

Some Theoretical Results of Learning Theory Based on Intuitionistic Fuzzy Random Samples

Abstract: In this paper, we discuss the statistical learning theory based on intuitionistic fuzzy random samples. Firstly, the definitions of intuitionistic fuzzy number and intuitionistic fuzzy random variable are introduced. Secondly, the concepts and some properties of the mathematical expectation and distance of intuitionistic fuzzy random variables are provided. Thirdly, the definitions of the intuitionistic fuzzy empirical risk functional, the intuitionistic fuzzy expected risk functional, and intuitionistic fuzzy empirical risk minimization principle are proposed. Finally, the key theorem of learning theory based on intuitionistic fuzzy random samples is proved, and the bounds of the rate of convergence of learning process based on intuitionistic fuzzy random samples are discussed.

Keywords: intuitionistic fuzzy numbers, intuitionistic fuzzy random variables, intuitionistic fuzzy empirical risk minimization principle, the key theorem, the bounds of the rate of convergence

1. Introduction

Statistical Learning Theory (SLT, for short) proposed in 1960s and fully established in 1990s by Vapnik et al. [46-48], has emerged as an interesting and sound theory that supports the development of laws of statistical learning for small data samples. It has provided effective solutions obtained in presence of small samples where such samples are inherently associated with crucial issues such as overfitting and underfitting, high-dimensionality of classification problems, existence of multiplicity of local minima and other important problems encountered in practice of machine learning methods and their architectures such as e.g., neural networks. In the late 1990s SLT had become one of the fastest-growing disciplines in machine learning. Its essence was to make the learning machines work effectively with the limited samples and then improve the generalization abilities of the learning machines. By doing this, we establish a meaningful theoretical framework for statistical learning for small data samples. Meanwhile, SLT gave rise to a new category of general learning algorithms, namely Support Vector Machine (SVM, for short). Currently, the SLT and SVM constitute interesting research avenues in machine learning [7,10,11,12,20,28,29,30,31,35,36,39,41,43,44,45,49,50,52,53,56].

The SLT embraces four main components [6], that is: (1) the necessary and sufficient conditions for the consistency of learning processes that minimize the empirical risk – which are represented in the form of the key theorem of the learning theory; (2) the bounds of generalization; (3) structural risk minimization principle based on these bounds; (4) the practical algorithm-Support Vector Machine realizing new structural risk minimization principle. The key theorem of learning theory and the bounds of the rate of convergence of learning process are the most important STL foundations. These two elements form the basis for pursuing further research on the pertinent topics such as e.g., structural risk minimization, support vector machines, and alike.

Though SLT has reached a substantial level of maturity, there are still a number of open issues such as e.g., the development of the SLT and SVM realized on a basis of probability space and the real-valued random samples (real-valued random variables). In real world, there are lots of non-probability spaces (such as Sugeno measure

space [51], credibility measure space [38], etc.) and non real-valued random samples (such as fuzzy samples [55], fuzzy random samples [32], fuzzy complex random samples [26], etc.). To alleviate these problems, it becomes imperative to move forward with extensions and generalizations along the line of dealing with the statistical learning theories established on non-probability space and based on non-real value random samples. Some research has been already realized along this line, cf. [21-27,42]. For example, Ha, Li, Bai, et al. [21-23] generalized the key theorem and the bounds on the rate of convergence of learning process of statistical learning theory from probability space to Sugeno measure space, quasi-probability measure space and credibility space which are three typical non-additive spaces; Ha, Tian, Pedrycz, et al. [24,26] constructed the theoretical foundations of statistical learning theory based on fuzzy random samples and fuzzy complex random samples; Lin and Wang [36] constructed support vector machine based on fuzzy random samples; Tomonori and Shigeo [43] presented some comparative study presenting a comparison between error correcting output codes and fuzzy support vector machines; Liu and Chen [39] discussed face recognition using total margin-based adaptive fuzzy support vector machines.

