



An uncertain support vector machine based on soft margin method

Qiqi Li¹ · Zhongfeng Qin^{1,2} · Zhe Liu³

Received: 12 December 2021 / Accepted: 6 June 2022 / Published online: 29 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Traditional support vector machines (SVMs) play an important role in the classification of precise data. However, due to various reasons, available data are sometimes imprecise. In this paper, uncertain variables are adopted to describe the imprecise data, and an uncertain support vector machine (USVM) is built for linearly α -nonseparable sets based on soft margin method, where a penalty coefficient is utilized as the trade-off between the maximum margin and the sum of slack variables. Then the equivalent crisp model is derived based on the inverse uncertainty distributions. Numerical experiments are designed to illustrate the application of the soft margin USVM. Finally, metrics, such as accuracy, precision, and recall are used to evaluate the robustness of the proposed model.

Keywords Uncertain variable · Uncertain support vector machine · Linearly α -nonseparable data set · Soft margin method · Uncertainty theory

1 Introduction

As one of the most important classification methods, SVMs have achieved great success in numerous real-world problems (Vapnik 1999; Burges 1998), such as image recognition (VenkateswarLal et al. 2019), disease diagnosis (Wang et al. 2018; Okwuashi and Ndehedehe 2020; Gautam et al. 2021), intrusion detection (Mukkamala et al. 2002; Priyadharsini and Chitra 2021) and so on. Traditionally, SVMs are designed to search a tube with the maximum margin based on precise data. A tube is uniquely determined by a hyperplane and the minimum distance from the data to the hyperplane. Different SVMs are constructed according to the complexity of the classification

problem. The hard margin method was first developed by Vapnik (1995) for linearly separable data sets, in which the data with positive labels and those with negative labels lie in the diverse half-spaces determined by a hyperplane. And the parameters of the hyperplane can be expressed as a linear combination of support vectors (Boser et al. 1992). Nevertheless, the majority of data sets are linearly nonseparable, and the SVM based on hard margin method failed in dealing with these problems. Thus, Cortes and Vapnik (1995) introduced slack variables for misclassification samples such that linearly nonseparable training sets can be classified at the maximum margin target, which is known as soft margin method.

However, when the observations are imprecise or the samples are not large enough, the results in the framework of probability theory are usually not satisfying. In order to handle such cases, Liu (2007) founded uncertainty theory, and then Liu (2009, 2010) further improved and perfected the theory. Following the idea, uncertain statistics are developed to handle the issues with imprecise observations. Yao and Liu (2018) proposed a least square method for regression by characterizing imprecise observations as uncertain variables. Then several other regression models (Hu and Gao 2020; Fang and Hong 2020; Zhang et al. 2020) were further studied, and parameter estimation methods (Liu and Yang 2020; Chen 2020; Li et al 2022) were discussed. In the meantime, confidence interval (Lio and Liu 2018), and hypothesis test (Lio and Liu 2020) for uncertain statistic were also introduced. For

✉ Zhongfeng Qin
qin@buaa.edu.cn

Qiqi Li
qiqili@buaa.edu.cn

Zhe Liu
zliu21@buaa.edu.cn

¹ School of Economics and Management, Beihang University, Beijing 100191, China

² Key Laboratory of Complex System Analysis, Management and Decision (Beihang University), Ministry of Education, Beijing 100191, China

³ School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China

time series analysis with imprecise observations, autoregressive models (Yang and Liu 2019) and autoregressive moving average models (Lu et al. 2020; Xin et al. 2021) were also introduced, respectively. Some applications in other fields were also explored, such as uncertain differential game (Zhang et al. 2021; Yang and Gao 2016), uncertain extensive game (Wang et al. 2017), COVID-19 spread (Lio 2021) and uncertain queueing model (Yao 2021).

For classification problems with imprecise observations, Qin and Li (2022) introduced an USVM based on hard margin method in the framework of uncertainty theory, which extended the traditional hard margin SVM. However, the hard margin USVM is only suitable for the linearly α -separable data sets with imprecise observations. In this paper, we propose an USVM based on soft margin method for the classification problem with linearly α -nonseparable data set. Similarly, we assume the imprecise observations as uncertain variables and formulate an optimization model for the soft margin USVM. Further, we conduct numerical experiments to illustrate the application of the proposed method and evaluate its performance.

The paper is organized as follows. Some definitions and theorems in uncertainty theory are given in Sect. 2. In Sect. 3, we formulate an USVM based on soft margin method for linearly α -nonseparable data sets in uncertain environment. Then, Sect. 4 presents two examples to show the application of the soft margin USVM. Finally, we give a conclusion of the paper in Sect. 5.

2 Preliminaries

In this section, we sketch some definitions and theorems used in this paper.

Let \mathcal{L} be a σ -algebra on a nonempty set Γ . A set function $\mathcal{M} : \mathcal{L} \rightarrow [0, 1]$ is called an uncertain measure (Liu 2007; 2009) if it satisfies: (1) $\mathcal{M}\{\Gamma\} = 1$ for the universal set Γ ; (2) $\mathcal{M}\{\Lambda\} + \mathcal{M}\{\Lambda^c\} = 1$ for any $\Lambda \in \mathcal{L}$; (3) For every countable sequence $\Lambda_1, \Lambda_2, \dots$, $\mathcal{M}\{\bigcup_{i=1}^\infty \Lambda_i\} \leq \sum_{i=1}^\infty \mathcal{M}\{\Lambda_i\}$, where the triple $(\Gamma, \mathcal{L}, \mathcal{M})$ is called an uncertainty space; (4) Let $(\Gamma_k, \mathcal{L}_k, \mathcal{M}_k)$ be uncertainty spaces for $k = 1, 2, \dots$. The product uncertain measure \mathcal{M} is an uncertain measure satisfying

$$\mathcal{M}\left\{\prod_{i=1}^\infty \Lambda_k\right\} = \bigwedge_{i=1}^\infty \mathcal{M}_k\{\Lambda_k\},$$

where Λ_k are arbitrarily chosen sets from \mathcal{L}_k for $k = 1, 2, \dots$, respectively.

An uncertain variable (Liu 2007) τ is a measurable function from an uncertainty space $(\Gamma, \mathcal{L}, \mathcal{M})$ to the set of real numbers, i.e., the set $\{\tau \in B\} = \{\gamma \in \Gamma \mid \tau(\gamma) \in B\}$ is an event in \mathcal{L} for any Borel set B . The function $Y(x) = \mathcal{M}\{\tau \leq x\}, x \in \mathfrak{R}$ is called the uncertainty distribution of τ .

