ABSTRACT
Regularized multiple-criteria linear programming (RMCLP) model is a new powerful method for classification and has been used in various real-life data mining problems. So far, current RMCLP implicitly assumes the training data to be known exactly. However, in practice, there are usually many measurement and statistical errors in the training data. In this paper, we propose a Robust Regularized Multiple-Criteria Linear Programming (called R-RMCLP) via second order cone programming formulations for classification. Preliminary numerical experiments show the robustness of our method.

Categories and Subject Descriptors
H.2.8 [Information Systems]: Database Management—Data mining

General Terms
Theory, Algorithms, Performance

Keywords
RMCLP; second order cone programming; SVM

1. INTRODUCTION
For the last decade, the researchers have extensively developed various optimization techniques to deal with classification problem in data mining or machine learning. Support Vector Machine(SVM) [28, 6] is one of the most popular methods. However, Applying optimization techniques to solve classification has seventy years history. Linear Discriminant Analysis(LDA) [8] was first proposed in 1936. Mangasarian [20] has proposed a large margin classifier based on linear programming in 1960’s. From 1980’s to 1990’s, Glover proposed a number of linear programming models to solve discriminant problems with a small sample size of data [9, 10]. Shi and his colleagues[21] extend Glover’s method into classification via multiple criteria linear programming (MCLP), and then various improved algorithms were proposed one after the other [25, 4, 16, 18, 31, 19, 14, 13]. These mathematical programming approaches to classification have been applied to handle many real world data mining problems, such as credit card portfolio management [26, 24], bioinformatics [30, 27], fraud management [22], information intrusion and detection [15], firm bankruptcy [17], and etc.

For the methods mentioned above, the parameters in the training sets are implicitly assumed to be known exactly. However, in real world applications, the parameters have perturbations since they are estimated from the data subject to measurement and statistical errors[11]. Goldfarb et al. pointed out that the solutions to optimization problems are typically sensitive to parameter perturbations, errors in the input parameters tend to get amplified in the decision function, which often results in misclassification. For an instance, for the fixed examples, original discriminants can correctly separate them(see Fig 1 (a)). When each example is allowed to move in a sphere, original decision function cannot separate samples in the worst case (see Fig 1 (b)). So the our goal is to explore a robust model which can deal with data set with measurement or statistical errors[2, 5](see Fig 1 (c)).

In this paper, we proposed a Robust formulation of Regularized Multiple Criteria Linear Programming(called Robust-RMCLP), which is represented as a second-order cone programming (SOCP). This method can deal with data with measurement noise and obtain a robust decision function, which is an useful extension of RMCLP[25].

The remaining parts of the paper are organized as follows. Section 2 introduces the basic formulation of MCLP and RMCLP; Section 3 describes in detail our proposed Algorithms: Robust-RMCLP; All experiment results are shown in section 4; In the last section the conclusions are given.

2. REGULARIZED MCLP FOR DATA MINING
We give a brief introduction of MCLP in the following,
For classification about the training data
\[ T = \{(x_i, y_i), \ldots, (x_i, y_i)\} \subset (\mathbb{R}^n \times \mathcal{Y})^l, \]
where \( x_i \in \mathbb{R}^n, y_i \in \mathcal{Y} = \{1, -1\}, i = 1, \ldots, l \), data separation can be achieved by two opposite objectives. The first objective separates the observations by minimizing the sum of the deviations (MSD) among the observations. The second maximizes the minimum distances (MMD) of observations from the critical value \([10]\). However, it is difficult to solve classification problem. These make RMCLP have stronger insensitivity to outliers.

