance [13]. Inspired by the Fisher discriminant analysis [7], minimizing the within-class scatter of points has been proved very useful in handling data sets with many outliers and noise points [4]. Hence, [1] followed this idea and proposed a FSVM model with within-class scatter. Note that, this model can produce more accurate classifications than other well-known FSVM models.

To deal with linearly inseparable data sets, a general way is to use kernel functions in SVM models [8, 12, 15, 27]. Note that, kernel functions map all points from original space to a higher dimensional space. Then SVM models separate these mapped points into two classes by a hyperplane in the new space. It is worth pointing out that the kernel and its corresponding parameters greatly influence the credit scoring results. However, there lacks a universal rule to automatically choose a proper kernel for a given data set. It may take a significant amount of computational time and effort to select a proper kernel and its parameters. To overcome these drawbacks of kernel-SVM models, a non-kernel quadratic surface SVM (QSSVM) model which directly uses a quadratic surface in the original space was proposed in [17]. The experimental results show that the non-kernel QSSVM model outperforms the SVM model with Gaussian or quadratic kernel in terms of classifying accuracy and efficiency. A least square QSSVM model was proposed in [3] for target disease classifications and achieved good performance. Moreover, to handle outliers and noise points well, a fuzzy QSSVM model with within-class scatter was proposed in [16] to improve the classifying performance of QSSVM model. Besides, a new approach based on the QSSVM model and fuzzy set was proposed in [21] to handle classification problems with mislabeled information. Because of the good performance and applicability, we choose non-kernel QSSVM model as the fundamental tool for credit scoring in this paper.

This paper mainly contributes to the field on the following four aspects: (1) We propose a new non-kernel nonlinear SVM model for direct credit scoring. Compared with the kernel-based SVM models, the non-kernel SVM model can save much computational time and improve the classifying accuracy; (2) A fuzzy non-kernel QSSVM model is proposed by incorporating the 2-norm measure of misclassification errors and “fuzzy within-class scatter” to handle the outliers and noise points. And an efficient way is developed to calculate the fuzzy memberships of training points by solving an LP problem; (3) To enhance the credit scoring performance, the T-test feature weighting strategy is also included in the proposed model to address the “curse of dimensionality” issue and extract the most useful information from input features; (4) The comprehensive numerical experiments, tested on two benchmark personal credit data sets and one real-world corporation credit data set, validate the superior performance of proposed method over eight well-known methods in terms of accuracy and robustness.

The rest of the paper is arranged as follows. Section 2 reviews the soft QSSVM model and develops an efficient way to calculate the fuzzy memberships. In Section 3, for credit scoring, we propose one fuzzy non-kernel QSSVM model by incorporating “fuzzy within-class scatter” of training points, the T-test feature weights and 2-norm of misclassification errors. Furthermore, the numerical results of the
The proposed model on two benchmark data sets of personal credit and one real-world data set of corporation credit are shown in Section 4. Finally, Section 5 concludes the paper.

2. Fuzzy membership function. In this section, we develop an efficient way to calculate the fuzzy membership for each training point.

2.1. Review of soft QSSVM model. Given the training data set \( \{(x^i, y^i), i = 1, \ldots, n\} \), where \( x^i = [x^i_1, x^i_2, \ldots, x^i_m]^T \in \mathbb{R}^m \), \( y^i = +1 \) or \(-1\) indicates the class label of point \( x^i \) for Class \( C_1 \triangleq \{x | y^i = +1, j = 1, \ldots, n\} \) or Class \( C_2 \triangleq \{x | y^i = -1, j = 1, \ldots, n\} \). The number of elements in \( C_1 \) and \( C_2 \) are denoted as \( n_1 \) and \( n_2 \), respectively, then \( n_1 + n_2 = n \). The soft QSSVM model is to determine the parameter set \((Q, f, c)\) of a quadratic surface

\[ q(x) \equiv \frac{1}{2} x^T Q x + f^T x + c = 0, \]

where \( Q = Q^T = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1m} \\ q_{12} & q_{22} & \cdots & q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{1m} & q_{2m} & \cdots & q_{mm} \end{bmatrix} \in \mathbb{R}^{m \times m}, \ f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} \in \mathbb{R}^m \]

and \( c \in \mathbb{R} \), that separates the \( n \) training points into two classes with maximum margin. Since there is no special restriction on \( Q, f, c \), the quadratic classifier can approximate many types of data very well. Above all, the soft QSSVM model [17] is formulated as:

\[
\begin{align*}
\min & \sum_{i=1}^{n} \|Qx^i + f\|_2^2 + \eta \sum_{i=1}^{n} \xi_i \\
\text{s.t.} & \ y^i \left( \frac{1}{2} (x^i)^T Q x^i + f^T x^i + c \right) \geq 1 - \xi_i, i = 1, \ldots, n, \\
& \xi_i \geq 0, i = 1, \ldots, n, \\
& Q = Q^T \in \mathbb{R}^{\frac{m(m+1)}{2}}, \ (f, c) \in \mathbb{R}^{m+1}.
\end{align*}
\]

where \( \xi_i \) denotes the misclassification error for point \( x^i \) and \( \eta > 0 \) is a penalty constant. For simplicity, the model (SQSSVM) can also be reformulated as follows [17]. First, the vector \( \bar{q} \) is formed by including the \( \frac{m^2 + m}{2} \) elements of upper triangle part of matrix \( Q \), i.e.,

\[ \bar{q} \triangleq [q_{11}, q_{12}, \cdots, q_{1m}, q_{22}, q_{23}, \cdots, q_{2m}, \cdots, q_{mm}]^T \in \mathbb{R}^{\frac{m^2 + m}{2}}. \]