Theorem 1 (Liu 2010) *Let $\xi_1, \xi_2, \dots, \xi_n$ be independent uncertain variables with inverse uncertainty distributions $\Phi_1^{-1}, \Phi_2^{-1}, \dots, \Phi_n^{-1}$, respectively. If $f(x_1, x_2, \dots, x_n)$ is strictly increasing with respect to x_1, x_2, \dots, x_m , and strictly decreasing with respect to $x_{m+1}, x_{m+2}, \dots, x_n$, then the uncertain variable $\xi = f(\xi_1, \xi_2, \dots, \xi_n)$ has an inverse uncertainty distribution*

$$\Psi^{-1}(u) = f(\Phi_1^{-1}(u), \Phi_2^{-1}(u), \dots, \Phi_m^{-1}(u), \Phi_{m+1}^{-1}(1-u), \dots, \Phi_n^{-1}(1-u)).$$

Theorem 2 (Liu 2015) *Assume the function $g(\mathbf{x}, \xi_1, \xi_2, \dots, \xi_n)$ is strictly increasing with respect to $\xi_1, \xi_2, \dots, \xi_k$ and strictly decreasing with respect to $\xi_{k+1}, \xi_{k+2}, \dots, \xi_n$. If $\xi_1, \xi_2, \dots, \xi_n$ are independent uncertain variables with inverse uncertainty distributions $\Phi_1^{-1}, \Phi_2^{-1}, \dots, \Phi_n^{-1}$, respectively, then the chance constraint*

$$\mathcal{M}\{g(\mathbf{x}, \xi_1, \xi_2, \dots, \xi_n) \leq 0\} \geq \alpha$$

holds if and only if

$$g(\mathbf{x}, \Phi_1^{-1}(\alpha), \dots, \Phi_k^{-1}(\alpha), \Phi_{k+1}^{-1}(1-\alpha), \dots, \Phi_n^{-1}(1-\alpha)) \leq 0. \tag{1}$$

3 Soft margin method for linearly α -nonseparable data sets

This section discusses linearly α -nonseparable data sets and presents a soft margin USVM for this situation.

Suppose the observed data set is $S = \{(\tilde{\mathbf{x}}_1, y_1), (\tilde{\mathbf{x}}_2, y_2), \dots, (\tilde{\mathbf{x}}_l, y_l)\}$, where $\tilde{\mathbf{x}}_i$ are uncertain vectors, and $y_i \in \{1, -1\}$ are crisp labels for $i = 1, 2, \dots, l$, respectively. Let $\alpha \in (0.5, 1)$ be a given confidence level. Qin and Li (2022) defined that S is linearly α -separable if there exists one hyperplane $\mathbf{w}^T \mathbf{x} + b = 0$ such that

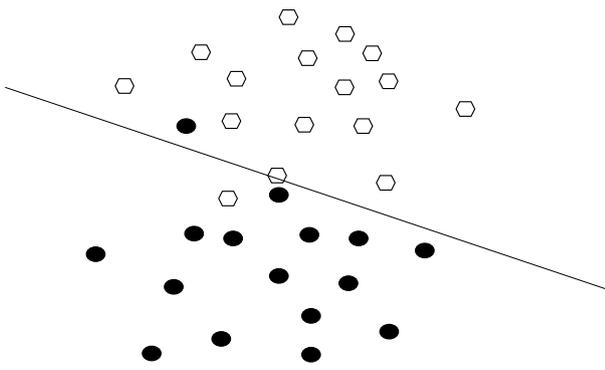
$$\mathcal{M}\{y_i \cdot (\mathbf{w}^T \tilde{\mathbf{x}}_i + b) \geq 0\} \geq \alpha, \quad i = 1, 2, \dots, l, \tag{2}$$

where $\mathbf{w} = (w_1, w_2, \dots, w_n)$ is an n -dimensional vector with an Euclidean norm $\|\mathbf{w}\|$. This definition implies that linearly α -separable data sets are those which can be classified by a hyperplane at confidence level α .

If there exists no hyperplane $\mathbf{w}^T \mathbf{x} + b = 0$ such that

$$\mathcal{M}\{y_i \cdot (\mathbf{w}^T \tilde{\mathbf{x}}_i + b) \geq 0\} \geq \alpha$$

for all $i = 1, 2, \dots, l$, then we call the data set S linearly α -nonseparable. For instance, Fig. 1 gives an example of a linearly α -nonseparable data set. There is no hyperplane that makes Inequality (2) hold. Thus the hard margin method (Qin and Li 2022) can not be applied in this case.



- Indicates the points with label -1
- Indicates the points with label 1

Fig. 1 A hyperplane and a linearly α -nonseparable data set

Suppose that S is linearly α -nonseparable. Then, for each observation $(\tilde{\mathbf{x}}_i, y_i)$, we may seek a nonnegative slack variable s_i such that

$$\mathcal{M}\{y_i \cdot (\mathbf{w}^T \tilde{\mathbf{x}}_i + b) \geq -s_i\} \geq \alpha. \tag{3}$$

Note that such an inequality holds if s_i is large enough. Therefore, we always want the slack variable s_i to be as small as possible. Note that if $s_i = 0$ for all i , then the set S degenerates into a linearly α -separable one.

Theorem 3 Suppose that the components $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in}$ of uncertain vector $\tilde{\mathbf{x}}_i$ are independent. Let Φ_{ij}^{-1} denote the inverse uncertainty distributions of \tilde{x}_{ij} for $i = 1, 2, \dots, l$ and $j = 1, 2, \dots, n$, respectively. Then Inequality (3) is equivalent to the following crisp form

$$y_i \left(\sum_{j=1}^n w_j \Upsilon_{ij}^{-1}(\alpha, -y_i w_j) + b \right) \geq -s_i, \tag{4}$$

where

$$\begin{aligned} \Upsilon_{ij}^{-1}(\alpha, -y_i w_j) &= \Phi_{ij}^{-1}(\alpha) \cdot I_{\{-y_i w_j \geq 0\}} \\ &+ \Phi_{ij}^{-1}(1 - \alpha) \cdot I_{\{-y_i w_j < 0\}}. \end{aligned} \tag{5}$$