A lot of empirical studies have shown that MCLP is a powerful tool for classification. However, we cannot ensure this model always has a solution under different kinds of training samples. To ensure the existence of solution, recently, Shi et al. proposed a RMCLP model by adding two regularized items \( \frac{1}{2} w^T H w + \frac{1}{2} \xi^{(1)}^T Q \xi^{(1)} \) on MCLP as follows (more theoretical explanation of this model can be found in \([25]\)):

\[
\min \quad \frac{1}{2} w^T H w + \frac{1}{2} \xi^{(1)}^T Q \xi^{(1)} + Ce^T \xi^{(1)} - De^T \xi^{(2)},
\]

\[
\text{s.t.} \quad (w \cdot x_i) - (\xi^{(1)}_i - \xi^{(2)}_i) = b, \quad \text{for } \{i|y_i = 1\}, \quad (w \cdot x_i) + (\xi^{(1)}_i - \xi^{(2)}_i) = b, \quad \text{for } \{i|y_i = -1\}\]

\[
\xi^{(1)}_i, \xi^{(2)}_i \geq 0,
\]

where \( C, D > 0 \), and \( e \in \mathbb{R}^l \) be vector whose all elements are 1, \( w \) and \( b \) are unrestricted, \( u_i \) is the overlapping and \( v_i \) the distance from the training sample \( x_i \) to the discriminator \( w \cdot x_i = b \) (classification separating hyperplane).

Figure 1: (a) The original examples and discriminants; (b) The effect of measurement noises; (c) The result of robust model.

### 3. ROBUST REGULARIZED MULTIPLE CRITERIA LINEAR PROGRAMMING (ROBUST-RMCLP)

#### 3.1 Linear Robust-RMCLP

We firstly give the formal representation of robust classification learning problem. Given a training set
\[ T = \{(X_i, y_i), \ldots, (X_i, y_i)\}, \]
where \( y_i \in \mathcal{Y} = \{1, -1\}, i = 1, \ldots, l \), and input set \( X_i \) is a sphere within \( r_i \) radius of the \( x_i \) center:
\[ X_i = \{\bar{x}_i|\bar{x}_i = x_i + r_i u_i, i = 1, \ldots, l, \quad ||u_i|| \leq 1\}, \]
\( \bar{x}_i \) is the true value of the training data, \( u_i \in \mathbb{R}^n, r_i \) is a given constant. The goal is to induce a real-valued function
\[ y = \text{sgn}(g(x)), \]

to infer the label \( y \) corresponding to any example \( x \) in \( \mathbb{R}^n \) space. Generally, Such problem is caused by measurement errors, where \( r_i \) reflects the measurement accuracy.

In order to obtain the optimization decision function of (10), we minimize the maximum misclassification of each example within their corresponding confidence balls. In this case, we choose \( H, Q \) to be identity matrix and add slack variable \( \xi^{(3)} \), (6)~(9) can be written as the following robust optimization problem:

\[
\min \quad \frac{1}{2} ||w||^2 + \frac{1}{2} ||\xi^{(1)}||^2 + C \sum_{i=1}^{l} \xi^{(1)}_i - D \sum_{i=1}^{l} \xi^{(2)}_i,
\]

\[
\text{s.t.} \quad y_i((w \cdot X_i + r_i u_i) - b) = \xi^{(1)}_i - \xi^{(2)}_i, \quad \forall ||u_i|| \leq 1, i = 1, \ldots, l,
\]
\[
\xi^{(1)}_i, \xi^{(2)}_i \geq 0, i = 1, \ldots, l,
\]
where $C, D$ are respectively given constants. 

Since

$$\min \{ y_i r_i (w \cdot u_i), ||u_i|| \leq 1 \} = -r_i ||w||,$$

(16)

problem (13)~(15) can be converted to

$$\min_{w,b,\xi^{(1)}, \xi^{(2)}} \frac{1}{2} ||w||^2 + \frac{1}{2} ||\xi^{(1)}||^2 + C \sum_{i=1}^{t} \xi^{(1)}_i - D \sum_{i=1}^{t} \xi^{(2)}_i,$$

s.t. 

