Article

A Quadratic Surface Minimax Probability Machine for Imbalanced Classification

Xin Yan [1],*, Zhouping Xiao [1] and Zheng Ma [2]

[1] School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China
[2] School of Management, Shanghai University of International Business and Economics, Shanghai 201620, China
* Correspondence: yanxin@suibe.edu.cn

Abstract: In this paper, a kernel-free minimax probability machine model for imbalanced classification is proposed. In this model, a quadratic surface is adopted directly for separating the data points into two classes. By using two symmetry constraints to define the two worst-case classification accuracy rates, the model of maximizing both the F_1 value of the minority class and the classification accuracy rate of all the data points is proposed. The proposed model corresponds to a fractional programming problem. Since the two worst-case classification accuracy rates are the symmetry, the proposed model can be further simplified. After this, the alternating descent algorithm is adopted for efficiently solving. The proposed method reduces the computational costs by both using the kernel-free technique and adopting the efficient algorithm. Some numerical tests on benchmark datasets are conducted to investigate the classification performance of the proposed method. The numerical results demonstrate that the proposed method performs better when compared with the other state-of-the-art methods, especially for classifying the imbalanced datasets. The better performance for the imbalanced classification is also demonstrated on a Wholesale customers dataset. This method can provide methodological support for the research in areas such as customer segmentation.

Keywords: imbalanced classification; quadratic surface; minimax probability machine

1. Introduction

Binary classification is an important branch in the field of machine learning. Minimax probability machine (MPM), which is first proposed in [1], is a competitive machine learning method for binary classification problems. It directly maximizes the correct classification probability of the worst-case with a given mean and covariance matrix [1]. In other words, MPM provides a clear lower bound on prediction accuracy without making any assumptions on data distribution. The robustness of the MPM-based methods have been shown with successful applications in the fields such as medical diagnosis [2], anomaly detection [3], and customer churn prediction [4].

Since the role of two classes are often symmetry, the same kind of modeling technique can be used. In the traditional MPM model, the lower bounds of correct classification probability for two symmetry categories are defined to be the same [1]. It may not be effective enough when a certain class of samples needs more attention. The minimum error minimax probability machine (MEMPM) [5] learns the classifier by maximizing the lower bound of the weighted correct classification probability of two categories, which provides different approaches to different classes of samples. After setting the weight to be $1/2$, a biased minimax probability machine is proposed in [6] to deal with the problem of imbalanced learning. Although the balanced parameter in MEMPM can deal with the different importance of two classes to some extent, only the accuracy rate is adopted as the objective function. Luo et al. [7] developed a new minimax probability machine for the F_β measure (MPMF), which provided a new perspective for the classification on imbalanced datasets. By using the lower bound of the

correct classification probability of the two classes, the F_β measure is defined as the objective function. To solve the MPMF model effectively, its equivalent form, which can be solved by an alternating descent method to learn a linear classifier, is derived. The kernel trick is employed to derive a nonlinear MPMF model to learn a nonlinear classifier. Although MPMF provides a good perspective for dealing with imbalanced datasets, the performance also depends on the choice of kernel functions and kernel parameters, a process that consumes a lot of time.

Some other MPM-based models have also been studied. Nath et al. proposed a maximum margin classifier [8] by considering the structural risk of the model. Maldonado et al. [9] proposed the regularized MPM and regularized MEMPM models, which minimized both the l_2 norm of the weight vector and the upper bound on the classification error rate. Drawing on the idea of twin support vector machines [10], Yang et al. [11] proposed the twin minimax probability machine model, which attempts to construct a pair of non-parallel decision hyperplanes such that each hyperplane separates one class of samples with a lower bound on the classification accuracy and is far from the other classes of samples. Although these methods show their good performance for classification, kernel functions are all needed to deal with nonlinear classification problems. This process may run into the efficiency issues of choosing kernel functions and their corresponding parameters.

For dealing with the issues of choosing a suitable kernel function and its optimal parameter, Dagher proposed a kernel-free quadratic surface support vector machine model [12] for binary classification. Luo et. al. extended it to a soft-margin quadratic surface SVM (SQSSVM) model [13]. Furthermore, more kernel-free SVM models are proposed in the literature. Bai et. al. proposed a new quadratic kernel-free least squares support vector machine for a binary classification problem [14]. Yan et. al. proposed a kernel-free semi-supervised quadratic surface support vector machine model [15]. Tian et. al. proposed a new approach based on the kernel-free quadratic surface support vector machine model to handle a binary classification problem with mislabeled information [16]. Luo et. al. proposed an unsupervised quadratic surface support vector machine and applied it to credit risk assessment [17]. Gao et. al. proposed a kernel-free soft quartic surface SVM model by utilizing the double well potential function for highly nonlinear binary classification [18]. Yan et. al. analyzed the equivalent form of the soft quadratic surface support vector machine and proposed a novel support vector machine model with feature mapping [19]. Based on the idea of a soft quadratic surface support vector machine [13], Wang et al. [20] proposed a kernel-free quadratic surface minimax probability machine (QSMPM), which attempts to find a quadratic surface to separate two classes of samples. QSMPM does not require use of the nonlinear mapping for dealing with nonlinear separable datasets, which saves a lot of time for tuning the parameters. Nevertheless, the QSMPM model still aims to maximize the lower bound of the correct classification probability of two classes, which is not suitable enough for dealing with imbalanced classification.

The existing minimax probability machines only focus on kernel-free schemes or imbalanced classification. Meanwhile, the existing optimization goal considers only one certain classification performance measure. For efficiently dealing with imbalanced classification problems, a new quadratic surface minimax probability machine model with the optimization goal that considers two measures (QSMPMFA) is proposed. The optimization goal consists of both the F_1 value of the minority class and the weight classification accuracy rate of the two classes. In this model, QSMPMFA tries to find a quadratic surface to separate the samples into two classes. Two symmetry constraints are used to define the lower bound of the probability of different classes of samples being correctly classified, i.e., worst-case classification accuracy rates. Since two worst-case classification accuracy rates are symmetry in the constraint, the proposed model can be further simplified. The original optimization problem of the QSMPMFA can then be transformed into an equivalent form, which can be solved by the alternating descent method. The advantages can be summarized as follows.

First, QSMPMFA is optimized by using the combination of the F_1 value and accuracy rate in the objective function, instead of considering only one certain classification performance measure, causing it to be better for handling imbalanced datasets.

Second, the proposed model can be further simplified by using the symmetry structure in the constraint. The alternating descent algorithm is adopted for efficiently solving.

Third, the numerical tests of QSMPMFA on several benchmark datasets show that QSMPMFA outperforms the other state-of-the-art methods in terms of both the F_1 value and accuracy rate, especially for the imbalanced datasets.

This paper is structured as follows. In Section 2, the related studies are briefly reviewed, including MPM, SQSSVM, and the formulation of the F_1 value. Section 3 provides the details of the proposed QSMPMFA model and the corresponding algorithm. Section 4 shows the numerical tests on different kinds of datasets and illustrates the comparison performance of the proposed method with other methods. Section 5 provides some conclusions and future work.

2. Related Work

This section provides a brief introduction to the minimax probability machine, soft quadratic surface support vector machine, and the performance measures.

Consider the dataset of $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$, where $x_i \in \mathbb{R}^d$ refers to the feature vector of the i-th sample and $y_i \in \{+1, -1\}$ represents this sample's class label for $i = 1, 2, \ldots, l$.

2.1. Minimax Probability Machine (MPM)

MPM attempts to separate the two classes of data points with the goal of maximizing a lower bound on the classification accuracy rate given the mean and covariance matrix of each class of data.

