An uncertain support vector machine based on soft margin method

Qiqi Li1 · Zhongfeng Qin1,2 · 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 $\alpha$-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 $\alpha$-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
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 queuing 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 $\alpha$-separable data sets with imprecise observations. In this paper, we propose an USVM based on soft margin method for the classification problem with linearly $\alpha$-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 $\alpha$-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 $\sigma$-algebra on a nonempty set $\Gamma$. 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 $\Gamma$; (2) $\mathcal{M}(\Lambda) + \mathcal{M}(\Lambda^c) = 1$ for any $\Lambda \in \mathcal{L}$; (3) For every countable sequence $\Lambda_1, \Lambda_2, \ldots, \mathcal{M}\{\bigcup_{i=1}^{\infty} \Lambda_i\} \leq \sum_{i=1}^{\infty} \mathcal{M}\{\Lambda_i\}$, 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 $k = 1, 2, \ldots$. The product uncertain measure $\mathcal{M}$ is an uncertain measure satisfying

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

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

An uncertain variable (Liu 2007) $\tau$ 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 $\Upsilon(x) = \mathcal{M}\{\tau \leq x\}, x \in \mathfrak{R}$ is called the uncertainty distribution of $\tau$.

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

$$\Psi^{-1}(u) = f(\Phi_1^{-1}(u), \Phi_2^{-1}(u), \ldots, \Phi_n^{-1}(u)), \quad \Phi_m^{-1}(1-u), \ldots, \Phi_n^{-1}(1-u)).$$

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

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

holds if and only if

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

3 Soft margin method for linearly $\alpha$-nonseparable data sets

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

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

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

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

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

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

for all $i = 1, 2, \ldots, l$, then we call the data set $S$ linearly $\alpha$-nonseparable. For instance, Fig. 1 gives an example of a linearly $\alpha$-nonseparable data set. There is no hyperplane that makes Inequality (2) hold. Thus the hard margin method (Qin and Li 2022) cannot be applied in this case.
Suppose that $S$ is linearly $\alpha$-nonseparable. Then, for each observation $\langle \tilde{x}_i, y_i \rangle$, we may seek a nonnegative slack variable $s_i$ such that

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

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 $\alpha$-separable one.

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

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

where

$$ Y^{-1}_{ij}(\alpha, -y_i w_j) = \Phi_{ij}^{-1}(\alpha) \cdot I_{[-y_i w_j \geq 0]} + \Phi_{ij}^{-1}(1 - \alpha) \cdot I_{[-y_i w_j < 0]}. \tag{5} $$

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

$$ M \{ w^T \tilde{x}_i + b \geq -s_i \} $$

$$ = M \{ -w^T \tilde{x}_i - b - s_i \leq 0 \} $$

$$ = M \left\{ -\sum_{j=1}^{n} w_j \tilde{x}_{ij} - b - s_i \leq 0 \right\} \geq \alpha. \tag{6} $$

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

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

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

$$ \sum_{j=1}^{n} w_j Y^{-1}_{ij}(a, -w_j) + b $$

$$ = y_i \left( \sum_{j=1}^{n} w_j Y^{-1}_{ij}(\alpha, -y_i w_j) + b \right) $$

$$ \geq -s_i. $$

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

$$ M \{ w^T \tilde{x}_i + b - s_i \leq 0 \} $$

$$ = M \left\{ \sum_{j=1}^{n} w_j \tilde{x}_{ij} + b - s_i \leq 0 \right\} \geq \alpha. \tag{7} $$

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

$$ F_{2i}^{-1}(u) = \sum_{j=1}^{n} w_j \left( \Phi_{ij}^{-1}(u) \cdot I_{[w_j \geq 0]} + \Phi_{ij}^{-1}(1 - u) \cdot I_{[w_j < 0]} \right) + b - s_i $$

$$ = \sum_{j=1}^{n} w_j Y^{-1}_{ij}(u, w_j) + b - s_i. $$

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

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

which is equivalent to Inequality (4), i.e.,
be the inverse uncertainty distributions of imprecise observations 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

\[
\begin{aligned}
\max_{\mathbf{w}, b, \xi} & \quad \min_i E \left[ \frac{||\mathbf{w}^T \mathbf{x}_i + b||}{||\mathbf{w}||} \right] - C \sum_{i=1}^{l} s_i \\
\text{s.t.} & \quad \mathcal{M} \left\{ y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq -s_i \right\} \geq \alpha, \\
& \quad s_i \geq 0, \quad i = 1, 2, \ldots, l.
\end{aligned}
\]  