$$y_i ((w \cdot x_i) - b) - r_i ||w|| = \xi^{(1)}_i - \xi^{(2)}_i,$$ 

$$\forall ||u|| \leq 1, i = 1, \ldots, t,$$

$$\xi^{(1)}_i, \xi^{(2)}_i \geq 0, i = 1, \ldots, t.$$ 

(17)

By introducing new variables $t_1, t_2$ and setting $||w|| \leq t_1$, $||\xi^{(1)}|| \leq t_2$. The above problem becomes

$$\min_{w,b,\xi^{(1)}, \xi^{(2)}} \frac{1}{2} t_1^2 + \frac{1}{2} t_2^2 + C \sum_{i=1}^{t} \xi^{(1)}_i - D \sum_{i=1}^{t} \xi^{(2)}_i,$$

s.t. 

$$y_i ((w \cdot x_i) - b) - r_i t_1 = \xi^{(1)}_i - \xi^{(2)}_i,$$ 

$$\forall ||u|| \leq 1, i = 1, \ldots, t,$$

$$\xi^{(1)}_i, \xi^{(2)}_i \geq 0, i = 1, \ldots, t. $$

$$||w|| \leq t_1, ||\xi^{(1)}|| \leq t_2.$$ 

(18)

(20)

(21)

(22)

(23)

(24)

For replacing $t_1^2, t_2^2$ in the objective function (20), we introduce new variables $u_1, u_2, v_1, v_2$ and satisfy the linear constraints $u_i + v_i = 1, i = 1, 2$ and second order cone constraints $\sqrt{t_1^2 + t_2^2} \leq u_i$. Therefore, problem (20)~(24) can be reformulated as the following Second Order Cone Program (SOCP):

$$\min \frac{1}{2} (u_1 - v_1) + \frac{1}{2} (u_2 - v_2) + C \sum_{i=1}^{t} \xi^{(1)}_i - D \sum_{i=1}^{t} \xi^{(2)}_i,$$

s.t. 

$$y_i ((w \cdot x_i) - b) - r_i t_1 = \xi^{(1)}_i - \xi^{(2)}_i,$$ 

$$\forall ||u|| \leq 1, i = 1, \ldots, t,$$

$$\xi^{(1)}_i, \xi^{(2)}_i \geq 0, i = 1, \ldots, t. $$

(25)

(26)

(27)

(28)

(29)

(30)

(31)

(32)

(33)

where $z = (w^T, b, \xi^{(1)}_i, \xi^{(2)}_i, u_1, u_2, v_1, v_2, t_1, t_2)^T$, $\xi^{(1)} = (\xi^{(1)}_1, \ldots, \xi^{(1)}_t)^T$, $\xi^{(2)} = (\xi^{(2)}_1, \ldots, \xi^{(2)}_t)^T$, $f = 1, 2$. The penalty parameters $C, D > 0$.

Now we derive the dual problem of problem (25)~(33).

By introducing its Lagrange function

$$L = \frac{1}{2} (u_1 - v_1) + \frac{1}{2} (u_2 - v_2) + C \sum_{i=1}^{t} \xi^{(1)}_i - D \sum_{i=1}^{t} \xi^{(2)}_i$$

$$- \sum_{i=1}^{t} \alpha_i (y_i ((w \cdot x_i) - b) - r_i t_1 - \xi^{(1)}_i + \xi^{(2)}_i)$$

$$- \sum_{i=1}^{t} \eta^{(1)}_i \xi^{(1)}_i - \sum_{i=1}^{t} \eta^{(2)}_i \xi^{(2)}_i - \beta_1 (u_1 + v_1 - 1)$$

$$- \beta_2 (u_2 + v_2 - 1) - z_{u_1} u_1 - z_{v_1} v_1 - \gamma_1 t_1 - z_{u_2} u_2 - z_{v_2} v_2 - \gamma_2 t_2 - z_{t_1} t_1 - z_{t_2} t_2 - z_{t_1}^{(1)} \xi^{(1)}_1,$$