Let $x \sim (\mu, \Sigma)$ be the random variable x with mean μ and covariance matrix Σ. Suppose the positive class sample $x_+ \sim (\mu_+, \Sigma_+)$, the negative class sample $x_- \sim (\mu_-, \Sigma_-)$, μ_+, $\mu_- \in \mathbb{R}^d$, and $\Sigma_+, \Sigma_- \in \mathbb{S}_+^d$. Using MPM tries to find the hyperplane:

$$g(x) = w^T x - b = 0, \tag{1}$$

where $w \in \mathbb{R}^d$, $b \in \mathbb{R}$. The optimization problem of MPM can be formulated as following:

$$
\begin{aligned}
\max_{w,b,\alpha} \quad & \alpha \\
s.t. \quad & \inf_{x_+ \sim (\mu_+, \Sigma_+)} Pr\{w^T x_+ \geq b\} \geq \alpha, \\
& \inf_{x_- \sim (\mu_-, \Sigma_-)} Pr\{w^T x_- \leq b\} \geq \alpha.
\end{aligned}
\tag{2}
$$

In Problem (2), $\alpha \in (0, 1)$ is the lower bound on the probability of the two classes of samples being correctly classified, which can also be explained as the worst-case classification accuracy rate. The structure of the two constraints are symmetry. By maximizing α in the objective function, the hyperplane with the best lower bound can be chosen.

Problem (2) is an optimization problem with probability constraints, which causes the problem to be difficult to deal with. The following two Lemmas are introduced for transformation.

Lemma 1. *Suppose x is a random vector, S is a given convex set, and $x \sim (\mu, \Sigma)$, then the supremum on the probability of $x \in S$ is equivalent to:*

$$\sup_{x \sim (\mu, \Sigma)} Pr\{x \in S\} = \frac{1}{1 + d^2}, \quad d^2 = \inf_{x \in S} (x - \mu)^T \Sigma^{-1} (x - \mu). \tag{3}$$

Lemma 2. *Given $w \neq 0, b, \mu, \Sigma$, the following condition:*

$$\inf_{x \sim (\mu, \Sigma)} Pr\{w^T x \leq b\} \geq \alpha \tag{4}$$

holds if and only if:

$$b - w^T x \geq k(\alpha)\sqrt{w^T \Sigma w},\tag{5}$$

where $k(\alpha) = \sqrt{\frac{\alpha}{1-\alpha}}$.

It is obvious that $k(\alpha)$ is a monotonically increasing function of α, so the original optimization problem of MPM can be transformed into the following formulation according to the above two lemmas.

$$\max_{w,b,\alpha} \quad k(\alpha)$$
$$s.t. \quad -b + w^T \mu_+ \geq k(\alpha)\sqrt{w^T \Sigma_+ w},\tag{6}$$
$$b - w^T \mu_- \geq k(\alpha)\sqrt{w^T \Sigma_- w}.$$

Problem (6) can be formulated as a second-order cone programming problem as the following:

$$\min_{w} \quad ||\Sigma_+^{\frac{1}{2}} w||_2 + ||\Sigma_-^{\frac{1}{2}} w||_2$$
$$s.t. \quad w^T(\mu_+ - \mu_-) = 1,\tag{7}$$

which can be solved by the interior point method. With the optimal value of w^*, the optimal value of b^* can be obtained by:

$$b^* = (w^*)^T \mu_- + \frac{||\Sigma_-^{\frac{1}{2}} w^*||_2}{||\Sigma_-^{\frac{1}{2}} w^*||_2 + ||\Sigma_+^{\frac{1}{2}} w^*||_2}$$
$$= (w^*)^T \mu_+ - \frac{||\Sigma_+^{\frac{1}{2}} w^*||_2}{||\Sigma_+^{\frac{1}{2}} w^*||_2 + ||\Sigma_-^{\frac{1}{2}} w^*||_2}.\tag{8}$$

A new data point $x \in \mathbb{R}^d$ can be assigned to positive class or negative class by calculating:

$$f(x) = sign\left((w^*)^T x + b^*\right).\tag{9}$$

For more details, please refer to [21].

2.2. Soft Quadratic Surface Support Vector Machine (SQSSVM)

The SQSSVM model intends to find the optimal parameter set (W, b, c) of a quadratic surface:

$$\frac{1}{2}x^T W x + b^T x + c = 0,$$

where $W = W^T \in \mathbb{R}^{d \times d}, b \in \mathbb{R}^d$, and $c \in \mathbb{R}$, which separates the training points into two classes by maximizing the distance from each data point to the separating surface. The SQSSVM model is as follows:

$$\min_{W,b,c} \quad \sum_{i=1}^{l} ||W x_i + b||_2^2 + C \sum_{i=1}^{l} \xi_i$$
$$s.t. \quad y_i\left(\frac{1}{2}x_i^T W x_i + b^T x_i + c\right) \geq 1 - \xi_i,, i = 1,\ldots,l,\tag{10}$$
$$\xi_i \geq 0, i = 1,\ldots,l,$$

where $C > 0$ is the penalty parameter.

The matrix $W = (\omega_{ij})$ can be vectorized to

$$\bar{W} = [\omega_{11}\ \omega_{12}\ \ldots\ \omega_{1d}\ \omega_{22}\ \omega_{23}\ \ldots\ \omega_{2d}\ \ldots\ \omega_{dd}]^T \in \mathbb{R}^{\frac{d^2+d}{2}}. \tag{11}$$

Denote $z = (\bar{W}^T, b^T)^T$. Problem (10) can be reformulated as follows:

$$\begin{aligned}
\min_{z,c,\xi}\quad & z^T \hat{G} z + C \sum_{i=1}^{l} \xi_i \\
s.t.\quad & y_i(s_i^T z + c) \geq 1 - \xi_i,\ i = 1, \ldots, l, \\
& \xi_i \geq 0,\ i = 1, \ldots, l.
\end{aligned} \tag{12}$$

In Problem (12), $\hat{G} = \sum_{i=1}^{n}([M^i, I])^T[M^i, I] \in \mathbb{R}^{(\frac{d^2+3d}{2}) \times (\frac{d^2+3d}{2})}$, where I is the m-dimensional identity matrix. $M^i \in \mathbb{R}^{d \times (\frac{d^2+d}{2})}$ is constructed by using the training point $x_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T \in \mathbb{R}^d$ as follows. For the j-th row of M^i in $\mathbb{R}^{\frac{d^2+d}{2}}$, $j = 1, \ldots, d$, check the elements of $\bar{W}$ one by one. If the p-th element of $\bar{W}$ is ω_{jk} or ω_{kj} for some $k = 1, 2, \ldots, d$, then assign the p-th element of the j-th row of M^i to be x_{ik}. Otherwise, assign it to be 0. Besides,

$$s_i = [\frac{1}{2}x_{i1}x_{i1}, \ldots, x_{i1}x_{id}, \frac{1}{2}x_{i2}x_{i2}, \ldots, x_{i2}x_{id}, \ldots, \frac{1}{2}x_{i,d-1}x_{i,d-1}, x_{i,d-1}x_{id}, \tag{13}$$

$$\frac{1}{2}x_{id}x_{id}, x_{i1}, x_{i2}, \ldots, x_{id}] \in \mathbb{R}^{\frac{d^2+d}{2}+d}. \tag{14}$$

Problem (12) is a convex quadratic programming problem. With the optimal value of z^* and c^*, the parameter set of (W^*, b^*, c^*) can then be obtained. A new data point $x \in \mathbb{R}^n$ can be assigned to positive class or negative class by calculating

$$f(x) = sign\left(\frac{1}{2}x^T W^* x + (b^*)^T x + c^*\right). \tag{15}$$

For more details, please refer to [13].