Then the \( m \times \left( \frac{m^2 + m}{2} \right) \) matrix \( M^i \) is constructed for training point \( x^i = [x^i_1, x^i_2, \cdots, x^i_m]^T \in \mathbb{R}^m \) as follows. We first check each element of \( \bar{q} \) in turn, then assign \( x^i_k \) as the \( p \)-th element of \( j \)-th row of \( M^i \) if the \( p \)-th element of \( \bar{q} \) is \( q_{jk} \) or \( q_{kj} \) for some \( k = 1, 2, \cdots, m \), otherwise assign 0. Afterwards, let \( H^i \triangleq [M^i, I] \in \mathbb{R}^{m \times (\frac{m^2 + m}{2} + m)}, i = 1, \cdots, n \) and then \( W \triangleq \sum_{i=1}^{n} (H^i)^T H^i \in \mathbb{R}^{(\frac{m^2 + 3m}{2}) \times (\frac{m^2 + 3m}{2})}, \) where \( I \) denotes the identity matrix. Note that, \( W \) is positive semidefinite. Finally,
we define the variable vector \( z = [q \ f] \in \mathbb{R}^{m^2 + 2m} \) and vector
\[
s^i = [\frac{1}{2} x_1^i x_1^i, \ldots, x_1^i x_m^i, \frac{1}{2} x_2^i x_2^i, \ldots, x_2^i x_m^i, \ldots, \frac{1}{2} x_m^i x_m^i, x_1^i x_m^i, \ldots, x_m^i x_m^i] \in \mathbb{R}^{(m+1)m + m}.
\]

Hence the model (SQSSVM) can be equivalently reformulated as
\[
\begin{align*}
\min_{z} & \quad z^T W z + \eta \sum_{i=1}^{n} \xi_i \\
\text{s.t.} & \quad y^i (s^i)^T z + c \geq 1 - \xi_i, i = 1, \ldots, n, \\
& \quad (z, c) \in \mathbb{R}^{m^2 + 2m + 1}, \xi_i \geq 0, i = 1, \ldots, n.
\end{align*}
\]
(SQSSVM')

It should be noted that the (SQSSVM') model is a convex problem which can be solved efficiently.

2.2. An LP model for approximating soft QSSVM model. If \( W \) is not positive definite, one perturbation can be appended such that the matrix \( W + \epsilon I \) \((\epsilon > 0)\) becomes positive definite. Without loss of generality, suppose that \( W \) is positive definite [17]. Then we can utilize the Lagrangian duality theory to formulate the dual problem of model (SQSSVM') as follows.
\[
\begin{align*}
\max_{\beta_i, i = 1, \ldots, n} & \quad \sum_{i=1}^{n} \beta_i - \frac{1}{4} (\sum_{i=1}^{n} \beta_i y_i s_i)^T W^{-1} (\sum_{i=1}^{n} \beta_i y_i s_i) \\
\text{s.t.} & \quad \sum_{i=1}^{n} \beta_i y_i = 0, \\
& \quad 0 \leq \beta_i \leq \bar{\eta}, i = 1, \ldots, n.
\end{align*}
\]
(DSQSSVM)

Suppose that the primal optimal solution of problem (SQSSVM') is \((z^*, c^*, \xi^*)\) and the dual optimal solution of problem (DSQSSVM) is \(\beta^* = (\beta_1^*, \beta_2^*, \ldots, \beta_n^*)\). Then the following relationship between the primal and dual optimal solutions can be obtained by the Lagrangian dual theory [5]:
\[
\begin{align*}
z^* &= \frac{1}{2} W^{-1} \sum_{i=1}^{n} \beta_i^* y_i s_i \\
\sum_{i=1}^{n} \beta_i^* = 1
\end{align*}
\]
(1)

The optimal values of the primal and dual problems are also equivalent to each other as follows.
\[
\begin{align*}
z^T W z^* + \eta \sum_{i=1}^{n} \xi_i^* &= \sum_{i=1}^{n} \beta_i^* - \frac{1}{4} (\sum_{i=1}^{n} \beta_i^* y_i s_i)^T W^{-1} (\sum_{i=1}^{n} \beta_i^* y_i s_i) \\
&= \sum_{i=1}^{n} \beta_i^* - z^T W z^*,
\end{align*}
\]
(2)