(34)

where $\alpha, \eta^{(1)}_i, \eta^{(2)}_i \in \mathbb{R}^n$, and $\beta_1, \beta_2, z_{u_1}, z_{v_1}, \gamma_1, z_{u_2}, z_{v_2}, \gamma_2, z_{t_1}, z_{t_2}, z_{t_1}^{(1)} \in \mathbb{R}^n$ are lagrange multipliers. In the following, we seek for the minimum value of $L$ about $w, b, \xi^{(1)}_i, \xi^{(2)}_i, u_1, u_2, v_1, v_2, t_1, t_2$ separately:

$$\nabla_{u_1} L = 0, \nabla_{v_1} L = 0, \nabla_{v_2} L = 0, \nabla_{\xi^{(1)}_1} L = 0, \nabla_{\xi^{(2)}_1} L = 0, \nabla_{\xi^{(2)}_2} L = 0.$$ (35)

From (35)~(36), we get

$$\beta_1 + z_{u_1} = \frac{1}{2}, \beta_2 + z_{u_2} = \frac{1}{2},$$

$$\beta_1 + z_{v_1} = -\frac{1}{2}, \beta_2 + z_{v_2} = -\frac{1}{2},$$

$$\sum_{i=1}^{t} \alpha_i = 0,$$

$$\sum_{i=1}^{t} \alpha_i \eta^{(1)}_i = 0,$$

$$\sum_{i=1}^{t} \alpha_i \eta^{(2)}_i = 0,$$

$$\sum_{i=1}^{t} z_{u_1} u_1 + z_{v_1} v_1 = 0,$$

$$\sum_{i=1}^{t} z_{u_2} u_2 + z_{v_2} v_2 = 0,$$

$$D + \alpha_i \eta^{(2)}_i = 0, i = 1, \ldots, l.$$ (36)

(37)

(38)

(39)

(40)

(41)

(42)

(43)

After deleting variables $z_{u_1}, z_{u_2}, z_{v_1}, z_{v_2}, \xi^{(1)}_1, \eta^{(2)}_2$, the dual problem of (25)~(33) can be reformulated as

$$\max \beta_1 + \beta_2$$

s.t. $\beta_1 + z_{u_1} = \frac{1}{2}, \beta_2 + z_{u_2} = \frac{1}{2},$ (44)

$$\beta_1 + z_{v_1} = -\frac{1}{2}, \beta_2 + z_{v_2} = -\frac{1}{2},$$ (45)

$$\sum_{i=1}^{t} \alpha_i = 0,$$ (46)

$$\sum_{i=1}^{t} \alpha_i \eta^{(1)}_i = 0,$$ (47)

$$\sum_{i=1}^{t} \alpha_i \eta^{(2)}_i = 0,$$ (48)

$$\gamma_1 \leq \sum_{i=1}^{t} z_{u_1} + \sqrt{\sum_{i=1}^{t} (C + \alpha_i - \eta^{(1)}_i)^2},$$ (49)

$$\gamma_2 \leq \sqrt{\sum_{i=1}^{t} (C + \alpha_i - \eta^{(2)}_i)^2},$$ (50)

$$\gamma_1^2 + \gamma_2^2 \leq z_{u_1}^2,$$ (51)

$$\eta^{(1)}_i \geq 0, i = 1, \ldots, l,$$ (52)

where $z = (\beta_1, \beta_2, z_{u_1}, z_{v_1}, \gamma_1, z_{u_2}, z_{v_2}, \gamma_2, \alpha^T, \eta^{(1)}^T, \eta^{(2)}^T)^T$.