2.3. F_1 Value

In this subsection, some performance measures are introduced. The true positive rate (TPR) and true negative rate (TNR) can be defined as follows.

$$\begin{aligned}
TPR &= 1 - FNR = P(\hat{y} = 1 \mid y = 1), \\
TNR &= 1 - FPR = P(\hat{y} = -1 \mid y = -1),
\end{aligned} \tag{16}$$

where y is the true value of the label of the data, $\hat{y}$ is the predicted value of the label of the data, FNR is the false negative rate, and FPR is the false positive rate. Then the precision rate can be defined as:

$$Precision = P(y = 1 \mid \hat{y} = 1) = \frac{p \cdot TPR}{p \cdot TPR + (1-p) \cdot FPR}, \tag{17}$$

where p denotes the proportion of positive class samples.

TPR is more concerned with the comprehensiveness for the recognition of positive class samples, while $Precision$ is more concerned with the accuracy rate of the recognition of positive class samples. Since the TPR and $Precision$ are two contradictory performance measures, some combinations are proposed. The F_1 value is the harmonic mean of TPR and $Precision$, i.e.:

$$F_1 = \frac{2 \cdot TPR \cdot Precision}{Precision + TPR}. \tag{18}$$

Since the classification of positive class samples has a great influence on the F_1 value, so the F_1 value is commonly used to evaluate the classification effect of imbalanced datasets. For more details, please refer to [7].

3. A Quadratic Surface Minimax Probability Machine with a Mixed Performance Measure

In this section, a quadratic surface minimax probability machine with a mixed performance measure (QSMPMFA) is proposed. The mixed performance measure considers both the F_1 value and the weighted accuracy rate. The F_1 measure is adopted for paying more attention on the class with less points. The weighted accuracy rate is adopted for balancing the importance between two classes. For better understanding, the proposed QSMPMFA is first introduced. After this, the alternating descend method for solving this problem is constructed.

3.1. The Proposed Model

Suppose the positive class sample $x_+ \sim (\mu_+, \Sigma_+)$ and the negative class sample $x_- \sim (\mu_-, \Sigma_-)$, $\mu_+, \mu_- \in \mathbb{R}^d$, $\Sigma_+, \Sigma_- \in \mathbb{S}_+^d$. Assume that the mean and covariance matrix of the positive and negative samples are all reliable.

For obtaining a robust classifier for imbalanced datasets, the improved minimax probability machine based on the soft quadratic surface support vector machine is proposed first as follows.

$$\min_{W,b,c,\alpha_1,\alpha_2} \quad \frac{1}{F_1} + C((1-p)\alpha_1 + p\alpha_2)$$

$$s.t. \quad \inf_{x_+ \sim (\mu_+, \Sigma_+)} Pr\{\frac{1}{2}x_+^T W x_+ + b^T x_+ - c \geq 0\} \geq 1 - \alpha_1, \tag{19}$$

$$\inf_{x_- \sim (\mu_-, \Sigma_-)} Pr\{\frac{1}{2}x_-^T A x_- + b^T x_- - c \leq 0\} \geq 1 - \alpha_2,$$

where α_1 denotes the classification error rate of the positive class, α_2 denotes the classification error rate of negative class, C is the penalty parameter, and p denotes the proportion of positive class points. It can be seen that α_1 is the false negative rate (FNR), while α_2 is the false positive rate (FPR) mentioned in Section 2.3. The first constraint in Problem (19) guaranteed that the infimum of the probability that the positive class sample x_+ is above the separating quadratic surface was not less than $1 - \alpha_1$. The second constraint in Problem (19) guaranteed that the infimum of the probability that the positive class sample x_- behind the separating quadratic surface was not less than $1 - \alpha_2$.

Since $\alpha_1 = FNR$ and $\alpha_2 = FPR$, the first item in the objective function can be represented as follows:

$$\frac{1}{F_1} = \frac{(1-p)\alpha_2 + p}{2p(1-\alpha_1)} + \frac{1}{2}. \tag{20}$$

By omitting the constant, Problem (19) can be transformed to the following problem:

$$\min_{W,b,c,\alpha_1,\alpha_2} \quad \frac{(1-p)\alpha_2 + p}{2p(1-\alpha_1)} + C((1-p)\alpha_1 + p\alpha_2)$$

$$s.t. \quad \inf_{x_+ \sim (\mu_+, \Sigma_+)} Pr\{\frac{1}{2}x_+^T W x_+ + b^T x_+ - c \geq 0\} \geq 1 - \alpha_1, \tag{21}$$

$$\inf_{x_- \sim (\mu_-, \Sigma_-)} Pr\{\frac{1}{2}x_-^T A x_- + b^T x_- - c \leq 0\} \geq 1 - \alpha_2.$$

Use the definition of $\bar{W}$ in formulation (11) and denote $z = (\bar{W}^T, b^T)^T$. Problem (21) can be reformulated as follows:

$$\min_{z,c,\alpha_1,\alpha_2} \quad \frac{(1-p)\alpha_2 + p}{2p(1-\alpha_1)} + C((1-p)\alpha_1 + p\alpha_2)$$

$$s.t. \quad \inf_{s_+ \sim (\mu_p, \Sigma_p)} Pr\{z^T s_+ - c \geq 0\} \geq 1 - \alpha_1, \tag{22}$$

$$\inf_{s_- \sim (\mu_n, \Sigma_n)} Pr\{z^T s_- - c \leq 0\} \geq 1 - \alpha_2,$$

where

$$s_+ = [\frac{1}{2}x_{+1}x_{+1}, \ldots, x_{+1}x_{+d}, \frac{1}{2}x_{+2}x_{+2}, \ldots, x_{+2}x_{+d}, \ldots, \frac{1}{2}x_{+,d-1}x_{+,d-1},$$

$$x_{+,d-1}x_{+d}, \frac{1}{2}x_{+d}x_{+d}, x_{+1}, x_{+2}, \ldots, x_{+d}] \in \mathbb{R}^{\frac{d^2+d}{2}+d}, \tag{23}$$

and

$$s_- = [\frac{1}{2}x_{-1}x_{-1}, \ldots, x_{-1}x_{-d}, \frac{1}{2}x_{-2}x_{-2}, \ldots, x_{-2}x_{-d}, \ldots, \frac{1}{2}x_{-,d-1}x_{-,d-1},$$

$$x_{-,d-1}x_{-d}, \frac{1}{2}x_{-d}x_{-d}, x_{-1}, x_{-2}, \ldots, x_{-d}] \in \mathbb{R}^{\frac{d^2+d}{2}+d}. \tag{24}$$

μ_p represents the mean of the dataset of s_+, Σ_p represents the covariance matrix of the dataset of s_+, μ_n represents the mean of the dataset of s_-, and Σ_n represents the covariance matrix of the dataset of s_-. The separation obtained in Problem (22) can be interpreted as follows. Each data point $x \in \mathbb{R}^d$ in the original dataset of T is first mapped into a new space, i.e., $s \in \mathbb{R}^{\frac{d^2+d}{2}+d}$. Next, a linear classifier is adopted for the separation.

Since the mean and covariance matrix are often unknown, $\hat{\mu}$ and $\hat{\Sigma}$ are used to estimate of the true value, i.e.:

$$\hat{\mu}_{p(n)} = \frac{1}{N_{+(-)}} \sum_{i=1}^{N_{+(-)}} s_i \in \mathbb{R}^{\frac{d^2+3d}{2}},$$

$$\hat{\Sigma}_{p(n)} = \frac{1}{N_{+(-)}} \sum_{i=1}^{N_{+(-)}} (s_i - \hat{\mu}_{p(n)})(s_i - \hat{\mu}_{p(n)})^T \in \mathbb{R}^{\frac{d^2+3d}{2}}, \tag{25}$$

where N_+ represents the number of positive points in the dataset and N_- represents the number of negative points in the data set.