Proof The argument breaks into two cases. Case I: When $y_i = 1$. In this case, Inequality (3) degenerates into

$$\begin{aligned} \mathcal{M}\{\mathbf{w}^T \tilde{\mathbf{x}}_i + b \geq -s_i\} &= \mathcal{M}\{-\mathbf{w}^T \tilde{\mathbf{x}}_i - b - s_i \leq 0\} \\ &= \mathcal{M}\left\{-\sum_{j=1}^n w_j \tilde{x}_{ij} - b - s_i \leq 0\right\} \geq \alpha. \end{aligned} \tag{6}$$

The inverse uncertainty distribution of $-\mathbf{w}^T \tilde{\mathbf{x}}_i - b - s_i = -\sum_{j=1}^n w_j \tilde{x}_{ij} - b - s_i$ is

$$\begin{aligned} F_{1i}^{-1}(u) &= -\sum_{j=1}^n w_j \left(\Phi_{ij}^{-1}(1 - u) \cdot I_{\{w_j \geq 0\}} \right. \\ &\quad \left. + \Phi_{ij}^{-1}(u) \cdot I_{\{w_j < 0\}} \right) - b - s_i \\ &= -\sum_{j=1}^n w_j \Upsilon_{ij}^{-1}(u, -w_j) - b - s_i. \end{aligned}$$

Thus, it follows from Theorem 2 that Inequality (6) is satisfied if and only if

$$F_{1i}^{-1}(\alpha) = -\sum_{j=1}^n w_j \Upsilon_{ij}^{-1}(\alpha, -w_j) - b - s_i \leq 0,$$

which is equivalent to Inequality (4), i.e.

$$\begin{aligned} &\sum_{j=1}^n w_j \Upsilon_{ij}^{-1}(\alpha, -w_j) + b \\ &= y_i \left(\sum_{j=1}^n w_j \Upsilon_{ij}^{-1}(\alpha, -y_i w_j) + b \right) \\ &\geq -s_i. \end{aligned}$$

Case II: When $y_i = -1$. In this case, Inequality (3) degenerates into

$$\begin{aligned} \mathcal{M}\{\mathbf{w}^T \tilde{\mathbf{x}}_i + b - s_i \leq 0\} &= \mathcal{M}\left\{\sum_{j=1}^n w_j \tilde{x}_{ij} + b - s_i \leq 0\right\} \\ &\geq \alpha. \end{aligned} \tag{7}$$

Similarly, the inverse uncertainty distribution of $\sum_{j=1}^n w_j \tilde{x}_{ij} + b - s_i$ is

$$\begin{aligned} F_{2i}^{-1}(u) &= \sum_{j=1}^n w_j \left(\Phi_{ij}^{-1}(u) \cdot I_{\{w_j \geq 0\}} \right. \\ &\quad \left. + \Phi_{ij}^{-1}(1 - u) \cdot I_{\{w_j < 0\}} \right) + b - s_i \\ &= \sum_{j=1}^n w_j \Upsilon_{ij}^{-1}(u, w_j) + b - s_i. \end{aligned}$$

Thus, Inequality (7) is satisfied if and only if

$$F_{2i}^{-1}(\alpha) = \sum_{j=1}^n w_j \Upsilon_{ij}^{-1}(\alpha, w_j) + b - s_i \leq 0,$$

which is equivalent to Inequality (4), i.e.,

$$\begin{aligned}
 & - \left(\sum_{j=1}^n w_j Y_{ij}^{-1}(\alpha, w_j) + b \right) \\
 & = y_i \left(\sum_{j=1}^n w_j Y_{ij}^{-1}(\alpha, -y_i w_j) + b \right) \\
 & \geq -s_i.
 \end{aligned}$$

□

It follows from Qin and Li (2022) that the distance from a data set S to the hyperplane $\mathbf{w}^T \mathbf{x} + b = 0$ is

$$\min_i E \left[\frac{|\mathbf{w}^T \tilde{\mathbf{x}}_i + b|}{\|\mathbf{w}\|} \right],$$

which determines the margin for the positive class and negative class. In order to maximize the margin and meanwhile minimize the sum of slack variables, we choose

$$\min_i E \left[\frac{|\mathbf{w}^T \tilde{\mathbf{x}}_i + b|}{\|\mathbf{w}\|} \right] - C \sum_{i=1}^l s_i$$

as the objective function, in which $C > 0$ is a penalty coefficient. Correspondingly, we formulate the following optimization model for the soft margin USVM as

$$\left\{ \begin{array}{ll} \max_{\mathbf{w}, b, s_1, s_2, \dots, s_l} & \min_i E \left[\frac{|\mathbf{w}^T \tilde{\mathbf{x}}_i + b|}{\|\mathbf{w}\|} \right] - C \sum_{i=1}^l s_i \\ \text{s.t.} & \mathcal{M}\{y_i(\mathbf{w}^T \tilde{\mathbf{x}}_i + b) \geq -s_i\} \geq \alpha, \\ & i = 1, 2, \dots, l \\ & s_i \geq 0, \quad i = 1, 2, \dots, l. \end{array} \right. \quad (8)$$

If $(\mathbf{w}, b, \mathbf{s})$ is an optimal solution to Model (8), then for any constant $\lambda > 0$, $(\lambda \mathbf{w}, \lambda b, \lambda \mathbf{s})$ is also an optimal solution to Model (8). To obtain a unique optimal solution, a constraint

$$\left\{ \begin{array}{ll} \max_{w_1, w_2, \dots, w_n, b, s_1, s_2, \dots, s_l} & \min_i \int_0^1 \left| \sum_{j=1}^n w_j Y_{ij}^{-1}(u, w_j) + b \right| du - C \sum_{i=1}^l s_i \\ \text{s.t.} & y_i \left(\sum_{j=1}^n w_j Y_{ij}^{-1}(\alpha, -y_i w_j) + b \right) \geq -s_i, \quad i = 1, 2, \dots, l \\ & \sum_{j=1}^n w_j^2 = 1 \\ & s_i \geq 0, \quad i = 1, 2, \dots, l, \end{array} \right. \quad (10)$$

on the coefficients is necessary. Without loss of generality, we insert a new constraint $\|\mathbf{w}\| = 1$ to Model (8). The objective function correspondingly becomes

$$\begin{aligned}
 & \min_i E \left[|\mathbf{w}^T \tilde{\mathbf{x}}_i + b| \right] - C \sum_{i=1}^l s_i \\
 & = \min_i E \left[\left| \sum_{j=1}^n w_j \tilde{x}_{ij} + b \right| \right] - C \sum_{i=1}^l s_i.
 \end{aligned}$$