By replacing the right hand of (2) with (1), we have
\[
z^T W z^* + \eta \sum_{i=1}^{n} \xi_i^* = \frac{1}{2} \left( \sum_{i=1}^{n} \beta_i^* + \eta \sum_{i=1}^{n} \xi_i^* \right)
\]
(3)

Hence, this quadratic programming problem (SQSSVM') can be transferred to one LP problem heuristically for simplicity [20]. The objective function in problem (SQSSVM') is first replaced by \(\sum_{i=1}^{n} \beta_i + \eta \sum_{i=1}^{n} \xi_i\) (similar as expression (3)); then \(z\) is substituted by \(\frac{1}{2} W^{-1} \sum_{i=1}^{n} \beta_i y_i s_i\) in the constraints of (SQSSVM') (similar as
expression (1)); finally the constraints \( \dot{\eta} \geq \beta_i \geq 0, i = 1, \ldots, n \), are added to obtain the following LP problem:

\[
\begin{align*}
\min_{\xi_i, \beta_i, i = 1, \ldots, n} & \quad \sum_{i=1}^{n} \beta_i + \dot{\eta} \sum_{i=1}^{n} \xi_i \\
\text{s.t.} \quad & \quad y^i (\frac{1}{2} \sum_{j=1}^{n} (y^j (s^j)^T W^{-1} s^j) \beta_j + c) \geq 1 - \xi_i, \\
& \quad \xi_i \geq 0, \dot{\eta} \geq \beta_i \geq 0, \ i = 1, \ldots, n.
\end{align*}
\]

Generally, the CPU time of solving problem (LPM) is much less than that of solving problem (SQSSVM'). After solving problem (LPM) to get the optimal solution \((\beta, \bar{c}, \hat{\xi})\), \(\bar{z}\) can be calculated by \(\bar{z} = \frac{1}{2} W^{-1} \sum_{i=1}^{n} \beta_i y^i s^i\). Then the parameter set \((\bar{z}, \bar{c})\) of a separating quadratic surface can be obtained.

2.3. Review of quadratic center surface and quadratic-margin distance.

In FSVM models [14, 25], authors utilize the Euclidean distance to measure the distance between two training points, then calculate the central point of each class as the mean of all training points in this class. However, for many general cases, if the training data set is not separable by one hyperplane but separable by one quadratic surface, then the mean of all training points in one class may be within the other class. Thus the mean of all training points in one class is not appropriate for being the center of the class. Therefore, it is necessary to introduce the quadratic center surface of one class and quadratic-margin distance in [16] as follows.

**Definition 2.1.** In the soft QSSVM settings, \(q(x) = \frac{1}{n_1} \sum_{j:y^j = +1} q(x^j)\) (or \(q(x) = \frac{1}{n_2} \sum_{j:y^j = -1} q(x^j)\)) is called as the quadratic center surface of Class \(C_1\) (or \(C_2\) with respect to \(q(x) = 0\).

**Definition 2.2.** In the soft QSSVM settings, for \(y^i\) being +1 (or \(-1\)), \(|q(x^i) - \frac{1}{n_1} \sum_{j:y^j = +1} q(x^j)|\) (or \(|q(x^i) - \frac{1}{n_2} \sum_{j:y^j = -1} q(x^j)|\)) is called as the quadratic-margin distance between point \(x^i\) and the quadratic center surface of its related class with respect to \(q(x) = 0\).

2.4. A way to calculate fuzzy memberships. In this paper, we calculate the fuzzy membership of each training point to obtain its relative importance, based on the quadratic-margin distance between each point and the quadratic center surface of its related class as follows. Problem (LPM) is first solved to obtain the parameters \((\bar{z}, \bar{c})\) of an effective separating surface. Then, with regard to this separating surface, we calculate the mean functional margins of all points in Class \(C_1\) and \(C_2\) as \(\alpha_1 \triangleq \frac{1}{n_1} \sum_{j:y^j = +1} (s^j)^T \bar{z} + \bar{c} \geq 0\) and \(\alpha_2 \triangleq -\frac{1}{n_2} \sum_{j:y^j = -1} (s^j)^T \bar{z} - \bar{c} \geq 0\), respectively. Also, as in Definition 2.2, the quadratic-margin distance \(d(x^i)\) between the training point \(x^i\) and its related quadratic center surface is calculated by

\[
d(x^i) = \begin{cases} 
|\bar{s}^i)^T \bar{z} + \bar{c} - \alpha_1|, & \text{if } y^i = +1, \\
|\bar{s}^i)^T \bar{z} - \bar{c} - \alpha_2|, & \text{if } y^i = -1.
\end{cases}
\]