Intuitionistic (IF, for short) set originated by Atanassov [1,5], is an important concept used to cope with imperfect and/or imprecise information. It is an intuitively straightforward extension of Zadeh's fuzzy sets [55]: while a fuzzy set gives a degree to which an element belongs to a set, an intuitionistic fuzzy set gives both a membership degree and a non-membership degree. The membership and non-membership values induce an indeterminacy index, which models the hesitancy of deciding the degree to which an object satisfies a particular property. As the basis for the study of IF set theory, many operations and relations over IF sets were introduced [1-5]. Many concepts in fuzzy set theory were also extended to IF set theory, such as IF relations, intuitionistic L-fuzzy sets, IF implication, IF logics, and the degree of similarity between IF sets, etc., [4,9,13,15,16,33,34,57]. For a further study for the structure of IF sets, construction theorems of IF sets, IF topology and the axiomatic characterization of IF sets have been investigated [8,14,18,40]. Recently, IF set theory has been successfully applied in decision analysis and pattern recognition (see, e.g., [19,37,54]).

However, the study for the combination of IF set theory and statistical learning theory is still blank. This paper discusses the statistical learning theory based on intuitionistic fuzzy random samples by combining intuitionistic fuzzy analysis and SLT, revisits the major parts of the SLT and establishes some ground material for further development of classifiers such as support vector machines. This paper is organized as follows: Section 2 introduces some basic definitions, elaborates on a number of properties of intuitionistic fuzzy random variables. In Section 3, we prove the key theorem of learning theory based on intuitionistic fuzzy random samples. In the sequel, in Section 4, the bounds of the rate of uniform convergence of learning process based on intuitionistic fuzzy random samples are discussed. The final section offers the conclusions and brings prospects of potential future developments.

2. Preliminaries

Throughout this paper, we let $(\Omega, \mathcal{A}, P)$ be a probability measure space and $\mathbb{R}$ be the real numbers field.

Definition 2.1 [32]. A fuzzy number is a fuzzy set $X : \mathbb{R} \rightarrow [0,1]$ with the following properties:

- (1) X is normal, i.e., there exists $x \in \mathbb{R}$ such that $X(x) = 1$;
- (2) X is upper semi-continuous;
- (3) $\text{supp } X = \text{cl}\{x \in \mathbb{R} : X(x) > 0\}$ is compact;

(4) X is a convex fuzzy set, i.e., $X(\lambda x + (1-\lambda)y) \geq \min(X(x), X(y))$, for $\forall x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$.

Let $F(\mathbb{R})$ be the family of all fuzzy numbers on $\mathbb{R}$. For a fuzzy set X , if we define

$$(X)_r = \begin{cases} \{x: X(x) \geq r\}, & 0 < r \leq 1 \\ \text{supp } X, & r = 0 \end{cases},$$

then it follows that X is a fuzzy number if and only if $(X)_1 \neq \emptyset$ and $(X)_r$ is a closed bounded interval for each $r \in [0, 1]$. Therefore, a fuzzy number X is completely determined by the intervals $(X)_r$.

Let $X, Y \in F(\mathbb{R})$, $\forall r \in [0, 1]$, $(X)_r = [(X)_r^-, (X)_r^+]$ and $(Y)_r = [(Y)_r^-, (Y)_r^+]$. Define

$$X \pm Y = \bigcup_{r \in [0, 1]} \left\{ [(X)_r^- \pm (Y)_r^-, (X)_r^+ \pm (Y)_r^+] \right\}.$$

Let us introduce a partially ordered relation on $F(\mathbb{R})$ as follow: for any $X, Y \in F(\mathbb{R})$, we say $X \leq Y$, iff $\forall r \in [0, 1]$, $(X)_r \leq (Y)_r$ i.e., $(X)_r^- \leq (Y)_r^-$ and $(X)_r^+ \leq (Y)_r^+$.