Further, we obtain the following model for the soft margin USVM,

$$\left\{ \begin{array}{ll} \max_{w_1, w_2, \dots, w_n, b, s_1, s_2, \dots, s_l} & \min_i E \left[\left| \sum_{j=1}^n w_j \tilde{x}_{ij} + b \right| \right] - C \sum_{i=1}^l s_i \\ \text{s.t.} & \mathcal{M} \left\{ y_i \left(\sum_{j=1}^n w_j \tilde{x}_{ij} + b \right) \geq -s_i \right\} \geq \alpha, \\ & i = 1, 2, \dots, l \\ & \sum_{j=1}^n w_j^2 = 1 \\ & s_i \geq 0, \quad i = 1, 2, \dots, l. \end{array} \right. \quad (9)$$

For the linearly α -separable case, we have $s_i = 0$ for $i = 1, 2, \dots, l$. In this case, Model (9) degenerates into the hard margin USVM proposed by Qin and Li (2022).

Next we give the crisp equivalent form of Model (9) when inverse uncertainty distributions of imprecise observations are given.

Theorem 4 *Suppose that the components $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{in}$ of each uncertain vectors $\tilde{\mathbf{x}}_i$ are independent for $i = 1, 2, \dots, l$. Let Φ_{ij}^{-1} be the inverse uncertainty distributions of \tilde{x}_{ij} for $i = 1, 2, \dots, l$ and $j = 1, 2, \dots, n$, respectively. Then Model (9) is equivalent to Model (10).*

where $Y_{ij}^{-1}(\alpha, -y_i w_j)$ is given in Equality (5), and

$$Y_{ij}^{-1}(u, w_j) = \Phi_{ij}^{-1}(u) \cdot I_{\{w_j \geq 0\}} + \Phi_{ij}^{-1}(1 - u) \cdot I_{\{w_j < 0\}}.$$

Proof It follows from Theorem 1 that the inverse uncertainty distributions of

$$w^T \tilde{x}_i + b = \sum_{j=1}^n w_j \tilde{x}_{ij} + b$$

are

$$\sum_{j=1}^n w_j Y_{ij}^{-1}(\alpha, w_j) + b$$

for $i = 1, 2, \dots, l$, respectively. Thus, the objective function becomes

$$\int_0^1 \left| \sum_{j=1}^n w_j Y_{ij}^{-1}(u, w_j) + b \right| du - C \sum_{i=1}^l s_i.$$

The constraints follow from Theorem 3 immediately. \square

4 Numerical experiments

In this section, we conduct two examples to show the application of the soft margin USVM. We suppose that all the imprecise observations are characterized by linear uncertain variables. The uncertainty distribution and the inverse uncertainty distribution of a linear uncertain variable $\mathcal{L}(a, b)$ is

Table 1 Data set S_1

i	A		i	B	
	\tilde{x}_{1i}	\tilde{x}_{2i}		\tilde{x}_{1i}	\tilde{x}_{2i}
1	$\mathcal{L}(17,21)$	$\mathcal{L}(71,74)$	31	$\mathcal{L}(26,28)$	$\mathcal{L}(61,63)$
2	$\mathcal{L}(9,10)$	$\mathcal{L}(4,5)$	32	$\mathcal{L}(31,33)$	$\mathcal{L}(54,57)$
3	$\mathcal{L}(12,14)$	$\mathcal{L}(7,8)$	33	$\mathcal{L}(29,31)$	$\mathcal{L}(63,66)$
4	$\mathcal{L}(10,11)$	$\mathcal{L}(12,14)$	34	$\mathcal{L}(25,27)$	$\mathcal{L}(47,49)$
5	$\mathcal{L}(21,23)$	$\mathcal{L}(72,74)$	35	$\mathcal{L}(30,33)$	$\mathcal{L}(31,33)$
6	$\mathcal{L}(24,26)$	$\mathcal{L}(72,75)$	36	$\mathcal{L}(37,39)$	$\mathcal{L}(76,79)$
7	$\mathcal{L}(17,19)$	$\mathcal{L}(42,45)$	37	$\mathcal{L}(20,22)$	$\mathcal{L}(0,1)$
8	$\mathcal{L}(18,20)$	$\mathcal{L}(77,80)$	38	$\mathcal{L}(28,30)$	$\mathcal{L}(57,60)$
9	$\mathcal{L}(23,25)$	$\mathcal{L}(44,46)$	39	$\mathcal{L}(21,23)$	$\mathcal{L}(39,41)$
10	$\mathcal{L}(18,20)$	$\mathcal{L}(46,48)$	40	$\mathcal{L}(30,32)$	$\mathcal{L}(68,71)$
11	$\mathcal{L}(15,17)$	$\mathcal{L}(17,19)$	41	$\mathcal{L}(30,32)$	$\mathcal{L}(17,19)$
12	$\mathcal{L}(12,14)$	$\mathcal{L}(7,9)$	42	$\mathcal{L}(34,36)$	$\mathcal{L}(50,53)$
13	$\mathcal{L}(17,19)$	$\mathcal{L}(31,33)$	43	$\mathcal{L}(24,26)$	$\mathcal{L}(34,36)$
14	$\mathcal{L}(12,13)$	$\mathcal{L}(35,37)$	44	$\mathcal{L}(31,33)$	$\mathcal{L}(49,52)$
15	$\mathcal{L}(12,14)$	$\mathcal{L}(35,37)$	45	$\mathcal{L}(28,30)$	$\mathcal{L}(60,62)$
16	$\mathcal{L}(12,13)$	$\mathcal{L}(38,40)$	46	$\mathcal{L}(36,38)$	$\mathcal{L}(72,75)$
17	$\mathcal{L}(14,15)$	$\mathcal{L}(7,8)$	47	$\mathcal{L}(32,34)$	$\mathcal{L}(63,66)$
18	$\mathcal{L}(14,16)$	$\mathcal{L}(26,28)$	48	$\mathcal{L}(26,28)$	$\mathcal{L}(42,44)$
19	$\mathcal{L}(22,24)$	$\mathcal{L}(43,46)$	49	$\mathcal{L}(26,28)$	$\mathcal{L}(71,74)$
20	$\mathcal{L}(9,10)$	$\mathcal{L}(7,8)$	50	$\mathcal{L}(24,26)$	$\mathcal{L}(28,30)$
21	$\mathcal{L}(10,11)$	$\mathcal{L}(2,3)$	51	$\mathcal{L}(25,26)$	$\mathcal{L}(13,14)$
22	$\mathcal{L}(15,16)$	$\mathcal{L}(48,50)$	52	$\mathcal{L}(34,35)$	$\mathcal{L}(70,72)$
23	$\mathcal{L}(18,19)$	$\mathcal{L}(63,65)$	53	$\mathcal{L}(30,31)$	$\mathcal{L}(45,46)$
24	$\mathcal{L}(17,18)$	$\mathcal{L}(58,60)$	54	$\mathcal{L}(38,39)$	$\mathcal{L}(41,42)$
25	$\mathcal{L}(11,12)$	$\mathcal{L}(46,48)$	55	$\mathcal{L}(34,35)$	$\mathcal{L}(40,42)$
26	$\mathcal{L}(13,14)$	$\mathcal{L}(52,54)$	56	$\mathcal{L}(22,23)$	$\mathcal{L}(30,31)$
27	$\mathcal{L}(14,15)$	$\mathcal{L}(22,23)$	57	$\mathcal{L}(28,29)$	$\mathcal{L}(37,38)$
28	$\mathcal{L}(15,16)$	$\mathcal{L}(65,66)$	58	$\mathcal{L}(26,27)$	$\mathcal{L}(24,25)$
29	$\mathcal{L}(20,21)$	$\mathcal{L}(28,29)$	59	$\mathcal{L}(32,33)$	$\mathcal{L}(25,26)$
30	$\mathcal{L}(19,20)$	$\mathcal{L}(55,56)$	60	$\mathcal{L}(35,36)$	$\mathcal{L}(60,62)$