Then, similar to the fuzzy membership function in [16], for each training point \(x^i\), the following way is introduced to calculate the fuzzy membership \(h^i\) to also
consider the affinity among training points:

\[
\begin{align*}
    h^i & \equiv \left\{ \begin{array}{ll}
    \frac{\hat{p}}{\alpha_1} - d(x^i) + 1 - \hat{p}, & \text{if } y^i = +1 \& d(x^i) \leq \alpha_1, \\
    \frac{1 - \hat{p}}{1 + d(x^i) - \alpha_1}, & \text{if } y^i = +1 \& d(x^i) > \alpha_1, \\
    \frac{\hat{p}}{\alpha_2} - d(x^i) + 1 - \hat{p}, & \text{if } y^i = -1 \& d(x^i) \leq \alpha_2, \\
    \frac{1 - \hat{p}}{1 + d(x^i) - \alpha_2}, & \text{if } y^i = -1 \& d(x^i) > \alpha_2,
    \end{array} \right.
\end{align*}
\]

where the constant \(0.5 < \hat{p} < 1\) is given in advance. Obviously, \(0 \leq h^i \leq 1, \ i = 1, \cdots, n\). From this fuzzy membership function, take training points with label +1 (Class \(C_1\)) as examples, if they are farther away from the related quadratic center surface \(q(x) = \alpha_1\), they are more likely to be noise (or outliers) and less important, hence they have smaller fuzzy memberships. With functional margins being larger than \(2\alpha_1\) or smaller than 0, the training points are supposed to be outliers so that the fuzzy memberships for them are much smaller than those for points with functional margins between 0 and \(2\alpha_1\).

3. Fuzzy non-kernel QSSVM with fuzzy within-class scatter for credit scoring. In this section, for credit scoring, we propose a fuzzy non-kernel QSSVM model which incorporates the fuzzy membership of each training point, “fuzzy within-class scatter” of training points, the 2-norm measure of misclassification errors and the T-test feature weights.

3.1. Review of fuzzy QSSVM model. By incorporating the fuzzy membership \(h^i, i = 1, \cdots, n\), of each training point into the soft QSSVM model, a fuzzy QSSVM model is given below

\[
\begin{align*}
    \min \ & \sum_{i=1}^{n} h^i \|Qx^i + f\|^2 + \hat{\eta}_1 \sum_{i=1}^{n} h^i \xi_i \\
    \text{s.t.} \ & y^i \left( \frac{1}{2} (x^i)^T Qx^i + f^T x^i + c \right) \geq 1 - \xi_i, \quad (FQS1) \\
    \ & \xi_i \geq 0, \ i = 1, \cdots, n, \\
    \ & Q = Q^T \in \mathbb{R}^{m \times m}, \ (f, c) \in \mathbb{R}^{m+1},
\end{align*}
\]

where the penalty constant \(\hat{\eta}_1 \geq 0\) needs to be chosen beforehand as a tradeoff between two terms in the objective function. The slack variable \(\xi_i\) is a measure of misclassification error and the fuzzy membership \(h^i\) is the relative importance of point \(x^i\) toward its corresponding class. The terms \(h^i \xi_i\) and \(h^i \|Qx^i + f\|^2\) can be deemed as a measure of \(\xi_i\) and \(\|Qx^i + f\|^2\) with the weight \(h^i\), respectively. If the training point \(x^i\) is more likely to be an outlier or noise, which indicates the less importance of \(x^i\), then the corresponding \(h^i\) is expected to be smaller to reduce the effect of parameter \(\xi_i\) and expression \(\|Qx^i + f\|^2\) in the model (FQS1).

3.2. Fuzzy non-kernel QSSVM with fuzzy within-class scatter for credit scoring. In this subsection, we first develop the fuzzy within-class scatter of training points. Then, for credit scoring, one fuzzy non-kernel QSSVM with fuzzy within-class scatter by incorporating the fuzzy within-class scatter, T-test feature weights and 2-norm of misclassification errors.
Motivated by the within-class scatter in [7], the fuzzy within-class scatter with respect to a quadratic surface is introduced by considering the relative importance of each training point as follows.

**Definition 3.1.** In the soft QSSVM settings, we define $S(Q, f, c) \triangleq \frac{1}{n} \left( \sum_{i:y_i=+1} \left( h^t q(x_i) - \frac{h^t}{n_1} \sum_{j:y_j=+1} q(x_j) \right)^2 + \sum_{i:y_i=-1} \left( h^t q(x_i) - \frac{h^t}{n_2} \sum_{j:y_j=-1} q(x_j) \right)^2 \right) \in \mathbb{R}$ as the fuzzy within-class scatter of training points with respect to $q(x) = 0$.

From this definition, we can see that the quadratic-margin distance between point $x_i$ and the quadratic center surface of its related class (see Definition 2.2) is scaled by the fuzzy membership $h^i$ of point $x_i$. To increase the class separability, we need to minimize $S(Q, f, c)$ for the following proposed model.