Definition 2.2 [32]. A fuzzy number valued function $\xi: \Omega \rightarrow F(\mathbb{R})$ is called fuzzy random variable if for every closed subset B of $\mathbb{R}$, the fuzzy set $\xi^{-1}(B)$ is measurable when considered as a function from Ω to $[0, 1]$, where $\xi^{-1}(B)$ denotes the fuzzy subset of Ω defined by $\xi^{-1}(B)(\omega) = \sup_{x \in B} \xi(\omega)(x)$ for every $\omega \in \Omega$.

Definition 2.3 [32]. A fuzzy random variable $\xi(\omega) = \{[(\xi_r^-(\omega), \xi_r^+(\omega))] | 0 < r \leq 1\}$ is called integrable if for each $r \in [0, 1]$, $\xi_r^-(\omega)$ and $\xi_r^+(\omega)$ are integrable, or equivalently $\int \|\xi\| dP < \infty$, where $\|\xi\| = \max\{|\xi_0^-|, |\xi_0^+|\}$. In this case, the expected value of ξ is defined in the following manner

$$E\xi = \int \xi dP = \left\{ \left[\int \xi_r^- dP, \int \xi_r^+ dP \right] | 0 < r \leq 1 \right\}.$$

Definition 2.4 [1]. Let U be a nonempty set. An intuitionistic fuzzy set A in U is an object having the form

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle | x \in U \},$$

where $\mu_A: U \rightarrow [0, 1]$ and $\nu_A: U \rightarrow [0, 1]$ satisfy $0 \leq \mu_A + \nu_A \leq 1$ for all $x \in U$, and $\mu_A(x)$ and $\nu(x)$ are, respectively, called the degree of membership and the degree of non-membership of the element $x \in U$ to A . The complement of an IF set A is denoted by $A^c = \{ \langle x, \nu(x), \mu_A(x) \rangle | x \in U \}$.

Obviously, every fuzzy set $A = \{ \langle x, A(x) \rangle | x \in U \} = \{ \langle x, \mu_A(x) \rangle | x \in U \}$ can be identified with the IF set of the form $\{ \langle x, \mu_A(x), 1 - \mu_A(x) \rangle | x \in U \}$ and is thus an IF set.

We introduce some basic operations about IF sets as follows [1–5,17,33]:

Let A and B be two IF sets, then

$$A \subseteq B \text{ iff } \mu_A(x) \leq \mu_B(x) \text{ and } \nu_A(x) \leq \nu_B(x) \text{ for all } x \in U,$$

$$A \supseteq B \text{ iff } B \subseteq A,$$

$$A = B \text{ iff } A \subseteq B \text{ and } B \subseteq A, \text{ i.e., } \mu_A(x) = \mu_B(x) \text{ and } \nu_A(x) = \nu_B(x) \text{ for any } x \in U,$$

$$A \cap B = \left\{ \langle x, \mu_A(x) \wedge \mu_B(x), \nu_A(x) \vee \nu_B(x) \rangle \mid x \in U \right\},$$

$$A \cup B = \left\{ \langle x, \mu_A(x) \vee \mu_B(x), \nu_A(x) \wedge \nu_B(x) \rangle \mid x \in U \right\},$$

$$\forall \alpha, \beta \in [0,1], \alpha + \beta \leq 1, (\alpha, \beta)A = \left\{ \langle x, \alpha \wedge \mu_A(x), \beta \vee \nu_A(x) \rangle \mid x \in U \right\}.$$

Definition 2.5. Let A be an IF set and $\alpha, \beta \in [0,1]$ with $\alpha + \beta \leq 1$, the (α, β) -level cut set of A , denoted by A_{α}^{β} , is defined as follows:

$$A_{\alpha}^{\beta} = \{x \in U \mid \mu_A(x) \geq \alpha, \nu_A(x) \leq \beta\}$$

$A_{\alpha} = \{x \in U \mid \mu_A(x) \geq \alpha\}$ and $A_{\alpha+} = \{x \in U \mid \mu_A(x) > \alpha\}$ are, respectively, called the α -level cut set and the strong α -level cut set of membership generated by A . And $A^{\beta} = \{x \in U \mid \nu_A(x) \leq \beta\}$ and $A^{\beta+} = \{x \in U \mid \nu_A(x) < \beta\}$ are, respectively, referred to as the β -level cut set and the strong β -level cut set of non-membership generated by A .