By using Lemma 2, the probability inequalities in Problem (22) can be equivalently replaced by two inequalities. Then Problem (22) can be formulated as follows:

$$\min_{z,c,\alpha_1,\alpha_2} \quad \frac{(1-p)\alpha_2 + p}{2p(1-\alpha_1)} + C((1-p)\alpha_1 + p\alpha_2)$$

$$s.t. \quad z^T \mu_p - c \geq \pi(\alpha_1)\sqrt{z^T \Sigma_p z}, \tag{26}$$

$$c - z^T \mu_n \geq \pi(\alpha_2)\sqrt{z^T \Sigma_n z},$$

where $\pi(\alpha_1) = \sqrt{\frac{1-\alpha_1}{\alpha_1}}$, $\pi(\alpha_2) = \sqrt{\frac{1-\alpha_2}{\alpha_2}}$.

Since the objective function is an increasing function with respect to α_1 or α_2, Problem (26) can be further formulated as the following problem:

$$\min_{z,c,\alpha_1,\alpha_2} \quad \frac{(1-p)\alpha_2 + p}{2p(1-\alpha_1)} + C((1-p)\alpha_1 + p\alpha_2)$$

$$s.t. \quad z^T \mu_p - c = \pi(\alpha_1)\sqrt{z^T \Sigma_p z}, \tag{27}$$

$$c - w^T \mu_n = \pi(\alpha_2)\sqrt{z^T \Sigma_n z}.$$

The equivalence of Problem (26) and Problem (27) can be derived similarly to Lemma 4 in [7]. By summarizing the two constraints, the variable c can be omitted, i.e.:

$$\min_{z,\alpha_1,\alpha_2} \quad \frac{(1-p)\alpha_2 + p}{2p(1-\alpha_1)} + C((1-p)\alpha_1 + p\alpha_2) \tag{28}$$
$$\text{s.t.} \quad \pi(\alpha_1)\sqrt{z^T\Sigma_p z} + \pi(\alpha_2)\sqrt{z^T\Sigma_n z} = z^T(\mu_p - \mu_n).$$

Furthermore, as shown in (28), α_1 and α_2 play the symmetry role in the equality constraint. Then, α_2 can be represented by α_1 and z as follows.

$$\alpha_2 = \frac{\sqrt{z^T\Sigma_n z}}{\sqrt{z^T\Sigma_n z} + (z^T(\mu_p - \mu_n) - \pi(\alpha_1)z^T\Sigma_p z)^2}. \tag{29}$$

Finally, by using this formulation, Problem (28) can be transformed to the following problem with respect to z and α_1.

$$\min_{z,\alpha_1} \quad M(z,\alpha_1) = \frac{\left(\frac{1-p}{2p(1-\alpha_1)} + Cp\right)\sqrt{z^T\Sigma_n z}}{\sqrt{z^T\Sigma_n z} + (z^T(\mu_p - \mu_n) - \pi(\alpha_1)z^T\Sigma_p z)^2}$$
$$+ \frac{1}{2(1-\alpha_1)} + C(1-p)\alpha_1 \tag{30}$$
$$\text{s.t.} \quad z^T(\mu_p - \mu_n) - \pi(\alpha_1)z^T\Sigma_p z \geq 0,$$
$$\|z\| = 1.$$

The constraint $\|z\| = 1$ is added to ensure the uniqueness of the solution. Problem (30) is a nonconvex optimization problem with a fractional objective function.

3.2. Algorithm for QSMPMFA

For efficiently solving this problem, an alternating descent algorithm [7] is adopted. The details of the algorithm are described as follows.

In the t-th iteration, suppose that the value of z_t satisfies the constraint of $\|z_t\| = 1$, then α_1^t can be updated by solving the following problem:

$$\alpha_1^t = \underset{\alpha_1}{argmin} \quad \frac{\left(\frac{1-p}{2p(1-\alpha_1)} + Cp\right)\sqrt{z^T\Sigma_n z}}{\sqrt{z^T\Sigma_n z} + (z^T(\mu_p - \mu_n) - \pi(\alpha_1)z^T\Sigma_p z)^2} + \frac{1}{2(1-\alpha_1)}$$
$$+ C(1-p)\alpha_1 \tag{31}$$
$$\text{s.t.} \ \alpha_1 \geq z_t^T\Sigma_p z_t/(z_t^T\Sigma_p z_t + (\mu_p - \mu_n)^2).$$

Since this problem is a univariate optimization problem, it can be solved by the line search method.

Now, z is updated by solving Problem (30) with $\alpha_1 = \alpha_1^t$, i.e.:

$$\max_z \quad \lambda_t(z) = \frac{z^T(\mu_p - \mu_n) - \pi(\alpha_1^t)\sqrt{z^T\Sigma_p z}}{\sqrt{z^T\Sigma_n z}}. \tag{32}$$
$$\text{s.t.} \quad \|z\| = 1.$$

To find a $\mathbf{z}_{t+1}$ that satisfies $\lambda_t(z_{t+1}) > \lambda_t(z_t)$, define the following function:

$$f_t(z) = z^T(\mu_p - \mu_n) - \pi(\alpha_1^t)\sqrt{z^T\Sigma_p z} - \lambda_t(z_t)\sqrt{z^T\Sigma_n z}. \tag{33}$$

The $f_t(z_t) = 0$. A z_{t+1} that satisfies $f_t(z_{t+1}) > 0$ should be found. The following nonlinear programming problem can be solved for obtaining the inexact solution of Problem (32).

$$\max_z \quad f_t(z) \tag{34}$$
$$\text{s.t.} \quad \|z\| = 1.$$

The gradient ascent method is used to solve Problem (34). Let $v_1 = z_t$, step $\gamma_k = 1/k, k = 1, 2, \ldots$, and, at the k-th step, v_k is updated according to the following equation:

$$u_k = v_k + \gamma_k\left((\boldsymbol{\mu}_p - \boldsymbol{\mu}_n) - \frac{\pi(\alpha_1^t)\Sigma_p v_k}{\sqrt{v_k^T \Sigma_p v_k}} - \frac{\lambda_t(z_t)\Sigma_n v_k}{\sqrt{v_k^T \Sigma_n v_k}}\right), \quad v_{k+1} = \frac{u_k}{\|u_k\|}. \tag{35}$$

Set the stopping condition of the above iterative process as:

$$f_t(v_{k+1}) \geq 0, \quad f_t(v_k) \geq 0, \quad \frac{f_t(v_{k+1}) - f_t(v_k)}{f_t(v_{k+1})} \leq \epsilon. \tag{36}$$

The ϵ in the above equation is the stopping error. The v^* that satisfied conditions (36) can be set to the updated z_{t+1}.

After t_0 rounds of the alternating descent algorithm, the value of z_{t_0} and b_{t_0} can be obtained. Correspondingly, c^* in the quadratic surface parameters are represented as:

$$c^* = z_{t_0}^T \boldsymbol{\mu}_p - \pi(\alpha_1^T)\sqrt{z_{t_0}^T \Sigma_p z_{t_0}}. \tag{37}$$

Furthermore, the parameters W^* and b^* can be determined. For a new sample x, the label can be further determined according to the following decision function:

$$f(x) = sign\left(\frac{1}{2}x^T W^* x + (b^*)^T x + c^*\right), \tag{38}$$

where $sign(\cdot)$ is a sign function.

The algorithm has been summarized in Algorithm 1. The convergence of this algorithm can be illustrated by using the analyze in [7].