(8)

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

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

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

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

(9)

For the linearly \( \alpha \)-separable case, we have \( s_i = 0 \) for \( i = 1, 2, \ldots, l \). 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 \( \tilde{x}_{1i}, \tilde{x}_{2j}, \ldots, \tilde{x}_{mi} \) of each uncertain vectors \( \tilde{x}_i \) are independent for \( i = 1, 2, \ldots, l \). Let \( \Phi_{ij}^{-1} \) be the inverse uncertainty distributions of \( \tilde{x}_{ij} \) for \( i = 1, 2, \ldots, l \) and \( j = 1, 2, \ldots, n \), respectively. Then Model (9) is equivalent to Model (10).

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

(10)
where $Y^{-1}_\alpha(a, -y, w_j)$ is given in Equality (5), and
\[ Y^{-1}_\alpha(u, w_j) = \Phi^{-1}(u) \cdot J_{\{w_j \geq 0\}} + \Phi^{-1}(1 - u) \cdot J_{\{w_j < 0\}}. \]

**Proof** It follows from Theorem 1 that the inverse uncertainty distributions of
\[ w_i \bar{x}_i + b = \sum_{j=1}^{n} w_j x_j i + b \]
for $i = 1, 2, \ldots, l$, respectively. Thus, the objective function becomes
\[ \int_{0}^{1} \left| \sum_{j=1}^{n} w_j Y^{-1}_\alpha(u, w_j) + b \right| \, du - C \sum_{j=1}^{l} s_j. \]
The constraints follow from Theorem 3 immediately. \qed

## 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 $L(a, b)$ is

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

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 $\bar{x}_i = (\bar{x}_{i1}, \bar{x}_{i2})$ with the label $y_i \in \{-1, 1\}, i = 1, 2, \ldots, 60$. Let $\Phi^{-1}_i$ and $\Phi^{-1}_j$ denote the inverse uncertainty distributions of $\bar{x}_{i1}$ and $\bar{x}_{i2}$ for $i = 1, 2, \ldots, 60$, respectively. The detailed data are shown in Table 1.

![Training data and test data in Example 1](image-url)}

Table 1 Data set $S_1$

| $x_{i1}$ | $x_{i2}$ | $y_i$ | $x_{i1}$ | $x_{i2}$ |
|---------|---------|------|---------|---------|
| 1       | 17,21   | -1   | 31      | 26,28   |
| 2       | 7,10    | -1   | 32      | 31,33   |
| 3       | 12,14   | -1   | 33      | 29,31   |
| 4       | 10,11   | -1   | 34      | 25,27   |
| 5       | 21,23   | -1   | 35      | 30,33   |
| 6       | 24,26   | -1   | 36      | 37,39   |
| 7       | 17,19   | -1   | 37      | 20,22   |
| 8       | 18,20   | -1   | 38      | 28,30   |
| 9       | 23,25   | -1   | 39      | 21,23   |
| 10      | 18,20   | -1   | 40      | 30,32   |
| 11      | 15,17   | -1   | 41      | 30,32   |
| 12      | 12,14   | -1   | 42      | 34,36   |
| 13      | 17,19   | -1   | 43      | 24,26   |
| 14      | 12,13   | -1   | 44      | 31,33   |
| 15      | 12,14   | -1   | 45      | 28,30   |
| 16      | 12,13   | -1   | 46      | 36,38   |
| 17      | 14,15   | -1   | 47      | 32,34   |
| 18      | 14,16   | -1   | 48      | 26,28   |
| 19      | 22,24   | -1   | 49      | 26,28   |
| 20      | 9,10    | -1   | 50      | 24,26   |
| 21      | 10,11   | -1   | 51      | 25,26   |
| 22      | 15,16   | -1   | 52      | 34,35   |
| 23      | 18,19   | -1   | 53      | 30,31   |
| 24      | 17,18   | -1   | 54      | 38,39   |
| 25      | 11,12   | -1   | 55      | 34,35   |
| 26      | 13,14   | -1   | 56      | 22,23   |
| 27      | 14,15   | -1   | 57      | 28,29   |
| 28      | 15,16   | -1   | 58      | 26,27   |
| 29      | 20,21   | -1   | 59      | 32,33   |
| 30      | 19,20   | -1   | 60      | 35,36   |