Proposition 2.1. Let $A \in IF(U)$, $\alpha, \beta \in [0,1]$ and $\alpha + \beta \leq 1$. Then

$$A = \bigcup (\alpha, \beta) A_{\alpha}^{\beta} = \bigcup (\alpha, \beta) A_{\alpha+}^{\beta} = \bigcup (\alpha, \beta) A_{\alpha}^{\beta+} = \bigcup (\alpha, \beta) A_{\alpha+}^{\beta+}.$$

Proof. We have only to prove that the equation $A = \bigcup (\alpha, \beta) A_{\alpha}^{\beta}$ holds for any $x \in U$. The other equations can be proved in a similar way. On one hand, we have

$$\begin{aligned} \mu_{\bigcup (\alpha, \beta) A_{\alpha}^{\beta}}(x) &= \vee \mu_{(\alpha, \beta) A_{\alpha}^{\beta}}(x) = \vee (\alpha \wedge A_{\alpha}^{\beta}(x)) \\ &= \left(\bigvee_{0 \leq \alpha \leq \mu_A(x)} (\alpha \wedge A_{\alpha}^{\beta}(x)) \right) \vee \left(\bigvee_{\mu_A(x) < \alpha \leq 1} (\alpha \wedge A_{\alpha}^{\beta}(x)) \right) \\ &= \bigvee_{0 \leq \alpha \leq \mu_A(x)} \alpha = \mu_A(x). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} \nu_{\bigcup (\alpha, \beta) A_{\alpha}^{\beta}}(x) &= \wedge \nu_{(\alpha, \beta) A_{\alpha}^{\beta}}(x) = \wedge (\beta \vee (1 - A_{\alpha}^{\beta}(x))) \\ &= \left(\bigwedge_{0 \leq \beta \leq \nu_A(x)} (\beta \vee (1 - A_{\alpha}^{\beta}(x))) \right) \wedge \left(\bigwedge_{\nu_A(x) < \beta \leq 1} (\beta \vee (1 - A_{\alpha}^{\beta}(x))) \right) \\ &= \bigwedge_{\nu_A(x) < \beta \leq 1} \beta = \nu_A(x). \end{aligned}$$

Thus $A = \bigcup (\alpha, \beta) A_{\alpha}^{\beta}$.

Definition 2.6. The IF set $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in \mathbb{R} \}$ is called an IF number if and only if μ_A and ν_A^c are two fuzzy numbers, where $\nu_A^c = 1 - \nu_A$.

The family of all intuitionistic fuzzy numbers is denoted by $IF(\mathbb{R})$.

Proposition 2.2. Let $A \in IF(\mathbb{R})$. Then the membership function $\mu_A(x)$ and non-membership function $\nu_A(x)$ have the following properties:

(1) $\mu_A : \mathbb{R} \rightarrow [0,1]$ is an upper semi-continuous function and $\mu_A : \mathbb{R} \rightarrow [0,1]$ is a lower semi-continuous function;

(2) $\mu_A(x) = 0, \forall x \in (-\infty, c] \cup [d, +\infty), \nu_A(x) = 1, \forall x \in (-\infty, e] \cup [f, +\infty);$

(3) $\mu_A(x) = 1, \nu_A(x) = 0, \forall x \in [a, b];$

(4) $\mu_A(x)$ is strictly monotonously increasing on $[c, a]$ and strictly monotonously decreasing on $[b, d];$

$\nu_A(x)$ is strictly monotonously decreasing on $[e, a]$ and strictly monotonously increasing on $[b, f].$

Definition 2.7. Some Operations on $IF(\mathbb{R})$ are defined as follows:

$\forall A, B \in IF(\mathbb{R}),$

$$A + B = \left\{ \left\langle z, \sup_{x+y=z} (\mu_A(x) \wedge \mu_B(y)), \inf_{x+y=z} (\nu_A(x) \vee \nu_B(y)) \right\rangle \mid x, y \in U \right\};$$

$$\rho A = \left\{ \left\langle z, \mu_A\left(\frac{z}{\rho}\right), \nu_A\left(\frac{z}{\rho}\right) \right\rangle \mid z = \rho x, x \in U \right\}, \quad \rho \neq 0;$$

If $\rho = 0$, then

$$\mu_{0A}(z) = \begin{cases} \sup \{ \mu_A(x) \mid x \in U \}, & z = 0 \\ 0, & z \neq 0 \end{cases}, \quad \nu_{0A}(z) = \begin{cases} \inf \{ \nu_A(x) \mid x \in U \}, & z = 0 \\ 1, & z \neq 0 \end{cases}.$$

Proposition 2.3. Let $A, B \in IF(\mathbb{R})$. And (α, β) -level cut set of A and B are denoted by

$$A_\alpha^\beta = [\mu_A^L, \mu_A^R, \nu_A^L, \nu_A^R, (\alpha, \beta)], B_\alpha^\beta = [\mu_B^L, \mu_B^R, \nu_B^L, \nu_B^R, (\alpha, \beta)]. \text{ Then we have}$$

$$A + B = \left\{ [\mu_A^L + \mu_B^L, \mu_A^R + \mu_B^R, \nu_A^L + \nu_B^L, \nu_A^R + \nu_B^R, (\alpha, \beta)] \mid \alpha, \beta \in [0,1], \alpha + \beta \leq 1 \right\},$$

$$\rho A = \begin{cases} \left\{ [\rho \mu_A^L, \rho \mu_A^R, \rho \nu_A^L, \rho \nu_A^R, (\alpha, \beta)] \mid \alpha, \beta \in [0,1], \alpha + \beta \leq 1 \right\}, & \text{if } \rho \geq 0 \\ \left\{ [\rho \mu_A^R, \rho \mu_A^L, \rho \nu_A^R, \rho \nu_A^L, (\alpha, \beta)] \mid \alpha, \beta \in [0,1], \alpha + \beta \leq 1 \right\}, & \text{if } \rho < 0 \end{cases}.$$

Definition 2.8. Let $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in U \}, B = \{ \langle x, \mu_B(x), \nu_B(x) \rangle \mid x \in U \} \in IF(\mathbb{R})$. Define $A \leq B$ if

$\mu_A \leq \mu_B$ and $\nu_A^c \leq \nu_B^c$; $A = B$ if $A \leq B$ and $B \leq A$; $A < B$ if $A \leq B$ and $A \neq B$.

Obviously, $(IF(\mathbb{R}), \leq)$ is a partially ordered set.

Definition 2.9. Let $S \subseteq IF(\mathbb{R})$ and $M \in IF(\mathbb{R})$. We say that M is the supremum of the set S if the following two conditions are satisfied:

- (1) for any $A \in S$, $A \leq M$, i.e., M is the upper bound of the set S ;
- (2) for any the upper bound N of S , $M \leq N$.

Similarly, we can define the infimum of the set S .

Let $\mathcal{K}(\mathbb{R})$ be a family of all the nonempty compact convex subsets of $\mathbb{R}$. If $A, B \in \mathcal{K}(\mathbb{R})$, then the Hausdorff metric is defined by

$$d_H(A, B) = \max \left\{ \sup_{x \in A} \inf_{y \in B} d(x, y), \sup_{y \in B} \inf_{x \in A} d(x, y) \right\},$$

where $d(x, y)$ denotes the distance between two real numbers x and y .