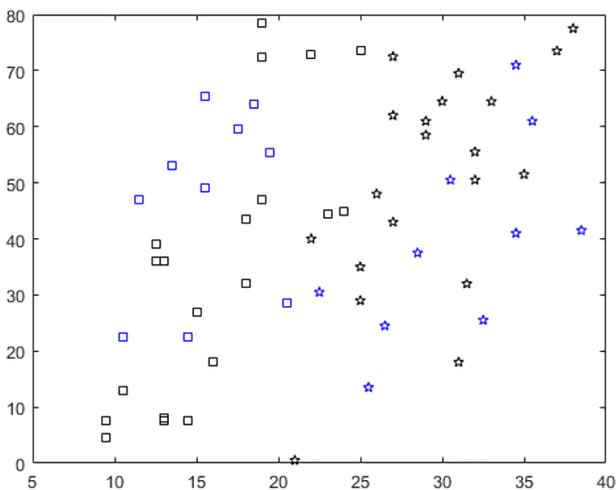


Fig. 2 Training data and test data in Example 1

$$\Phi(x) = \begin{cases} 0, & \text{if } x \leq a \\ \frac{x-a}{b-a}, & \text{if } a < x \leq b \\ 1, & \text{if } b < x \end{cases}$$

and

$$\Phi^{-1}(u) = (1 - u)a + ub,$$

respectively.

Example 1 We consider the first data set S_1 which contains 60 observations. The data is denoted as $(\tilde{x}_{i1}, \tilde{x}_{i2})$ with the label $y_i \in \{1, -1\}$, $i = 1, 2, \dots, 60$. Let Φ_{i1}^{-1} and Φ_{i2}^{-1} denote the inverse uncertainty distributions of \tilde{x}_{i1} and \tilde{x}_{i2} for $i = 1, 2, \dots, 60$, respectively. The detailed data are shown in Table 1.

Table 2 Optimal solutions (w_1, w_2, b) under different penalty coefficients and confidence levels

	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.99$
C=0	(0.9941, -0.1089, -19.5222)	(0.9940, -0.1092, -19.5023)	(0.9986, -0.0531, -22.3406)
C=1	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=2	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=3	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=4	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=10	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=20	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=40	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)

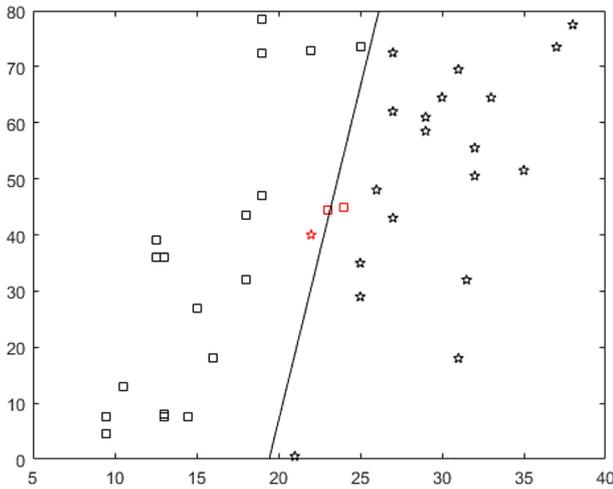


Fig. 3 Misclassified data in S_1

For brevity, denote by A the negative class and denote by B the positive one. The first 20 data in A and the first 20 data in B are utilized as the training data, while the rest are utilized as the test data. The training data and the test data are plotted in Fig. 2, where black squares represent the training data in A, black stars represent the training data in B, blue squares represent the test data in A, and blue stars represent the test data in B. It can be seen that S_1 is a linearly α -nonseparable data set.

For the data set S_1 , Model (10) is reformulated and we obtain Model (11).

$$\left\{ \begin{array}{l} \max_{w_1, w_2, b, s_i, i \in I} \min_{i \in I} \int_0^1 |w_1 Y_{i1}^{-1}(u, w_1) + w_2 Y_{i2}^{-1}(u, w_2) + b| du - C \sum_{i \in I} s_i \\ \text{s.t. } y_i (w_1 Y_{i1}^{-1}(\alpha, -y_i w_1) + w_2 Y_{i2}^{-1}(\alpha, -y_i w_2) + b) \geq -s_i, \quad i \in I \\ w_1^2 + w_2^2 = 1 \\ s_i \geq 0, \quad i \in I, \end{array} \right. \tag{11}$$

Table 3 Predicting the data label

No.	21	22	23	24	25
Forecasted label	-1	-1	-1	-1	-1
No.	26	27	28	29	30
Forecasted label	-1	-1	-1	-1	-1
No.	51	52	53	54	55
Forecasted label	+1	+1	+1	+1	+1
No.	56	57	58	59	60
Forecasted label	-1	+1	+1	+1	+1

Table 4 Confusion matrix in Example 1

Real label	Forecasted label	
	Positive	Negative
Positive	9	1
Negative	0	10

where $I = \{1, 2, \dots, 20, 31, 32, \dots, 50\}$.