Algorithm 1: Training process of QSMPMFA

Input: Training set $T = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{m_1+m_2}, y_{m_1+m_2})\}$, C, $t = 0$.
Output: W^*, b^*, c^*.
1. Transform the training set T to $\tilde{T} = \{(s_1, y_1), (s_2, y_2), \ldots, (s_{m_++m_-}, y_{m_++m_-})\}$ by using the formulations (23) and (24).
2. Calculate the mean and covariance matrices $(\hat{\boldsymbol{\mu}}_p, \hat{\Sigma}_p)$, $(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_n)$ according to formula (25).
3. Denote $z_t = (\hat{\boldsymbol{\mu}}_p - \hat{\boldsymbol{\mu}}_n)/\|\hat{\boldsymbol{\mu}}_p - \hat{\boldsymbol{\mu}}_n\|$.
4. **for** t =1, 2, $\cdots$ **do**
5.   Calculate α_1^t by solving Problem (31).
6.   Calculate the objective function of Problem (30), i.e., $M(z_t, \alpha_1^t)$.
7.   Let $v_k = z_t$, $k = 1$.
8.   **While** the condition (36) is not met
9.   update v_{k+1} by using formula (35).
10.   **End**
11.   Let $z_{t+1} = v_{k+1}$, calculate the objective function of Problem (30), i.e., $M(z_{t+1}, \alpha_1^t)$.
12.   **If** $(M(z_t, \alpha_1^t) - M(z_{t+1}, \alpha_1^t))/M(z_t, \alpha_1^t) \leq \epsilon$
13.   terminate
14.   **End**
15. **End**
16. Calculate W^*, b^*, c^* according to Equations (11) and (37).

---

4. Numerical Tests

In this section, the classification performance of the QSMPMFA model is investigated. For comparison, a minimax probability machine with a linear kernel (MPM-l), polynomial

kernel (MPM-p), and Gaussian kernel (MPM-g) [1]; a minimax probability machine for F measure with a linear kernel (MPMF-l), polynomial kernel (MPMF-p), and Gaussian kernel (MPMF-g) [7]; and a quadratic surface minimax probability machine (QSMPM) [22] are tested.

All the experiments are implemented in MATLAB R2018a on a PC with 2.1 GHz CPU and 16 GB RAM. Both the F_1 value of the minority class samples and accuracy rate are considered for measuring the classification performance. The F_1 value is defined in Section 2.3. The accuracy rate is defined as the ratio of correctly classified numbers to the total number. The penalty parameter C is chosen in the range of $\{2^{-10}, 2^{-5}, 2^0, 2^5\}$. The kernel function parameter of the kernel-based methods is chosen in the range of $\{2^{-5}, 2^{-6}, \ldots, 2^8\}$. They are all trained by using a 5-fold cross-validation [23]. All the methods are tested ten times by randomly selecting points in two classes to form training sets. For each dataset, a random selection of 70% of the samples in the dataset is used for training and the remaining 30% are used for testing.

In order to illustrate the classification performance of the proposed QSMPMFA, comparative experiments on 12 benchmark datasets extracted from the UCI Machine Learning Repository [24] are conducted. The information of all the datasets is shown in Table 1. In this table, the number of positive samples, negative samples, total samples, and features are listed. The last column in Table 1 records the imbalanced ratio, which is calculated by the ratio of the negative data numbers and positive data numbers. The larger the imbalanced ratio, the higher the imbalanced degree.

Table 1. Information of the benchmark datasets.

Data Set	Positive #	Negative #	Total #	Feature #	Imbalanced Ratio
Iris	50	50	100	4	1
Seeds	70	70	140	7	1
Glass	51	163	214	9	3.20
Wholesale customers	142	298	440	7	2.10
WDBC	212	357	569	30	1.68
Balance	288	288	576	4	1
Breast cancer	241	458	699	9	1.90
Banknote	610	762	1372	4	1.25
Segment	330	1980	2310	19	6
Rice	1630	2180	3810	7	1.34
Dry Bean	1322	12289	13,611	16	9.30
Skin	50,859	194,198	245,057	3	3.82

Table 2 shows the F_1 value of each method on all the datasets. The bold ones are the best results among all the methods. "-" indicates that the memory exceeds the limit in the training process. From Table 2, it can be seen that QSMPMFA outperforms the comparison methods for classifying most datasets, especially on datasets with higher imbalance ratios such as Segment, Dry Bean, Skin, and so on. For the other datasets that QSMPMFA performs not optimally on, it can be seen that the best results often occur in the methods with Gaussian kernel functions. The reason may be that the Gaussian kernel is more proper for dealing with these datasets. For showing the results more intuitively, the boxplot of the F_1 value of the positive class of each method for classifying all the datasets is illustrated in Figure 1.

For comparison, the classification accuracy rate of each method is also listed in Table 3. Except for the WDBC and Segment datasets, the performance of QSMPMFA is similar to that in Table 2. For the WDBC dataset, QSMPMFA performs almost the same as the best method. The reasons for explaining these results may be that the F_1 value is adopted in the cross-validation of choosing parameters. It can be seen that the accuracy rates are often larger than the F_1 value on most of the datasets. The boxplot of the accuracy rates of all the methods on each dataset are illustrated in Figure 2.

Table 2. F_1 value for each method on each data set.

Data Set	MPM-l	MPM-g	MPM-p	MPMF-l	MPMF-g	PMPF-p	QSMPM	QSMPMFA
Iris	**1.0000± 0.0000**	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	0.9968 ± 0.0102	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**
Seeds	0.9530 ± 0.0309	**0.9907 ± 0.0120**	0.9887 ± 0.0130	0.9839 ± 0.0188	0.9643 ± 0.0454	0.9497 ± 0.0291	0.9699 ± 0.295	0.9862 ± 0.0160
Glass	0.7728 ± 0.0702	0.8456 ± 0.0472	0.8722 ± 0.0558	0.8158 ± 0.0452	**0.8795 ± 0.0478**	0.8604 ± 0.0504	0.8259 ± 0.0463	0.8534 ± 0.0564
Wholesale customers	0.8158 ± 0.0413	0.8452 ± 0.0421	0.8425 ± 0.0164	0.8017 ± 0.1277	0.8451 ± 0.0430	0.7256 ± 0.1361	0.8608 ± 0.0327	**0.8660 ± 0.0379**
WDBC	0.9357 ± 0.0172	0.9495 ± 0.0192	0.9524 ± 0.0171	0.9495 ± 0.0195	0.9548 ± 0.0194	0.9519 ± 0.0108	0.9598 ± 0.0162	**0.9603 ± 0.0116**
Balance	0.9448 ± 0.0147	**0.9901 ± 0.0087**	0.9849 ± 0.0126	0.9605 ± 0.0105	0.9885 ± 0.0066	0.9596 ± 0.0099	0.9878 ± 0.0080	0.9818 ± 0.0105
Breast cancer	0.9427 ± 0.0158	0.9558 ± 0.0156	0.9503 ± 0.0210	0.9500 ± 0.0200	0.9556 ± 0.0142	0.9471 ± 0.0294	0.9527 ± 0.0158	**0.9567 ± 0.0117**
Banknote	0.8619 ± 0.0253	0.9895 ± 0.0042	0.9820 ± 0.0050	0.9850 ± 0.0060	0.9908 ± 0.0071	0.9823 ± 0.0069	0.9773 ± 0.0070	**0.9957 ± 0.0057**
Segment	0.5791 ± 0.0164	0.6265 ± 0.0714	0.7106 ± 0.0521	0.9660 ± 0.0120	0.9771 ± 0.0312	0.9110 ± 0.0359	0.9302 ± 0.0209	**0.9814 ± 0.0153**
Rice	0.9047 ± 0.0094	0.9142 ± 0.0237	0.9073 ± 0.0313	0.9098 ± 0.0074	0.9063 ± 0.0117	0.9131 ± 0.0078	0.9135 ± 0.0074	**0.9142 ± 0.0080**
Dry Bean	0.5263 ± 0.0121	0.8361 ± 0.0436	0.8117 ± 0.0393	0.8790 ± 0.0109	0.8639 ± 0.0078	0.8617 ± 0.0082	0.8353 ± 0.0135	**0.8867 ± 0.0097**
Skin	0.8098 ± 0.0017	-	-	0.8793 ± 0.0013	-	-	0.8562 ± 0.0012	**0.8824 ± 0.0178**
Average rank	7.08	3.83	4.79	4.79	3.96	5.46	4.08	2.00

Table 3. Accuracy rate for each method on each dataset.