Let us define a consistent Hausdorff metric in $IF(\mathbb{R})$ to be in the following form

$$\begin{aligned} d(A, B) &= \sup_{\substack{\alpha, \beta \in [0,1] \\ \alpha + \beta \leq 1}} \max \{ d_H(A_\alpha, B_\alpha), d_H(A^\beta, B^\beta) \} \\ &= \max \{ d(\mu_A, \mu_B), d(\nu_A, \nu_B) \} \\ &= \max \{ d(\mu_A, \mu_B), d(\nu_A^c, \nu_B^c) \} \end{aligned}$$

where $A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle \mid x \in U \}$, $B = \{ \langle x, \mu_B(x), \nu_B(x) \rangle \mid x \in U \} \in IF(\mathbb{R})$.

Proposition 2.4. If $A, B, C \in IF(\mathbb{R})$, then $d(A, B) + d(B, C) \geq d(A, C)$.

Proposition 2.5. If $A, B, C \in IF(\mathbb{R})$ and $A \geq B \geq C$, then $d(A, C) \geq d(A, B)$ and $d(A, C) \geq d(B, C)$.

Definition 2.10. Let $(\Omega, \mathcal{A}, P)$ be a probability measure space. An intuitionistic fuzzy number valued mapping

$$\xi : \Omega \rightarrow IF(\mathbb{R}),$$

$$\omega \rightarrow \xi(\omega) = \{ \langle x, \mu_{\xi(\omega)}(x), \nu_{\xi(\omega)}(x) \rangle \mid x \in \mathbb{R} \}$$

is called an intuitionistic fuzzy random variable if μ_ξ and ν_ξ^c are two fuzzy random variables defined on

$(\Omega, \mathcal{A}, P)$, where $\mu_\xi : \omega \rightarrow \mu_{\xi(\omega)}$, $\nu_\xi : \omega \rightarrow \nu_{\xi(\omega)}$ and $\nu_\xi^c = 1 - \nu_\xi$.

Definition 2.11 Let $\xi(\omega) = \{ \langle x, \mu_{\xi(\omega)}(x), \nu_{\xi(\omega)}(x) \rangle \mid x \in \mathbb{R} \}$ be the intuitionistic fuzzy random variable defined

on $(\Omega, \mathcal{A})$. We say that

$$F(Z) = P\{\xi(\omega) \leq Z\} = P\{\mu_{\xi(\omega)} \leq \mu_Z, \nu_{\xi(\omega)} \geq \nu_Z\}$$

is the distribution function of ξ , where $Z = \{\langle x, \mu_Z(x), \nu_Z(x) \rangle | x \in \mathbb{R}\} \in IF(\mathbb{R})$.

Definition 2.12. An intuitionistic fuzzy random variable $\xi(\omega) = \{\langle x, \mu_{\xi(\omega)}(x), \nu_{\xi(\omega)}(x) \rangle | x \in \mathbb{R}\}$ is called

integrable if μ_ξ and ν_ξ^c are integrable. In this case, the mathematical expectation of ξ is defined in the following manner

$$E(\xi) = \int \xi dP = \left\{ \langle x, E(\mu_\xi)(x), E(\nu_\xi)(x) \rangle | x \in \mathbb{R} \right\} = \left\{ \left[\int \mu_\xi^L dP, \int \mu_\xi^R dP, \int \nu_\xi^L dP, \int \nu_\xi^R dP, (\alpha, \beta) \right] | \alpha, \beta \in [0, 1], \alpha + \beta \leq 1 \right\},$$

where $E(\nu_\xi) = 1 - E(\nu_\xi^c)$.

Proposition 2.6. If ξ is an intuitionistic fuzzy random variable, then the following equalities

$$1. \{E(\xi)\}_\alpha^\beta = E(\xi_\alpha^\beta) = E(\xi_\alpha) \cap E(\xi^\beta) \text{ for } 0 < \alpha < \beta \leq 1;$$

$$2. E(c\xi) = cE\xi, \text{ whenever } c \in \mathbb{R};$$

$$3. E(\xi_1 \pm \xi_2) = E\xi_1 \pm E\xi_2$$

hold true.