For better performance of credit scoring, we first generate the T-test based feature weights $(\hat{\omega}_j, j = 1, \ldots, n)$ as follows. For $j = 1, 2, \ldots, m$, define $v_j^+$ and $v_j^-$ as the variance of the sets $\{x_i : y_i = +1\}$ and $\{x_i : y_i = -1\}$, respectively; and let $\rho_j^+$ and $\rho_j^-$ be the mean of the sets $\{x_i : y_i = +1\}$ and $\{x_i : y_i = -1\}$, respectively. Then the weight for feature $j$ based on T-test of the training data set is defined as:

$$\hat{\omega}_j = \frac{|\rho_j^+ - \rho_j^-|/\sqrt{v_j^+ n_1 + v_j^- n_2}}{\sum_{k=1}^m (|\rho_k^+ - \rho_k^-|/\sqrt{v_k^+ n_1 + v_k^- n_2})}, \quad j = 1, 2, \ldots, m.$$  

Notice that, the reason why we adopt this method is that it is the most effective feature weighting strategy in the literature [27].

Then, to propose the fuzzy non-kernel QSSVM model with fuzzy within-class scatter for credit scoring, an LP model is first solved to find the parameters of an effective separating surface via incorporating the T-test based feature weights $(\hat{\omega}_j, j = 1, \ldots, n)$. Similar to the proposed model (LPM) for approximating the SQSSVM model in Section 2.2, the following LP model is considered:

$$\min_{\xi, \beta, i = 1, \ldots, n} \sum_{i=1}^n \beta_i + \eta \sum_{i=1}^n \xi_i$$  

s.t. $y^i \left( \frac{1}{2} \sum_{j=1}^m \left( (y^j (\tilde{s}^i)^T \tilde{W}^{-1} \tilde{s}^j) \beta_j + c \right) \right) \geq 1 - \xi_i,$  

$$\xi_i \geq 0, \beta_i \geq 0, \quad i = 1, \ldots, n,$$

where $\tilde{s}^i$ and $\tilde{W}$ are generated, respectively, after substituting $\hat{\omega}_j x_j^i$ for $x_j^i$ in $s^i$ and $W$ (shown in model (SQSSVM$^\prime$) of Section 2.1) for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Similarly, the fuzzy memberships $(h^i, i = 1, \ldots, n)$ of all training points can be obtained by substituting $\hat{\omega}_j x_j^i$ for $x_j^i$ in the fuzzy membership function in Section 2.4. For better credit scoring accuracy [27], we first replace the 1-norm of slack variable vector $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ with the square of its 2-norm in the objective function of fuzzy QSSVM model (FQS1) in Section 3.1, then incorporate the T-test based feature weights $(\hat{\omega}_j, j = 1, \ldots, n)$ and fuzzy within-class scatter $\tilde{S}(Q, f, c)$ (which is generated by replacing $x_j^i$ with $\tilde{s}^i$) to model (FQS1) to
propose the following model:

\[
\begin{align*}
\min & \sum_{i=1}^{n} h_i^2 \|Q(\hat{\omega} \otimes x^i) + f\|_2^2 + \hat{n}_1 \sum_{i=1}^{n} h_i^2 \xi_i^2 + \hat{n}_2 \hat{S}(Q, f, c) \\
\text{s.t.} & \ y^i \left( \frac{1}{2} (\hat{\omega} \otimes x^i)^T Q(\hat{\omega} \otimes x^i) + f^T (\hat{\omega} \otimes x^i) + c \right) \geq 1 - \xi_i, \ i = 1, \ldots, n, \quad \text{(FQS2)}
\end{align*}
\]

where \( \hat{\omega} = (\hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_m) \in \mathbb{R}^m \) is the feature weight vector, \( \hat{\omega} \otimes x^i \triangleq (\hat{\omega}_1 x_1^i, \hat{\omega}_2 x_2^i, \ldots, \hat{\omega}_m x_m^i) \in \mathbb{R}^m \). Similar to the reformulation of model (SQSSVM) in Section 2.1, and let \( \hat{W} \triangleq \sum_{i=1}^{n} h_i (H_i)^T H_i \in \mathbb{R}^{(m^2+2m)} \times (m^2+2m) \) and \( \hat{S} \triangleq \frac{1}{n} \left( \sum_{i:y_i=-1} (h_i s_i^1 - \frac{h_i}{m_1} \sum_{j:y_j=+1} s_j^1) (h_i s_i^1 - \frac{h_i}{m_1} \sum_{j:y_j=+1} s_j^1)^T + \sum_{i:y_i=-1} (h_i s_i^1 - \frac{h_i}{m_2} \sum_{j:y_j=-1} s_j^1) (h_i s_i^1 - \frac{h_i}{m_2} \sum_{j:y_j=-1} s_j^1)^T \right) \in \mathbb{R}^{(m^2+2m)} \times (m^2+2m) \), (FQS2) can be reformulated as

\[
\begin{align*}
\min & \ z^T (\hat{W} + \hat{S}) z + \hat{n}_1 \sum_{i=1}^{n} h_i \xi_i^2 \\
\text{s.t.} & \ y^i ((\hat{S})^T z + c) \geq 1 - \xi_i, \ i = 1, \ldots, n, \quad \text{(FNKSVM-FWS)}
\end{align*}
\]