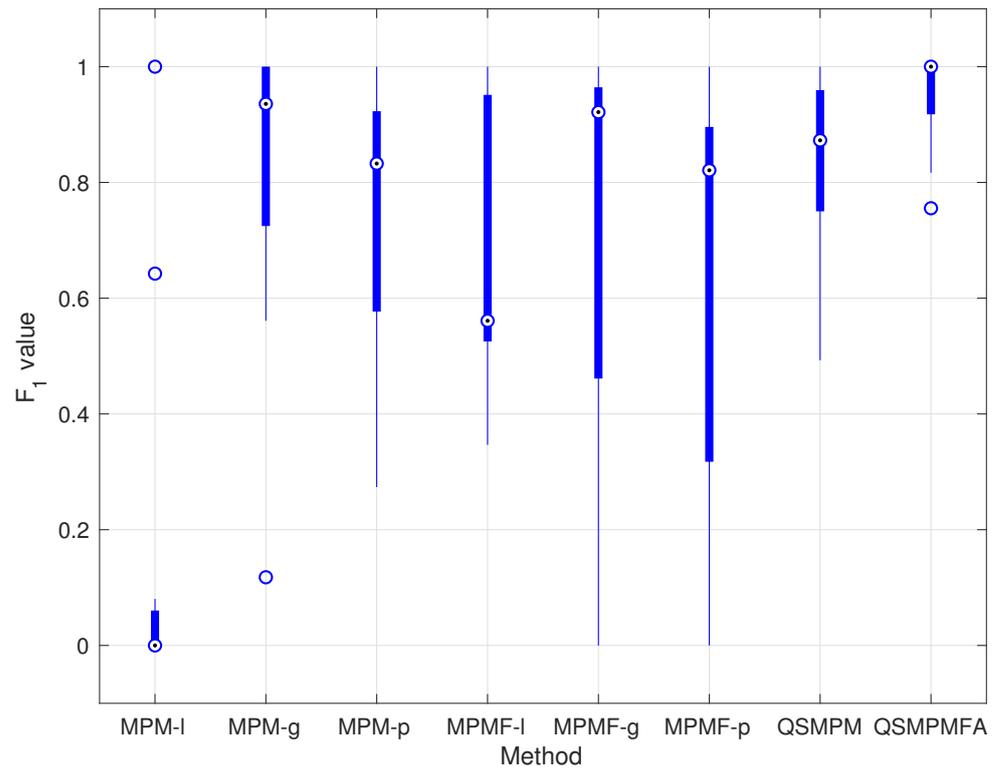| Data Set	MPM-l	MPM-g	MPM-p	MPMF-l	MPMF-g	PMPF-p	QSMPM	QSMPMFA
Iris	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	0.9967 ± 0.0105	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**
Seeds	0.9524 ± 0.0317	**0.9905 ± 0.0123**	0.9893 ± 0.0144	0.9833 ± 0.0196	0.9619 ± 0.0504	0.9476 ± 0.0310	0.9690 ± 0.0298	0.9857 ± 0.0166
Glass	0.8938 ± 0.0328	0.9281 ± 0.0223	0.9406 ± 0.0280	0.9125 ± 0.0235	**0.9406 ± 0.0242**	0.9375 ± 0.0156	0.9141 ± 0.0247	0.9328 ± 0.0256
Wholesale customers	0.8864 ± 0.0250	0.8992 ± 0.0305	0.8909 ± 0.0148	0.8341 ± 0.1572	0.8962 ± 0.0297	0.7561 ± 0.1686	0.9076 ± 0.0220	**0.9099 ± 0.0221**
WDBC	0.9520 ± 0.0132	0.9626 ± 0.0144	**0.9709 ± 0.0137**	0.9632 ± 0.0138	0.9661 ± 0.0148	0.9637 ± 0.0087	0.9696 ± 0.0126	0.9708 ± 0.0087
Balance	0.9453 ± 0.0143	**0.9901 ± 0.0087**	0.9849 ± 0.0127	0.9593 ± 0.0110	0.9884 ± 0.0067	0.9581 ± 0.0104	0.9878 ± 0.0080	0.9814 ± 0.0109
Breast cancer	0.9603 ± 0.0108	0.9689 ± 0.0111	0.9656 ± 0.0141	0.9646 ± 0.0138	0.9689 ± 0.0099	0.9636 ± 0.0199	0.9670 ± 0.0109	**0.9694 ± 0.0082**
Banknote	0.8789 ± 0.0214	0.9914 ± 0.0021	0.9872 ± 0.0042	0.9864 ± 0.0055	0.9917 ± 0.0064	0.9840 ± 0.0063	0.9794 ± 0.0065	**0.9961 ± 0.0051**
Segment	0.7924 ± 0.0144	0.8658 ± 0.0347	0.8831 ± 0.0305	0.9902 ± 0.0036	**0.9953 ± 0.0095**	0.9729 ± 0.0116	0.9788 ± 0.0066	0.9947 ± 0.0044
Rice	0.9182 ± 0.0082	0.9248 ± 0.0127	0.9094 ± 0.0216	0.9190 ± 0.0068	0.9158 ± 0.0113	0.9227 ± 0.0068	0.9241 ± 0.0067	**0.9253 ± 0.0071**
Dry Bean	0.8412 ± 0.0065	0.9490 ± 0.0208	0.9571 ± 0.0202	0.9761 ± 0.0022	0.9771 ± 0.0032	0.9705 ± 0.0047	0.9634 ± 0.0034	**0.9775 ± 0.0022**
Skin	0.9096 ± 0.0008	-	-	0.9430 ± 0.0007	-	-	0.9342 ± 0.0006	**0.9448 ± 0.0089**
Average rank	6.92	4.08	4.46	4.83	3.83	5.63	4.08	2.17