Definition 2.13. Suppose that $\xi_n = \{\langle x, \mu_{\xi_n}(x), \nu_{\xi_n}(x) \rangle | x \in U\}, n = 1, 2, \dots$ is a sequence of intuitionistic fuzzy random variables.

(1) If the fuzzy random variables sequence $\mu_{\xi_n}, n = 1, 2, \dots$ and $\nu_{\xi_n}^c, n = 1, 2, \dots$ are respectively mutually independent, then we say that $\xi_n, n = 1, 2, \dots$ is a sequence of mutually independent intuitionistic fuzzy random variables.

(2) If $\mu_{\xi_n}, n = 1, 2, \dots$ is a sequence of identically distributed fuzzy random vectors and $\nu_{\xi_n}^c, n = 1, 2, \dots$ is also a sequence of identically distributed fuzzy random vectors, then we say that $\xi_n, n = 1, 2, \dots$ is a sequence of identically distributed intuitionistic fuzzy random variables.

Definition 2.14. Suppose that $\xi_n, n = 1, 2, \dots$ is a sequence of intuitionistic fuzzy random variables and ξ is an intuitionistic fuzzy random variable (or an intuitionistic fuzzy number). If $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\{\omega | d(\xi_n(\omega), \xi(\omega)) > \varepsilon\} = 0,$$

then we say that ξ_n converge in probability P to ξ , denoted by

$$\xi_n \xrightarrow{P} \xi, n \rightarrow \infty, \text{ or } \lim_{n \rightarrow \infty} \xi_n = \xi.$$

Theorem 2.1. Let $\xi_n = \{\langle x, \mu_{\xi_n}(x), \nu_{\xi_n}(x) \rangle | x \in U\}, n = 1, 2, \dots$ be a sequence of intuitionistic fuzzy random variables

and $\xi = \{\langle x, \mu_\xi(x), \nu_\xi(x) \rangle | x \in U\}$ be an intuitionistic fuzzy random variable. Then $\xi_n \xrightarrow{P} \xi$ if and only if

$$\mu_{\xi_n} \xrightarrow{P} \mu_\xi \text{ and } \nu_{\xi_n} \xrightarrow{P} \nu_\xi.$$

Proof. By the definition of Hausdorff metric in $IF(\mathbb{R})$, we have

$$d(\mu_{\xi_n}, \mu_{\xi}) \leq d(\xi_n, \xi), d(v_{\xi_n}, v_{\xi}) \leq d(\xi_n, \xi), d(\xi_n, \xi) \leq d(\mu_{\xi_n}, \mu_{\xi}) + d(v_{\xi_n}, v_{\xi}).$$

Hence,

$$P\{d(\xi_n, \xi) > \varepsilon\} \leq P\{d(\mu_{\xi_n}, \mu_{\xi}) + d(v_{\xi_n}, v_{\xi}) > \varepsilon\} \leq P\left\{d(\mu_{\xi_n}, \mu_{\xi}) > \frac{\varepsilon}{2}\right\} + P\left\{d(v_{\xi_n}, v_{\xi}) > \frac{\varepsilon}{2}\right\}.$$

The theorem is easily proven.

Theorem 2.2 (The strong law of large numbers of fuzzy random variables) [32]. Let $\{\xi_n, n = 1, 2, \dots\}$ be a sequence

of independent and identically distributed fuzzy random variables with $E\|\xi_1\| < \infty$. Then we have

$$\frac{1}{n} \sum_{j=1}^n \xi_j \xrightarrow{P} E\xi_1, n \rightarrow \infty.$$

Theorem 2.3 (The strong law of large numbers of intuitionistic fuzzy random variables). Let $\{\xi_n, n = 1, 2, \dots\}$ be a sequence of independent and identically distributed intuitionistic fuzzy random variables with $E\|\mu_{\xi_1}\| < \infty$ and $E\|v_{\xi_1}\| < \infty$. Then

$$\frac{1}{n} \sum_{j=1}^n \xi_j \xrightarrow{P} E\xi_1, n \rightarrow \infty.$$

Proof. It follows from Theorem 2.2 and Proposition 2.3.