**Theorem 3.1.** Suppose that $z^*$ is a solution about the dual problem (44)~(52), where $z^* = (\beta_1^*, \beta_2^*, z_{u_1}^*, z_{v_1}^*, \gamma_1^*, z_{u_2}^*, z_{v_2}^*, \gamma_2^*, \alpha^T, \eta^{(1)}^T, \eta^{(2)}^T)^T$. If there exists $\xi^{(1)}_j = 0$, we will obtain
the solution \((w^*, b^*)\) to the primal problem (25)\(~\sim\) (33):

\[
w^* = \frac{\gamma_1}{(\gamma_1 - \sum_{i=1}^{l} \alpha_i r_i)} \sum_{i=1}^{l} \alpha_i y_i x_i, \tag{53}
\]
\[
b^* = \frac{\gamma_1}{(\gamma_1 - \sum_{i=1}^{l} \alpha_i r_i)} \sum_{i=1}^{l} \alpha_i y_i (x_i \cdot x_j) + y_j r_j \gamma_1. \tag{54}
\]

**Proof.** Introduce the dual problem’s lagrange function

\[
L = -\beta_1 - \beta_2 - t_1 \left( \sum_{i=1}^{l} \alpha_i r_i - \gamma_1 \right) - w^T \left( \sum_{i=1}^{l} y_i x_i \right) -
\]
\[
t_2(-\gamma_1) - \sum_{i=1}^{l} \xi_i^{(1)} (C + \alpha_i - \eta_i^{(1)}) + u_1 (\beta_1 + z_{u_1} - \frac{1}{2} t_2) +
\]
\[
u_2 (\beta_2 + z_{u_2} - \frac{1}{2}) + v_1 (\beta_1 + v_1 + \frac{1}{2}) + v_2 (\beta_2 + v_2 + \frac{1}{2}) +
\]
\[
+ l \sum_{i=1}^{l} \rho_i \eta_i^{(1)} - z_{1_u} - z_{2_u} - \gamma_1 = \gamma_2. \tag{55}
\]

According to the KKT conditions in the infinite-dimensional space [3], we know that there exist lagrange multipliers satisfying:

\[
-\xi_i^{(1)} r_i - y_i (w^* \cdot x_i) - \xi_i^{(1)} r_i^* + b^* y_i = 0, \tag{56}
\]
\[-1 + u_1^* + v_1^* = 0, \quad -1 + u_2^* + v_2^* = 0, \tag{57}
\]
\[
\eta_i^{(1)} \geq 0, \quad \rho_i^* \geq 0, \quad \rho_i^* \eta_i^{(1)} = 0, \quad i = 1, \cdots, l, \tag{63}
\]
\[
\left( \begin{array}{c}
\xi_i^{(1)} \\
u_1^* \\
u_2^* \\
t_1^* \\
t_1^*
\end{array} \right) = \left( \begin{array}{c}
z_i^{(1)} \\
z_i^{(1)} \\
z_i^{(1)} \\
z_i^{(1)} \\
z_i^{(1)}
\end{array} \right), \quad \left( \begin{array}{c}
u_2^* \\
u_2^* \\
u_2^* \\
u_2^* \\
u_2^*
\end{array} \right) = \left( \begin{array}{c}
z_i^{(1)} \\
z_i^{(1)} \\
z_i^{(1)} \\
z_i^{(1)} \\
z_i^{(1)}
\end{array} \right). \tag{59}
\]
\[
\eta_i^{(1)} \geq 0, \quad \rho_i^* \geq 0, \quad \rho_i^* \eta_i^{(1)} = 0, \quad i = 1, \cdots, l, \tag{63}
\]
\[
\left( \begin{array}{c}
t_1^* \\
t_1^*
\end{array} \right) \in L^{n+1}, \quad t_1^* \in \mathcal{R}, \quad w^* \in \mathcal{R}^n, \tag{64}
\]
\[
\left( \begin{array}{c}
t_2^* \\
t_2^*
\end{array} \right) \in L^{n+1}, \quad t_2^* \in \mathcal{R}, \quad \xi^{(1)*} \in \mathcal{R}^n, \tag{67}
\]
\[
\left( \begin{array}{c}
u_1^* \\
u_1^* \\
u_1^* \\
u_1^* \\
u_1^*
\end{array} \right) = \left( \begin{array}{c}
z_{u_1}^{(1)} \\
z_{u_1}^{(1)} \\
z_{u_1}^{(1)} \\
z_{u_1}^{(1)} \\
z_{u_1}^{(1)}
\end{array} \right), \quad \left( \begin{array}{c}
u_2^* \\
u_2^* \\
u_2^* \\
u_2^* \\
u_2^*
\end{array} \right) = \left( \begin{array}{c}
z_{u_2}^{(1)} \\
z_{u_2}^{(1)} \\
z_{u_2}^{(1)} \\
z_{u_2}^{(1)} \\
z_{u_2}^{(1)}
\end{array} \right). \tag{68}
\]
\[
\xi^{(1)*} \geq 0, \quad \rho_i^* \geq 0, \quad \rho_i^* \eta_i^{(1)} = 0, \quad i = 1, \cdots, l, \tag{63}
\]