Setting confidence levels $\alpha = 0.90, 0.95, 0.99$, respectively, we employ ‘fmincon’ package in Matlab to seek the optimal solution to Model (11) under different parameters. The coefficients for the optimal hyperplanes are reported in Table 2. It can be seen that the optimal hyperplanes appear to change slightly under different penalty coefficients C . And a larger C seems to produce more reliable results. However, the results are slightly different for distinct confidence levels.

When the parameter $C = 1$ and confidence level $\alpha = 0.95$, the optimal hyperplane determined by Model (11) for S_1 is $0.9966x_1 - 0.0820x_2 = 19.9543$. The misclassified data are the 9th, 19th, and 39th observations, and their corresponding slack variables are 1.0278, 0.1051, and 2.2807, respectively. These misclassified data are pointed in red in Fig. 3, where the data \tilde{x}_i with label $y_i = 1$ are indicated as squares, and the data \tilde{x}_i with label $y_i = -1$ are indicated as stars for $i = 1, 2, \dots, 40$.

Next, we evaluate the generalization ability of the soft margin USVM by following the classification approach given by Qin and Li (2022). We evaluate Model (11) by the test data with the confidence level $\alpha = 0.90$. Denote the number of the points in positive class for right classification as true positive (TP), the number of the points in positive class for the false classification as false negative (FN), the number of the points in negative class for the right classification as true negative (TN), and the number of points in negative class for the false classification as false positive (FP). The results are presented in Table 3 and summarized in a confusion matrix which is listed in Table 4.

Further, commonly used evaluation indicators for binary classification problem are accuracy, precision, recall, and F-score, which are defined as follows,

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}},$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

and

$$F = \frac{2}{\text{Recall}/2 + \text{Precision}/2} = \frac{\text{TP}}{\text{TP} + (\text{FP} + \text{FN})/2}.$$

We expect a higher number of the right classification for positive class as well as negative class. That is, above metrics indicate a good classification result if they are close to 1. Accordingly, we obtain

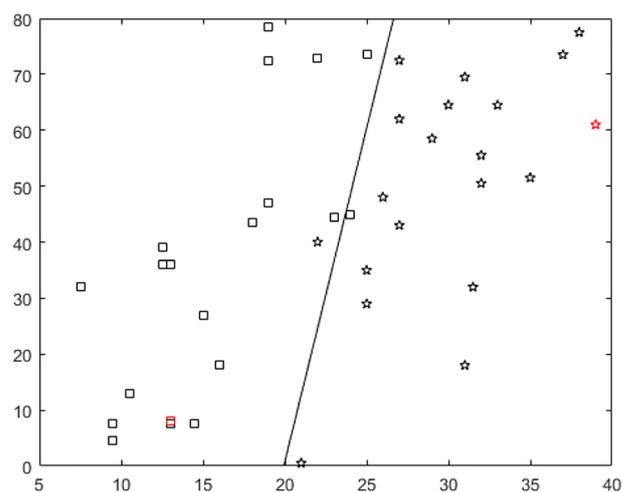


Fig. 4 Optimal hyperplane for data set S_2

$$\text{Accuracy} = \frac{9 + 10}{9 + 1 + 0 + 10} = 95.0\%,$$

$$\text{Precision} = \frac{9}{9 + 1} = 90.0\%,$$

$$\text{Recall} = \frac{9}{9 + 0} = 100.0\%,$$

and

$$F = \frac{9}{9 + (0 + 1)/2} = 94.7\%.$$

The results show that the soft margin USVM performs reasonably well for the data set S_1 .

Next, we explore the robustness of the soft margin USVM by repeating the analysis and changing original data. Suppose that the uncertainty distributions of the 12th and the 35th data become $(\mathcal{L}(12, 14), \mathcal{L}(7, 9))$ and $(\mathcal{L}(38, 40), \mathcal{L}(60, 62))$, respectively, and their labels remain unchanged. Then, we obtain a new data set S_2 . Running ‘fmincon’ function in Matlab, we obtain the optimal solutions under confidence levels 0.90, 0.95 and 0.99, respectively, which are reported in Table 5.

Table 5 Optimal solutions (w_1, w_2, b) under different penalty coefficients and confidence levels with two modified data

	$\alpha=0.90$	$\alpha=0.95$	$\alpha=0.99$
C=0	(0.9936, -0.1133, -18.2130)	(0.9894, -0.1450, -16.7865)	(1.0000, -0.0034, -23.3601)
C=1	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=2	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=3	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=4	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=10	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=20	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)
C=40	(0.9966, -0.0821, -20.0578)	(0.9966, -0.0820, -19.9543)	(0.9965, -0.0836, -19.8671)

Table 6 Optimal solutions (w_1, w_2, b) under different penalty coefficients and confidence levels for data set S_3

	$\alpha=0.90$	$\alpha=0.95$	$\alpha=0.99$
C=0	(0.1104, 0.9939, -35.5314)	(0.1104, 0.9939, -35.5309)	(0.1104, 0.9939, -35.5309)
C=1	(0.1104, 0.9939, -35.5314)	(0.1104, 0.9939, -35.5311)	(0.1104, 0.9939, -35.5313)
C=2	(0.1104, 0.9939, -35.5313)	(0.1104, 0.9939, -35.5315)	(0.1104, 0.9939, -35.5313)
C=3	(0.1104, 0.9939, -35.5309)	(0.1104, 0.9939, -35.5313)	(0.1104, 0.9939, -35.5313)
C=4	(0.1104, 0.9939, -35.5313)	(0.1104, 0.9939, -35.5314)	(0.1105, 0.9939, -35.5333)
C=10	(0.1104, 0.9939, -35.5314)	(0.1092, 0.9940, -35.5005)	(0.1617, 0.9868, -36.9185)
C=20	(0.1104, 0.9939, -35.5314)	(0.1623, 0.9867, -36.9345)	(0.1104, 0.9939, -35.5316)
C=40	(0.1104, 0.9939, -35.5315)	(0.1623, 0.9867, -36.9342)	(0.1618, 0.9868, -36.9206)