where \( \hat{W} \) and \( \hat{S} \) are obtained by substituting \( \hat{\omega}_j x_j^i \) for \( x_j^i, \forall i, j \) in \( \hat{W} \) and \( \hat{S} \), respectively. Notice that, this model is a convex problem, whose global optimal solution can be generated efficiently. Moreover, the model (FNKSVM-FWS) is a fuzzy QSSVM model which uses a quadratic surface \( q(x) = 0 \) to measure the fuzzy within-class scatter. The tradeoff between classification margin \( z^T \hat{W} z \) and the fuzzy within-class scatter \( z^T \hat{S} z \) is controlled by the nonnegative constant \( \hat{n}_2 \). In general, a large \( \hat{n}_2 \) will reduce the fuzzy within-class scatter and narrow the classification margin.

### 4. Numerical tests on credit data.

In this section, we investigate the credit scoring performance of the proposed fuzzy non-kernel QSSVM model on some real-world credit data sets. For comparisons, we mainly test three groups of credit scoring methods: The first group includes commonly-used non-SVM methods such as the logistic regression (denoted by “LOG_REG”) and feed-forward backpropagation (FFBP) neural networks (denoted by “FFBP_NN”); the second group includes commonly-used kernel-based SVM methods such as the soft SVM model with Gaussian kernel (denoted by “SVM_GausKer”), the weighted 2-norm SVM model with Gaussian or Quadratic kernel [27] (denoted by “W2NSVM_GausKer” or “W2NSVM_QuadKer”), well-known FSVM model with within-class scatter and Gaussian kernel [1] (denoted by “FSVMWCS_GausKer”) and clustered SVM method [11] (denoted by “Clu_SVM”); the third group includes two non-kernel SVM models such as Dagher’s QSV model [6] and soft QSSVM model [17]. For all SVM models, the grid method is utilized to search the best penalty parameter \( \hat{\eta} \) as the following: \( \log_2 \hat{\eta} \in [2, 3, \ldots, 19, 20] \).

All numerical tests in this paper are conducted via MATLAB (R2014a) on a personal laptop equipped with Intel Core i5 2.40 GHz CPU, 2.5GB usable RAM and Microsoft Windows 7 Professional. The logistic regression and neural networks are implemented using the MATLAB module “glmfit” and “newff”, respectively. The group of non-kernel SVM models are implemented using the interior point
Table 1. Credit Data Sets

| data set  | # of features | Class C₁ | # of points | Class C₂ | # of points |
|-----------|---------------|----------|-------------|----------|-------------|
| German    | 20            | Creditworthy | 700         | Non-creditworthy | 300         |
| Australian| 14            | Non-default | 383         | Default   | 307         |
| Chinese   | 7             | Good credit | 58          | Bad credit | 48          |

For numerical experiments, we utilize two types of credit data sets: One is about the personal credit information, such as the German and Australian applicants from the UCI machine learning repository database [2]; the other one is about the corporation credit information, such as the suppliers from the state grid corporation of China. It is worth pointing out that the information of these suppliers is gathered from Nanping City in Fujian Province. Moreover, to protect the privacies of these suppliers, we delete their names and invited 21 experienced experts to classify them as good suppliers or bad ones based on their different features. For more details, the profiles of three credit data sets are shown in Table 1.

For German and Australian data sets of personal credit information, the common features include personal sex, marriage, age, salary, credit history, saving account, present residence, present employment and so forth. For Chinese real-world data set, there are some features about the credit information of some corporations, such as the number of cooperation within three years, number of good credit record, number of black list, number of bad delivery, and so forth.

For these three credit data sets, the information in all ordinal and continuous variables are not changed. To reflect the real world, all nominal variables without the logic relationships are transferred into one or several categories accordingly. Then for each category, one dumb variable is used to indicate the related applicant in the specific state (i.e., 1 indicates yes while 0 indicates no). Moreover, for the prepossessed data points \( \{x^i, i = 1, \ldots, n\} \) with all numerical variables, where \( x^i = [x^i_1, x^i_2, \ldots, x^i_m]^T \in \mathbb{R}^m \), we linearly scaled all input features to \([0, 1] \) to avoid the dominance of features with greater values over those with smaller values, i.e.,

\[
x^i_j = \frac{x^i_j - \min_{k \in \{1, \ldots, n\}} x^k_j}{\max_{k \in \{1, \ldots, n\}} x^k_j - \min_{k \in \{1, \ldots, n\}} x^k_j}, \quad i = 1, \ldots, n, j = 1, \ldots, m.
\]

4.2. Numerical experiments on three credit data sets. After prepossessing the three credit data sets, the commonly-used 10-fold cross validation method is utilized to validate the performance of all models. For the German credit data set, the 10-fold cross validation procedure is repeated for 100 times by randomly partitioning the data set into 10 equal-sized parts. For all models, the mean and standard deviation of the calculated misclassification rates and average CPU time of the 100 independent tests are reported in Table 2. Moreover, the reported CPU time of solving the proposed fuzzy non-kernel QSSVM model or weighted 2-norm SVM model includes the time of computing fuzzy memberships of all training points. However, the reported CPU time of all methods doesn’t include the time of tuning the parameters in the neural network, the penalty parameter \( \hat{\eta} \) or kernel parameters in SVM.
Furthermore, using the same experiment procedures as those designed for German credit data set, we conduct experiments for all models on Australian and Chinese credit data set, and record the computational results in Tables 3 and 4, respectively.