From above conditions (56)\(~\sim\) (69), we easily get that \((w^*^T, b, \xi^{(1)*^T}, \xi^{(2)*^T}, u_1^*, u_2^*, v_1^*, v_2^*, t_1^*, t_2^*)^T\) is a feasible solution to problem (25)\(~\sim\) (33).

Since

\[
\left( \begin{array}{c}
t_1^* \\
t_1^*
\end{array} \right) \in L^{n+1}, \quad \left( \begin{array}{c}
\sum_{i=1}^{l} \alpha_i r_i - \gamma_1 \\
\sum_{i=1}^{l} \alpha_i y_i x_i
\end{array} \right) \in L^{n+1}, \tag{70}
\]
\[
\left( \begin{array}{c}
t_2^* \\
t_2^*
\end{array} \right) \in L^{n+1}, \quad \left( C + \alpha^* - \eta^{(1)*} \right) \in L^{n+1}, \tag{71}
\]

according to Lemma15 of [1], we can obtain

\[
\left( \sum_{i=1}^{l} \alpha_i r_i - \gamma_1 \right) + \sum_{i=1}^{l} \alpha_i y_i (w^* - x_i) = 0, \tag{72}
\]
\[
\left( \sum_{i=1}^{l} \alpha_i y_i x_i \right) + \left( \sum_{i=1}^{l} \alpha_i r_i - \gamma_1 \right) w^* = 0, \tag{73}
\]
\[
\left( \sum_{i=1}^{l} \alpha_i r_i - \gamma_1 \right) (C + \alpha^* - \eta^{(1)*}) = 0, \tag{74}
\]
\[
\left( \sum_{i=1}^{l} \alpha_i \right) (C + \alpha^* - \eta^{(1)*}) + (-\gamma_2) \xi^{(1)*} = 0. \tag{75}
\]

Furthermore, according to Lemma15 of [1], (68), (69) are equivalent to the following formulas

\[
u_1^* z_{u_1}^* + v_1^* z_{u_1}^*, \quad t_1^* = 0. \tag{76}
\]
\[
u_1^* \left( \frac{z_{u_1}^*}{\gamma_1} \right) + z_{u_1} = 0. \tag{77}
\]
\[
u_1^* \left( \frac{\gamma_1}{\gamma_1 - \sum_{i=1}^{l} \alpha_i r_i} \right) \sum_{i=1}^{l} \alpha_i y_i x_i. \tag{79}
\]

By (56)\(~\sim\) (69), (78) and (79), we can get

\[
\frac{1}{2} (u_1^* - v_1^*) + \frac{1}{2} (u_2^* - v_2^*) + C \sum_{i=1}^{l} \xi^{(1)*} - D \sum_{i=1}^{l} \xi^{(2)*} = \beta_1^* + \beta_2^* \tag{80}
\]