Table 7 Optimal solutions (w_1, w_2, b) under different penalty coefficients and confidence levels for data set S_4

	$\alpha = 0.90$	$\alpha = 0.95$	$\alpha = 0.99$
C=0	(0.1504, 0.9886, -36.6204)	(0.1104, 0.9939, -35.5314)	(0.1104, 0.9939, -35.5317)
C=1	(0.1104, 0.9939, -35.5313)	(0.1104, 0.9939, -35.5314)	(0.1104, 0.9939, -35.5312)
C=2	(0.1104, 0.9939, -35.5314)	(0.1104, 0.9939, -35.5314)	(0.1104, 0.9939, -35.5313)
C=3	(0.1104, 0.9939, -35.5305)	(0.1104, 0.9939, -35.5313)	(0.1104, 0.9939, -35.5313)
C=4	(0.1104, 0.9939, -35.5314)	(0.1104, 0.9939, -35.5312)	(0.1104, 0.9939, -35.5314)
C=10	(0.1104, 0.9939, -35.5313)	(0.1528, 0.9883, -36.6848)	(0.1105, 0.9939, -35.5338)
C=20	(0.1104, 0.9939, -35.5314)	(0.1102, 0.9939, -35.5246)	(0.1523, 0.9883, -36.6700)
C=40	(0.1535, 0.9881, -36.7032)	(0.1108, 0.9938, -35.5427)	(0.1104, 0.9939, -35.5315)

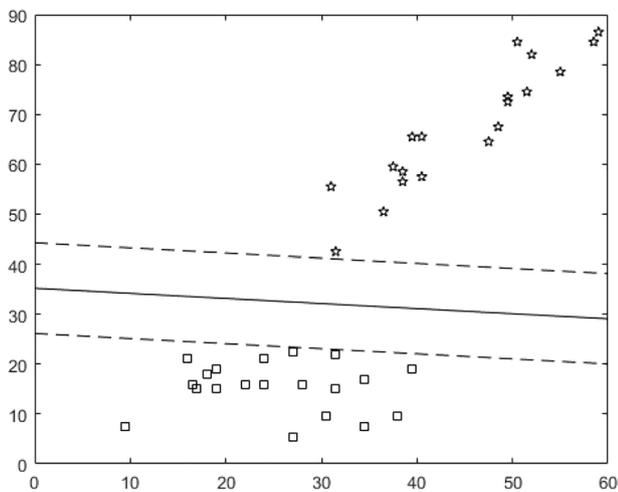


Fig. 5 Optimal hyperplane for data set S_3

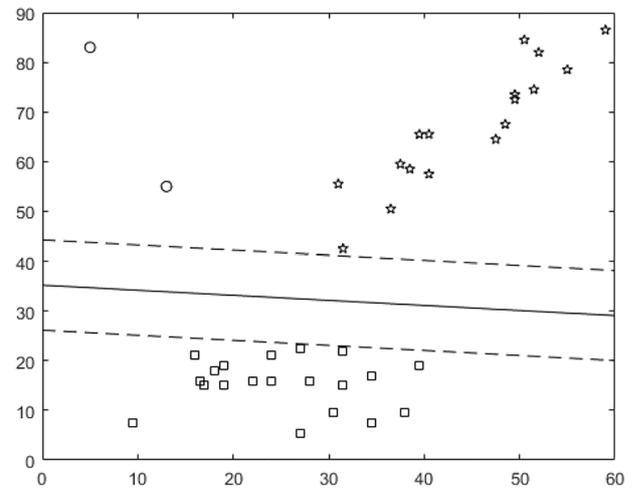


Fig. 6 Optimal hyperplane for data set S_4

When the penalty coefficient $C = 1$ and the confidence level $\alpha = 0.95$, the optimal hyperplane determined by Model (11) for the linearly α -nonseparable data set S_2 is $0.9966x_1 - 0.0820x_2 = 19.9543$. Figure 4 shows the optimal hyperplane as a line and two modified data in red. The misclassified data are the 9th, the 19th, and the 39th data, and the corresponding slack variables are 1.0278, 0.1051 and 2.2807, respectively.

Table 5 indicates that the optimal solution will change when the penalty coefficient C from 0 to 1. But when C continues increasing, the optimal solutions are unchanged under data modification. This is possibly because the penalty term in the objective function plays a leading role.

Example 2 In this example, we examine the soft margin USVM based on the linearly α -separable data set given by Qin and Li (2022). Denote the data set as S_3 , whose elements are denoted as $(\tilde{x}_{i1}, \tilde{x}_{i2})$ with label $y_i \in \{1, -1\}$,

$i = 1, 2, \dots, 40$. Let Φ_{i1}^{-1} and Φ_{i2}^{-1} denote the inverse uncertainty distributions of \tilde{x}_{i1} and \tilde{x}_{i2} for $i = 1, 2, \dots, 40$, respectively.

To classify the data in S_3 , Model (10) is reformulated, and we obtain Model (12) with a penalty coefficient. This calls for an examination of the classification power of Model (12) from the perspective of different penalty coefficients as well as confidence levels. As before, we explore the optimal solutions for confidence levels $\alpha \in \{0.90, 0.95, 0.99\}$, and penalty coefficient $C \in \{1, 2, 3, 4, 10, 20, 40\}$. The optimal solution (w_1, w_2, b) to Model (12) can be obtained by employing ‘fmincon’ function in Matlab. We list the optimal solutions in Table 6.

$$\left\{ \begin{array}{l} \max_{w_1, w_2, b, s_1, s_2, \dots, s_{40}} \min_i \int_0^1 |w_1 Y_{i1}^{-1}(u, w_1) + w_2 Y_{i2}^{-1}(u, w_2) + b| du - C \sum_{i=1}^{40} s_i \\ \text{s.t. } y_i (w_1 Y_{i1}^{-1}(\alpha, -y_i w_1) + w_2 Y_{i2}^{-1}(\alpha, -y_i w_2) + b) \geq -s_i, \quad i = 1, 2, \dots, 40 \\ w_1^2 + w_2^2 = 1 \\ s_i \geq 0, \quad i = 1, 2, \dots, 40. \end{array} \right. \quad (12)$$

When the penalty coefficient C increases, the results tend to be unchanged. Letting $C = 10$ and $\alpha = 0.90$, we plot the results in Fig. 5 to show that the optimal hyperplane correctly classifies the data set. Suppose that the uncertainty distributions of the 26th and the 30th data become $(\mathcal{L}(1, 3), \mathcal{L}(81, 83))$ and $(\mathcal{L}(3, 4), \mathcal{L}(55, 58))$, respectively, and their labels remain unchanged. Then we obtain a new data set S_4 . As indicated by Fig. 6, two new data are indicated in circles.