Figure 1. Box plot of the F_1 value of each method on all the datasets.
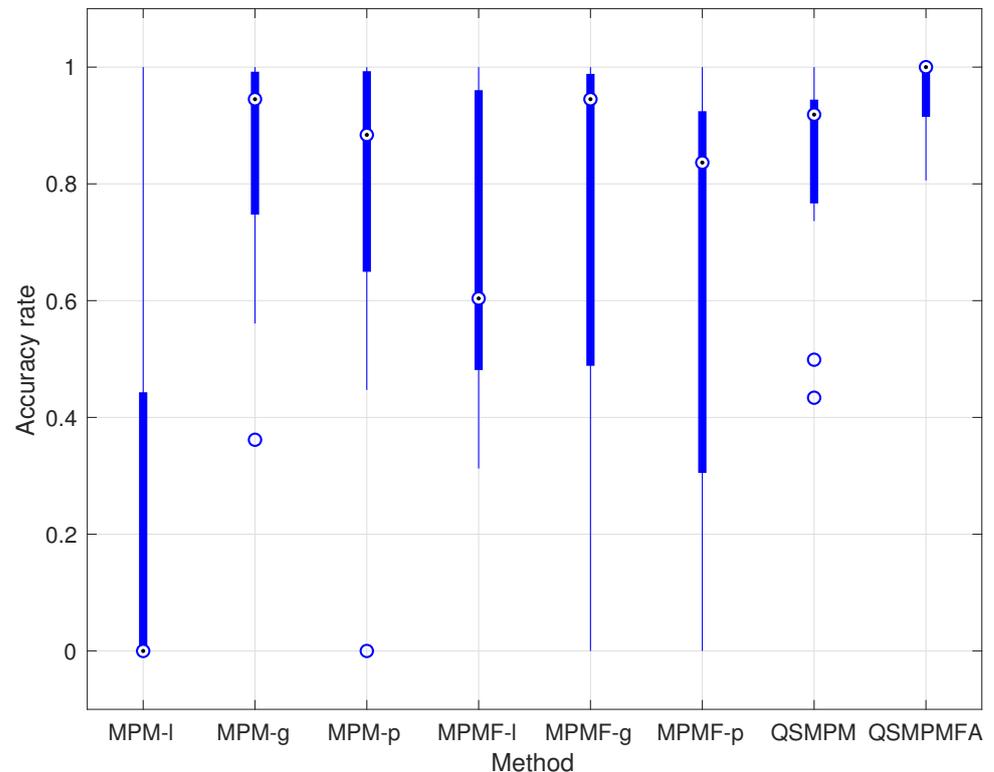

Figure 2. Box plot of the accuracy rate of each method on all the datasets.

From Tables 2 and 3, it can be seen that QSMPMFA performs better than QSMPM for classifying each dataset except for the Balance dataset. QSMPM only aims to maximize the lower bound of the correct classification probability of two classes. Since both QSMPM and QSMPMFA are kernel-free methods, the better performance of the proposed method

demonstrates the effectiveness of setting both the F_1 value and accuracy rate as objective functions. MPMF with different kernel functions adopts F_β as the optimization goal, which focuses on the classification of imbalanced datasets. When compared with MPMF, the better performance also illustrates the effectiveness of using two measures and the kernel-free setting.

For illustrating the performance of the proposed method on imbalanced datasets, more tests on the wholesale customers dataset are conducted. The wholesale customer dataset refers to the information of clients of a wholesale distributor. It includes the annual spending in monetary units on diverse product categories. Some positive samples are chosen with a different ratio of 1, 0.9, 0.6, and 0.2 randomly. Next, the chosen samples with different ratios and all the negative samples are combined to construct four datasets. A similar training process is utilized for testing these datasets.

The comparison results of MPM-l, MPM-g, MPM-p, MPMF-l, MPMF-g, MPM-p, QSMPM, and QSMPMFA on the four datasets are illustrated in Figures 3 and 4. It can be seen that the proposed QSMPMFA performs the best for classifying all the datasets. As shown in Figure 3, the F_1 values of each method on each dataset show a decreasing trend overall when the ratio of positive points decreases (i.e., the imbalance degree increases). The variation of the F_1 values of the proposed QSMPMFA is relatively stable. The other methods, especially MPM-g, MPM-p, and QSMPM, decrease more significantly when the ratio is small. From Figure 4, it can be seen that the accuracy rate of each method on each dataset increase overall when the ratio of positive points decreases (i.e., the imbalance degree increases). The performance of QSMPMFA is also relatively stable. Although the accuracy rate is not a proper measure for analyzing the imbalanced dataset, it still demonstrates the better performance of the proposed method. When compared with QSMPM, the relatively large growth rate shows the better performance of the proposed method, which further demonstrates the superior performance for classifying imbalanced datasets.
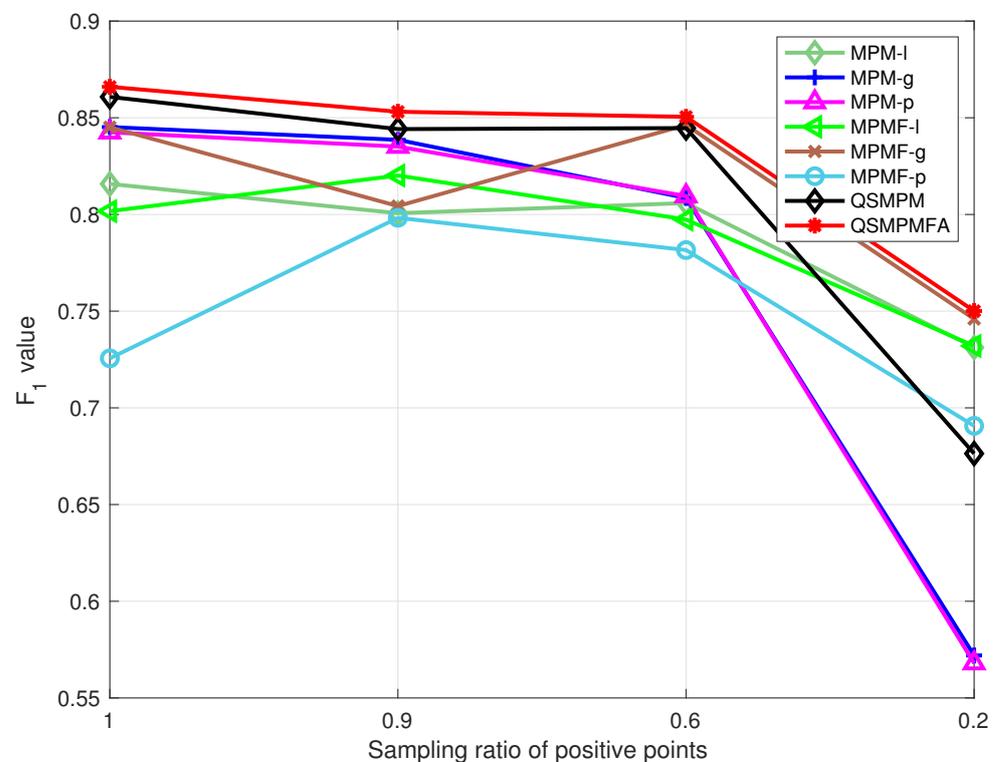

Figure 3. Imbalanced tests for the wholesale customers dataset (F_1 value).
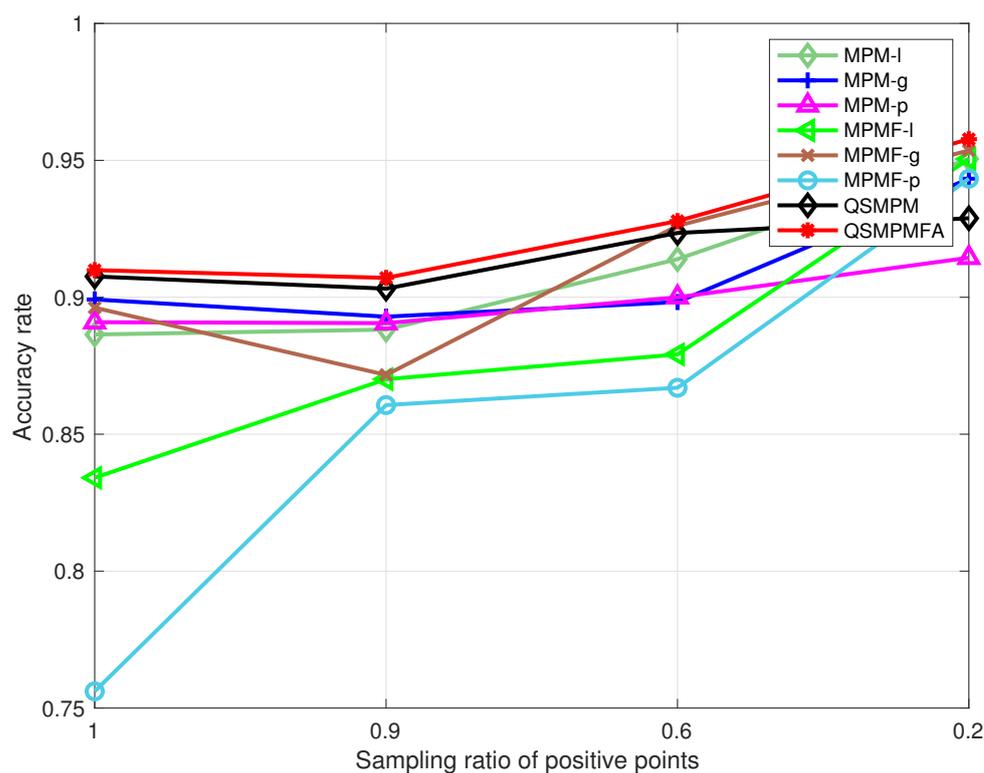
Figure 4. Imbalanced tests for the wholesale customers dataset (accuracy rate).