Equation (80) shows that the object function value of primal problem (25)\(~\sim\) (33) about \((w^*^T, b, \xi^{(1)*^T}, \xi^{(2)*^T}, u_1^*, u_2^*, v_1^*, v_2^*, t_1^*, t_2^*)^T\) is equal to the one of dual problem (44)\(~\sim\) (52) about \((\beta_1^*, \beta_2^*, z_{u_1}^*, z_{u_2}^*, \gamma_1^*, \gamma_2^*, \alpha^*, \eta^{(1)*^T})^T\). According to Theorem 4 in [7], we can get \((w^*^T, b, \xi^{(1)*^T}, \xi^{(2)*^T}, u_1^*, u_2^*, v_1^*, v_2^*, t_1^*, t_2^*)^T\) is the solution to the primal problem. On the other hand, according to duality theory, we can obtain that \(w^*\) is the unique solution to primal problem (25)\(~\sim\) (33). Hence equation (79) is proved.

If there exist a training point \(x_j\) classified correctly, then \(\xi_j^{(1)*} = 0\), according to (56), we obtain the corresponding

\[
b^* = \frac{\gamma_1}{(\gamma_1 - \sum_{i=1}^{l} \alpha_i r_i)} \sum_{i=1}^{l} \alpha_i y_i (x_i \cdot x_j) + y_j r_j \gamma_1. \tag{81}
\]

Next we are in a position to establish the following Algorithm 1 based on the theorem above.

### 3.2 Nonlinear Robust-RMCLP

The above discussion is restricted to the linear case. Here, we will analyze nonlinear robust-RMCLP by introducing
Obtain the decision function

\[
\begin{align*}
\text{Output:} & \quad \text{Obtain the decision function} \\
& \quad f(x) = \text{sgn}(\sum_{i=1}^{l} \alpha_i^* y_i (x_i \cdot x) + b^*). \\
& \quad \text{where } x \in \mathcal{H}, \mathcal{H} \text{ is the Hilbert space. So the training set (33) becomes} \\
& \quad T = \{(X_i, y_i), \ldots, (X_l, y_l)\}, \\
& \quad \text{where } X_i = \{\Phi(\hat{x}_i)|\hat{x}_i \text{ is in the sphere of the radius } r \text{ and the center } x_i\}. \\
& \quad \text{So when } ||\hat{x}_i - x|| \leq r, \text{ we have} \\
& \quad ||\Phi(\hat{x}_i) - \Phi(x)||^2 = (\Phi(\hat{x}_i) - \Phi(x_i)) \cdot (\Phi(\hat{x}_i) - \Phi(x_i)) \\
& \quad = K(\hat{x}_i, x_i) - 2K(\hat{x}_i, x_i) + K(x_i, x_i) \\
& \quad = 2 - 2\exp(-||\hat{x}_i - x||^2/2\sigma^2) \\
& \quad \leq r^2, \\
& \quad \text{where} \\
& \quad r_i = \sqrt{2 - 2\exp(-||\hat{x}_i - x||^2/2\sigma^2)}. \\
& \quad \text{Thus } X_i \text{ becomes a sphere of the center } \Phi(x_i) \text{ and the radius } r_i \\
& \quad X_i = \{\hat{x}||\hat{x} - \Phi(x)|| \leq r_i\}. \\
\end{align*}
\]
This leads to the following algorithm.

4. **NUMERICAL EXPERIMENT**

Our algorithm code was wrote in MATLAB 2010. The experiment environment: Intel Core i5 CPU, 2 GB memory. The ScDuMiI\(^1\) software is employed to solve the second cone programming problem related to this paper.

We only modify Algorithm 1 as follows:

1. Introduce to Gauss kernel function
   \[
   K(x, x') = \exp(-||x - x'||^2/2\sigma^2),
   \]
   The inner product \((x_i \cdot x_j)\) and \((x_i \cdot x)\) in Algorithm 1 are replaced by \(K(x_i \cdot x_j)\) and \(K(x_i \cdot x)\);
2. \(r_i\) in the second order cone Programming (25)~(33) becomes \(r_i = \sqrt{2 - 2\exp(-||\hat{x}_i - x||^2/2\sigma^2)}\).