Similarly, we may obtain the corresponding optimal hyperplane and their coefficients are presented in Table 7. The results are not sensitive to the confidence level α and the penalty coefficient C . From Fig. 6, we observe that all the data with positive labels are lying in the same half-plane determined by the optimal hyperplane, and the data with negative labels are lying in the contrary part. In the meantime, the widest margin is manifested by the hyperplane. The results are consistent with the main conclusions in Qin and Li (2022). We can see that the soft margin USVM is not only capable of handling the classification problem with linearly α -separable data sets, but also capable of the case with linearly α -nonseparable data sets.

5 Conclusions

This paper proposed an USVM based on soft margin method for linearly α -nonseparable data set with imprecise observations in the framework of uncertainty theory. The equivalent

crisp model was derived based on the inverse uncertainty distributions. Two numerical examples were conducted to demonstrate the application and effectiveness of the proposed model. The modelling idea on soft margin USVM shed a light on some future research directions. We may further study the nonlinear classification problems, multi-classification problem and so on in the case with imprecise observations.

Acknowledgements This work was supported by National Natural Science Foundation of China (Nos. 72071008 and 71771011).

Declarations

Conflict of interest The author declares that there is no conflict of interests regarding the publication of this paper.

Ethics statement This work did not involve any active collection of human data.

References

Boser B, Guyon I, Vapnik V (1992) A training algorithm for optimal margin classifiers. In: Fifth annual workshop on computational learning theory, vol 5. ACM, Pittsburgh, pp 144–152

Burges C (1998) A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 2(2):121–167

Chen D (2020) Tukey’s biweight estimation for uncertain regression model with imprecise observations. *Soft Comput* 24:16803–16809

Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297

Fang L, Hong Y (2020) Uncertain revised regression analysis with responses of logarithmic, square root and reciprocal transformations. *Soft Comput* 24:2655–2670

Gautam N, Singh A, Kumar K et al (2021) Investigation on performance analysis of support vector machine for classification of abnormal regions in medical image. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-021-02965-9>

Hu Z, Gao J (2020) Uncertain Gompertz regression model with imprecise observations. *Soft Comput* 24:2543–2549

Li Q, Qin Z, Liu Z (2022) Uncertain support vector regression with imprecise observations. *J Intell Fuzzy Syst* 43:3403–3409

Lio W (2021) Uncertain statistics and COVID-19 spread in China. *J Uncertain Syst* 14(1):2150008

Lio W, Liu B (2018) Residual and confidence interval for uncertain regression model with imprecise observations. *J Intell Fuzzy Syst* 35(2):2573–2583

- Lio W, Liu B (2020) Uncertain maximum likelihood estimation with application to uncertain regression analysis. *Soft Comput* 24:9351–9360
- Liu B (2007) *Uncertainty theory*, 2nd edn. Springer, Berlin
- Liu B (2009) Some research problems in uncertainty theory. *J Uncertain Syst* 3(1):3–10
- Liu B (2010) *Uncertainty theory: a branch of mathematics for modeling human uncertainty*. Springer, Berlin
- Liu B (2015) *Uncertainty theory*, 4th edn. Springer, Berlin
- Liu Z, Yang Y (2020) Least absolute deviations uncertain regression with imprecise observations. *Fuzzy Optim Decis Mak* 19:33–52
- Lu J, Peng J, Chen J et al (2020) Prediction method of autoregressive moving average models for uncertain time series. *Int J Gen Syst* 49(5):546–572
- Mukkamala S, Janoski G, Sung A (2002) Intrusion detection using neural networks and support vector machines. In: *Proceedings of the 2002 international joint conference on neural networks*, vol 2, pp 1702–1707
- Okwuashi O, Ndehedehe C (2020) Deep support vector machine for hyperspectral image classification. *Pattern Recognit* 103:107298
- Priyadharsini N, Chitra D (2021) A kernel support vector machine based anomaly detection using spatio-temporal motion pattern models in extremely crowded scenes. *J Ambient Intell Human Comput* 12:5225–5234
- Qin Z, Li Q (2022) An uncertain support vector machine with imprecise observations. Technical Report
- Vapnik V (1995) *The nature of statistical learning theory*. Springer, New York
- Vapnik V (1999) An overview of statistical learning theory. *IEEE Trans Neural Netw* 10(5):988–999
- VenkateswarLal P, Nitta G, Prasad A (2019) Ensemble of texture and shape descriptors using support vector machine classification for face recognition. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-019-01192-7>
- Wang Y, Luo S, Gao J (2017) Uncertain extensive game with application to resource allocation of national security. *J Ambient Intell Human Comput* 8:797–808
- Wang H, Zheng B, Yoon S et al (2018) A support vector machine-based ensemble algorithm for breast cancer diagnosis. *Eur J Oper Res* 267(2):687–699
- Xin Y, Yang X, Gao J (2021) Least squares estimation for the high-order uncertain moving average model with application to carbon dioxide emissions. *Int J Gen Syst* 50(6):724–740
- Yang X, Gao J (2016) Linear-quadratic uncertain differential game with application to resource extraction problem. *IEEE Trans Fuzzy Syst* 24(4):819–826
- Yang X, Liu B (2019) Uncertain time series analysis with imprecise observations. *Fuzzy Optim Decis Mak* 18(3):263–278
- Yao K (2021) An uncertain single-server queueing model. *J Uncertain Syst* 14(1):2150001
- Yao K, Liu B (2018) Uncertain regression analysis: an approach for imprecise observations. *Soft Comput* 22(17):5579–5582
- Zhang C, Liu Z, Liu J (2020) Least absolute deviations for uncertain multivariate regression model. *Int J Gen Syst* 49(4):449–465
- Zhang Y, Gao J, Li X et al (2021) Two-person cooperative uncertain differential game with transferable payoffs. *Fuzzy Optim Decis Making* 20:1–28

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.