5. Conclusions

In this paper, a new quadratic surface minimax probability machine model considering both the F_1 value and accuracy rate as the optimization goal is proposed. On the one hand, the proposed QSMPMFA combines the kernel-free technique with the MPM model, which causes it to be more efficient to deal with large-scale datasets. On the other hand, by adopting two performance measures in the objective function of the proposed model, QSMPMFA is more flexible for dealing with the imbalanced dataset. Some public benchmark datasets are adopted for validating the effectiveness of the proposed method. From the numerical tests, it can be seen that QSMPMFA performs the best when compared with other state-of-the-art MPM methods. It can also be concluded that the proposed QSMPMFA enjoys the superiority on dealing with imbalanced datasets.

In the proposed model, both the F_1 value of the minority class and the weighted classification accuracy rate of the two classes are considered as the optimization goal. In our settings, the parameter for balancing the importance between the two items in the objective function is tuned by using cross-validation. The parameter in the weighted classification accuracy rate is fixed by using the ratio of the minority class. Actually, the choice of these parameters affects the classification performance and efficiency of the proposed model. In the future, more research on how to choose proper parameters should be performed. More applications of this method can be explored in other areas [25]. Besides, methods with prior information will be considered.

Author Contributions: Conceptualization, X.Y.; methodology, Z.X.; validation, X.Y.; writing—original draft preparation, Z.X.; writing—review and editing, X.Y., Z.M.; supervision, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All of the benchmark datasets used in our numerical experiments are from the UCI Machine Learning Repository, which are available at http://archive.ics.uci.edu/ml, accessed on 1 January 2022.

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

References

1. Lanckriet, G.R.G.; El Ghaoui, L.; Bhattacharyya, C.; Jordan, M.I. Minimax probability machine. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*; MIT Press: Cambridge, MA, USA, 2001; pp. 801–807.
2. Huang, K.; Yang, H.; King, I.; Lyu, M.R.; Chan, L. Biased minimax probability machine for medical diagnosis. In Proceedings of the 8th International Symposium on Artificial Intelligence & Mathematics, Fort Lauderdale, FL, USA, 4–6 January 2004.
3. Jiang, B.; Guo, Z.; Zhu, Q.; Huang, G. Dynamic minimax probability machine-based approach for fault diagnosis using pairwise discriminate analysis. *IEEE Trans. Control. Syst. Technol.* **2017**, *27*, 806–813. [CrossRef]
4. Maldonado, S.; López, J.; Vairetti, C. Profit-based churn prediction based on minimax probability machines. *Eur. J. Oper. Res.* **2020**, *284*, 273–284. [CrossRef]
5. Huang, K.; Yang, H.; King, I.; Lyu, M.R.; Chan, L. The minimum error minimax probability machine. *J. Mach. Learn. Res.* **2004**, *5*, 1253–1286.
6. Huang, K.; Yang, H.; King, I.; Lyu, M.R. Imbalanced learning with a biased minimax probability machine. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2006**, *36*, 913–923. [CrossRef] [PubMed]
7. Luo, J.; Qiao, H.; Zhang, B. A minimax probability machine for nondecomposable performance measures. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**. [CrossRef] [PubMed]
8. Nath, J.S.; Bhattacharyya, C. Maximum margin classifiers with specified false positive and false negative error rates. In Proceedings of the 2007 SIAM International Conference on Data Mining, Minneapolis, MN, USA, 26–28 April 2007; pp. 35–46.
9. Maldonado, S.; Carrasco, M.; López, J. Regularized minimax probability machine. *Knowl.-Based Syst.* **2019**, *177*, 127–135. [CrossRef]
10. Jayadeva; Khemchandani, R.; Chandra, S. Twin support vector machines for pattern classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 905–910.
11. Yang, L.; Wen, Y.; Zhang, M.; Wang, X. Twin minimax probability machine for pattern classification. *Neural Netw.* **2020**, *131*, 201–214. [CrossRef] [PubMed]
12. Dagher, I. Quadratic kernel-free non-linear support vector machine. *J. Glob. Optim.* **2008**, *4*, 15–30. [CrossRef]
13. Luo, J.; Fang, S.C.; Deng, Z.; Guo, X. Soft quadratic surface support vector machine for binary classification. *Asia-Pac. J. Oper. Res.* **2016**, *33*, 1650046. [CrossRef]
14. Bai, Y.; Han, X.; Chen, T.; Yu, H. Quadratic kernel-free least squares support vector machine for target diseases classification. *J. Comb. Optim.* **2015**, *30*, 850–870. [CrossRef]
15. Yan, X.; Bai, Y.; Fang, S.C.; Luo, J. A kernel-free quadratic surface support vector machine for semi-supervised learning. *J. Oper. Res. Soc.* **2016**, *67*, 1001–1011. [CrossRef]
16. Tian, Y.; Sun, M.; Deng, Z.; Luo, J.; Li, Y. A new fuzzy set and nonkernel SVM approach for mislabeled binary classification with applications. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 1536–1545. [CrossRef]
17. Luo, J.; Yan, X.; Tian, Y. Unsupervised quadratic surface support vector machine with application to credit risk assessment. *Eur. J. Oper. Res.* **2020**, *280*, 1008–1017. [CrossRef]
18. Gao, Z.; Fang, S.C.; Luo, J.; Medhin, N. A kernel-free double well potential support vector machine with applications. *Eur. J. Oper. Res.* **2021**, *290*, 248–262. [CrossRef]
19. Yan, X.; Zhu, H. A novel robust support vector machine classifier with feature mapping. *Knowl.-Based Syst.* **2022**, *257*, 109928. [CrossRef]
20. Wang, Y.; Yang, Z.; Yang, X. Kernel-free quadratic surface minimax probability machine for a binary classification problem. *Symmetry* **2021**, *13*, 1378. [CrossRef]
21. Lanckriet, G.R.; Ghaoui, L.E.; Bhattacharyya, C.; Jordan, M.I. A robust minimax approach to classification. *J. Mach. Learn. Res.* **2002**, *3*, 555–582.
22. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
23. Zhou, Z. *Machine Learning*; Tsinghua University Press: Beijing, China, 2016.
24. Dua, D.; Graff, C. UCI Machine Learning Repository. 2017. Available online: http://archive.ics.uci.edu/ml (accessed on 1 January 2022).
25. Chang, J.; He, L.; Chen, L.; Shen, Z.; Chuah, L.F.; Bokhari, A.; Klemeš, J.J.; Han, N. Numerical simulation of liquid jet atomization in subsonic crossflow. *Energy* **2022**, *257*, 124676.