To demonstrate the capabilities of our algorithm, we report results on six data sets from the UCI datasets. They are respectively “Hepatitis”, “BUPA liver”, “Heart-Statlog”, “Heart-c”, “Votes” and “WPBC”. In the experiments, the datasets are normalized to \(-1\) and 1. For simplicity, we set all \(r_i\) in (11) to be a constant \(r\). The noise \(u_i\) is generated randomly from the normal distribution and scaled on the unit sphere. We add many noises for the datasets by \(\hat{x}_i = x_i + ru_i\).

The numerical results are given in Table 1 and Fig.2. The testing accuracies for our method are computed using standard 10-fold cross validation \([6]\). The kernel parameter \(C\) and the RBF kernel parameter \(\sigma\) is selected from the set \(\{2^{i}\} i = -7, \ldots, 7\) \((C, D) \text{ in RMCLP and R-RMCLP models are also selected in the same range}) \text{ by 10-fold cross validation on the tuning set comprising of random 10}\% \text{ of the training data. Once the parameters are selected, the tuning set was returned to the training set to learn the final decision function.}

For comparison, the results corresponding to the original RMCLP and Robust SVM\([12]\) are also listed in these tables.

From the results of Table 1 and Fig.2, we can find that the performance of the Robust-RMCLP and Robust SVM is consistently better than that of the original RMCLP. The accuracy of Robust-RMCLP is comparable with Robust SVM. In addition, with the increase of the noise, the accuracy for the original model is much lower than the Robust model.

5. **CONCLUSION**

In this paper, a new Robust Regularized Multiple Criteria Linear Programming has been proposed. All experiments in datasets with noises show that the performance of the robust model is better than that of the original model. However, since Robust-RMCLP need to solve Second Order Cone Programming, its computing speed is slower than original RMCLP. In the future work, we will develop more efficient robust algorithm for classification.

6. **ACKNOWLEDGMENT**

This work has been partially supported by grants from National Natural Science Foundation of China (NO.70921061, NO.10601064), the CAS/SAFEA International Partnership Program for Creative Research Teams, Major International (Ragional) Joint Research Project(NO.71110107026), the President Fund of GUCA, and the National Technology Support Program 2009BAH14B02.

7. **ADDITIONAL AUTHORS**
Table 1: The percentage of tenfold testing correctness for datasets with noises in the case of rbf kernel

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Ins</th>
<th>Dim</th>
<th>Model</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>Hepatitis</td>
<td>155</td>
<td>19</td>
<td>RMCLP</td>
<td>0.7882</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-SVM</td>
<td>0.7891</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-RMCLP</td>
<td>0.7934</td>
</tr>
<tr>
<td>BUPA liver</td>
<td>345</td>
<td>6</td>
<td>RMCLP</td>
<td>0.6412</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-SVM</td>
<td>0.6510</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-RMCLP</td>
<td>0.6555</td>
</tr>
<tr>
<td>Heart-Statlog</td>
<td>270</td>
<td>14</td>
<td>RMCLP</td>
<td>0.8021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-SVM</td>
<td>0.8123</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-RMCLP</td>
<td>0.8252</td>
</tr>
<tr>
<td>Heart-c</td>
<td>303</td>
<td>14</td>
<td>RMCLP</td>
<td>0.8328</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-SVM</td>
<td>0.8521</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-RMCLP</td>
<td>0.8467</td>
</tr>
<tr>
<td>Votes</td>
<td>435</td>
<td>16</td>
<td>RMCLP</td>
<td>0.9211</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-SVM</td>
<td>0.9321</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-RMCLP</td>
<td>0.9467</td>
</tr>
<tr>
<td>WPBC</td>
<td>198</td>
<td>34</td>
<td>RMCLP</td>
<td>0.8128</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-SVM</td>
<td>0.8278</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R-RMCLP</td>
<td>0.8214</td>
</tr>
</tbody>
</table>

8. REFERENCES

[8] R. A. Fisher. The Use of Multiple Measurements in