From Tables 2-4, we have the following observations and conclusions:
• For the two credit data sets of personal information, the proposed FNKSVM-FWS model produces much more accurate credit scorings than all compared well-known methods.

• For the Chinese credit data set of corporate information, the proposed FNKSVM-M-FWS model produces more accurate credit scorings than all other compared well-known methods besides the logistic regression. The logistic regression slightly outperforms the fuzzy 2-norm QSSVM in terms of accuracy, but the CPU time of logistic regression is longer than that of fuzzy 2-norm QSSVM.

• The proposed FNKFSVM-FWS model outperforms the non-kernel SVM model such as Dagher’s QSVM or soft QSSVM model in terms of credit scoring accuracy. The CPU time of proposed model is less than that of Dagher’s model, although the CPU time of proposed model is more than that of soft QSSVM model since the CPU time of calculating fuzzy memberships is included for proposed model. Notice that, if an LP model is not proposed for approximating soft QSSVM model to improve the efficiency of calculating the fuzzy memberships, the CPU time of proposed model should be at least twice as that of soft QSSVM model.

• All in all, the two smallest CPU time of all methods are that of logistic regression and clustered SVM method. Besides this, the CPU time of FFBP neural networks, weighted 2-norm SVM model with Gaussian or Quadratic kernel, FSVM model with within-class scatter, Dagher’s QSVM model, FNKSVM-FWS model is longer than that of soft SVM model with Gaussian kernel or soft QSSVM model.

• Since the recorded CPU time of each kernel-based SVM model doesn’t include the time of choosing kernel functions and tuning their parameters, the non-kernel SVM model costs much less CPU time than the kernel-based SVM model for real-world credit scoring.

In addition, the robustness of all models is further investigated by utilizing the 10-fold cross validation method on Australian credit data with outliers as follows. First, the Australian credit data set is randomly partitioned into 10 equal-sized parts after the same prepossess procedure in Section 4.1. Then, for each training data set including nine parts of Australian data, the outliers are generated by randomly selecting 35% of all points and changing the labels of these training points. Finally, using the generated training data set with 35% outliers, the 10-fold cross validation procedure is repeated for 100 times. For all models, the mean of calculated misclassification rates of the 100 independent tests are recorded in Table 5. It should be noted that there supposed to be no outliers (or noise) in the original Australian data so that the numerical results of experiments with or without outliers (from Table 3) are also compared in Table 5, to investigate the robustness of all models. From Table 5, we can see that the proposed FNKFSVM-FWS model is the most robust one among all tested models. Moreover, Dagher’s QSVM model produces much less inaccurate credit scores than other tested models mainly due to its lack of slack variables. And the proposed model is more robust than the SQSSVM model, which indicates that the fuzzy memberships and fuzzy within-class scatter can increase the robustness.
### Table 5. Robustness of Models on Australian Credit Data

| model                  | mean of misclassification rates (%) without outliers | mean of misclassification rates (%) with outliers |
|------------------------|------------------------------------------------------|--------------------------------------------------|
| LOG_REG                | 13.56                                                | 17.87                                            |
| FFBP_NN                | 14.42                                                | 15.94                                            |
| SVM_GausKer            | 15.00                                                | 15.80                                            |
| W2NSVM_GausKer         | 14.87                                                | 15.65                                            |
| W2NSVM_QuadKer         | 14.59                                                | 15.36                                            |
| FSVMWCS_GausKer        | 14.63                                                | 18.43                                            |
| Clu_SVM                | 14.34                                                | 17.84                                            |
| Dagher’s QSVM          | 26.42                                                | 53.21                                            |
| SQSSVM                 | 14.57                                                | 15.58                                            |
| FNKSVM-FWS             | 11.96                                                | 12.61                                            |

### 5. Conclusions and discussions.

In this paper, we first develop a new way to efficiently calculate the fuzzy memberships of training points. Then the fuzzy within-class scatter is developed to handle the outliers and noise points well. Most importantly, one fuzzy non-kernel QSSVM model with fuzzy within-class scatter is proposed for credit scoring by incorporating the fuzzy within-class scatter, T-test based feature weights and 2-norm of the slack vector. Finally, the proposed fuzzy 2-norm QSSVM credit scoring method is tested on personal and corporation credit data sets to investigate its performance. Our major findings are summarized as below.