$\Sigma$ Springer
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).

\[
\begin{align*}
\max_{w_1, w_2, b, s_1, i \in I} & \quad \min_{\text{ref}} \int_0^1 \left| w_1 Y_{\alpha^{-1}}(u, w_1) + w_2 Y_{\alpha^{-1}}(u, w_2) + b \right| du - C \sum_{i \in I} s_i \\
\text{s.t.} & \quad y_i \left( w_1 Y_{\alpha^{-1}}(u, -y_iw_1) + w_2 Y_{\alpha^{-1}}(u, -y_iw_2) + b \right) \geq -s_i, \quad i \in I \\
& \quad w_1^2 + w_2^2 = 1 \\
& \quad s_i \geq 0, \quad i \in I.
\end{align*}
\]

(11)

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

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.

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

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

Table 3 Predicting the data label

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

Table 4 Confusion matrix in Example 1

|                  | Positive | Negative | Total |
|------------------|----------|----------|-------|
| Positive         | 9        | 1        | 10    |
| Negative         | 0        | 10       | 10    |

Fig. 3 Misclassified data in $S_1$
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 $\bar{x}_i$ with label $y_i = 1$ are indicated as squares, and the data $\bar{x}_i$ with label $y_i = -1$ are indicated as stars for $i = 1, 2, \ldots, 40$.

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:

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

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},
\]

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},
\]

and

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

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 $S_2$](image)

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) |
When the penalty coefficient $C = 1$ and the confidence level $\alpha = 0.95$, the optimal hyperplane determined by Model (11) for the linearly $\alpha$-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 $\alpha$-separable data set given by Qin and Li (2022). Denote the data set as $S_3$, whose elements are denoted as $(\tilde{x}_i, y_i)$ with label $y_i \in \{1, -1\}$. Table 6 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.
\[ i = 1, 2, \ldots, 40. \] Let \( \Phi^{-1}_{i1} \) and \( \Phi^{-1}_{i2} \) denote the inverse uncertainty distributions of \( \tilde{x}_{i1} \) and \( \tilde{x}_{i2} \) for \( i = 1, 2, \ldots, 40 \), respectively.

To classify the data in \( S_1 \), 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.

\[
\begin{align*}
\max_{w_1, w_2, b, s_1, s_2, \ldots, s_{40}} & \quad \min_i \int_0^1 \left| w_1 Y^{-1}_{i1}(u, w_1) + w_2 Y^{-1}_{i2}(u, w_2) + b \right| du - C \sum_{i=1}^{40} s_i \\
\text{s.t.} & \quad y_i(w_1 Y^{-1}_{i1}(\alpha, -y_i w_1) + w_2 Y^{-1}_{i2}(\alpha, -y_i w_2) + b) \geq -s_i, \quad i = 1, 2, \ldots, 40 \\
& \quad w_1^2 + w_2^2 = 1 \\
& \quad s_i \geq 0, \quad i = 1, 2, \ldots, 40.
\end{align*}
\]

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 \((L(1, 3), L(81, 83))\) and \((L(3, 4), L(55, 58))\), 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 \( \alpha \) 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 \( \alpha \)-separable data sets, but also capable of the case with linearly \( \alpha \)-nonseparable data sets.

5 Conclusions

This paper proposed an USVM based on soft margin method for linearly \( \alpha \)-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 modeling 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

Springer