- For credit scoring, without increasing much computational time, the proposed fuzzy non-kernel QSSVM model is much more accurate and robust than other state-of-the-art credit scoring methods including logistic regression, neural networks, non-kernel and kernel-based SVM methods.
- Unlike well-known kernel-based SVM models, the proposed model does not need to use any kernel function or to tune the parameters in the kernel function for credit scoring. Therefore, compared to kernel-based SVMs, the new method saves much more effort and time.

For future research, we are interested in developing fuzzy semi-supervised non-kernel QSSVM methods [24] for credit scoring and other issues of risk management.

### Funding.

Jian’s research has been supported by the National Natural Science Foundation of China Grant # 71701035 and # 71831003. Tian’s research has been supported by the Fundamental Research Funds for the Central Universities # JBK1805005.

### REFERENCES

[1] W. An and M. Liang, Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises, *Neurocomputing*, 110 (2013), 101–110.

[2] K. Bache and M. Lichman, Uci machine learning repository, http://archive.ics.uci.edu/ml, 2013.

[3] Y. Bai, X. Han, T. Chen and H. Yu, Quadratic kernel-free least squares support vector machine for target diseases classification, *Journal of Combinatorial Optimization*, 30 (2015), 850–870.

[4] G. Baudat and F. Anouar, Generalized discriminant analysis using a kernel approach, *Neural Computation*, 12 (2000), 2385–2404.
[5] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, 2004.

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

[7] R. Fisher, The use of multiple measurements in taxonomic problems, *Annals of Human Genetics*, 7 (1936), 179–188.

[8] T. Gestel, B. Baesens and J. Garcia, A support vector machine approach to credit scoring, *Journal of Bank and Finance*, 2 (2003), 73–82.

[9] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd edition, Morgan Kaufmann, San Francisco, CA, 2006.

[10] L. Han and H. Zhao, Orthogonal support vector machine for credit scoring, *Engineering Applications of Artificial Intelligence*, 26 (2013), 848–862.

[11] T. Harris, Credit scoring using the clustered support vector machine, *Expert Systems with Applications*, 42 (2015), 741–750.

[12] C. Huang, M. Chen and C. Wang, Credit scoring with a data mining approach based on support vector machines, *Expert Systems with Applications*, 33 (2007), 847–856.

[13] X. Jiang, Y. Zhang and J. Lv, Fuzzy svm with a new fuzzy membership function, *Neural Computing and Applications*, 15 (2006), 268–276.

[14] C. Lin and S. Wang, Fuzzy support vector machines, *IEEE Transactions on Neural Networks*, 13 (2002), 464–471.

[15] F. Liu and X. Xue, Subgradient-based neural network for nonconvex optimization problems in support vector machines with indefinite kernels, *Journal of Industrial and Management Optimization*, 12 (2016), 285–301.

[16] J. Luo, S.-C. Fang, Y. Bai and Z. Deng, Fuzzy quadratic surface support vector machine based on fisher discriminant analysis, *Journal of Industrial and Management Optimization*, 12 (2016), 357–373.

[17] J. Luo, S.-C. Fang, Z. Deng and X. Guo, Quadratic surface support vector machine for binary classification, *Asia-Pacific Journal Of Operational Research*, 33 (2016), 1650046.

[18] A. Marques, V. Garcia and J. Sanchez, On the suitability of resampling techniques for the class imbalance problem in credit scoring, *Journal of the Operational Research Society*, 64 (2013), 1060–1070.

[19] D. Martin, Early warning of bank failure: a logistic regression approach, *Journal of Banking and Finance*, 1 (1977), 249–276.

[20] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, 2002.

[21] Y. Tian, M. Sun, Z. Deng, J. Luo and Y. Li, A new fuzzy set and non-kernel svm approach for mislabeled binary classification with applications, *IEEE Transactions on Fuzzy Systems*, 25 (2017), 1536–1545.

[22] W. Tunga, C. Queka and P. Cheng, Genso-ews: A novel neural-fuzzy based early warning system for predicting bank failures, *Neural Networks*, 17 (2004), 567–587.

[23] J. Wiginton, A note on the comparison of logic and discriminant models of customer credit behavior, *Journal of Financial and Quantitative Analysis*, 15 (1980), 757–770.

[24] X. Yan, Y. Bai, S.-C. Fang and J. Luo, A kernel-free quadratic surface support vector machine for semi-supervised learning, *Journal of the Operational Research Society*, 67 (2016), 1001–1011.

[25] X. Zhang, X. Xiao and G. Xu, Fuzzy support vector machine based on affinity among samples, *Journal of Software*, 17 (2006), 951–958.

[26] H. Zhong, C. Miao, Z. Shen and Y. Feng, Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings, *Neurocomputing*, 128 (2014), 285–295.

[27] L. Zhou, K. Lai and J. Yen, Credit scoring models with auc maximization based on weighted svm, *International Journal of Information Technology and Decision Making*, 4 (2009), 677–696.

Received August 2018; revised March 2019.

E-mail address: luojian546@dufe.edu.cn
E-mail address: SherryYangxq@outlook.com
E-mail address: yetian@swufe.edu.cn
E-mail address: 18840804432@163.com