3. The key theorem of learning theory based on intuitionistic fuzzy random samples

The key theorem forms the important foundation of SLT. It relaxes the requirement of the strict consistency of the empirical risk minimization method with the one of existence of uniform one-sided convergence of means to their mathematical expectations or, in other words, with convergence of some empirical process. In this section, we introduce and prove the key theory of the learning theory based on intuitionistic fuzzy random variables. First, we introduce some basic concepts.

Let $\xi_j, j = 1, 2, \dots, l$ be a sequence of independent and identically distributed intuitionistic fuzzy random samples

whose distribution is given as $F(\xi)$.

Definition 3.1. We call $R_{if}(\alpha) = E[Q(\xi, \alpha)] = \int Q(\xi, \alpha) dP, \alpha \in \Lambda$, where Λ is an index set and $Q(\xi, \alpha)$ denotes the loss functional, the expected risk functional based on intuitionistic fuzzy random samples. It could be referred to as the intuitionistic fuzzy expected risk functional.

Definition 3.2. We call the expression $R_{ifemp}(\alpha) = \frac{1}{l} \sum_{j=1}^l Q(\xi_j, \alpha), \alpha \in \Lambda$ the intuitionistic fuzzy empirical risk functional.

Let the minimum of the risk functional be obtained at $Q(\xi, \alpha_0)$ and the minimum of the empirical risk functional

be obtained at $Q(\xi, \alpha_l)$.

Definition 3.3. We minimize $R_{ifemp}(\alpha)$ instead of minimizing $R_{if}(\alpha)$ and consider the function $Q(\xi, \alpha_l)$ as an approximation to the function $Q(\xi, \alpha_0)$. This principle is called the intuitionistic fuzzy empirical risk minimization principle of statistical learning theory. (IFERM)

Definition 3.4. We say that the intuitionistic fuzzy empirical risk minimization principle is consistent for the set of functions $Q(\xi, \alpha), \alpha \in \Lambda$ and for the distribution functional $F(\xi)$ if the following two sequences converge in probability to the same limit:

$$R_{if}(\alpha_l) \xrightarrow[l \rightarrow \infty]{P} \inf_{\alpha \in \Lambda} R_{if}(\alpha),$$

$$R_{ifemp}(\alpha_l) \xrightarrow[l \rightarrow \infty]{P} \inf_{\alpha \in \Lambda} R_{if}(\alpha).$$

Definition 3.5. We say that the intuitionistic fuzzy empirical risk minimization principle is strictly consistent for the set of function $Q(\xi, \alpha), \alpha \in \Lambda$ and the distribution function $F(\xi)$ if for any nonempty subset $\Lambda(C), C \in IF(\mathbb{R})$ of this set of functions such that $\Lambda(C) = \left\{ \alpha : \int Q(\xi, \alpha) dP \geq C \right\}$ the convergence

$$\inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha) \xrightarrow[l \rightarrow \infty]{P} \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)$$

holds.

Remark 3.1. We only consider the sequence of intuitionistic fuzzy random variables that are partially ordered " $\leq$ ".

We denote by $d\left(\left(R_{if}(\alpha), R_{ifemp}(\alpha)\right)_{\geq}\right)$ the Hausdorff metric when $R_{ifemp}(\alpha) \leq R_{if}(\alpha)$.

Theorem 3.1 (The key theorem of learning theory based on intuitionistic fuzzy random samples). Suppose that there exist two intuitionistic fuzzy numbers W_1 and W_2 , such that for all functions in the set $Q(\xi, \alpha), \alpha \in \Lambda$, and for a given distributed function $F(\xi)$, the inequalities $W_1 \leq \int Q(\xi, \alpha) dP \leq W_2$ hold true. Then the sufficient and necessary conditions for the strict consistency of IFERM principle is that the convergence

$$P\left\{\sup_{\alpha \in \Lambda} d\left(\left(R_{if}(\alpha), R_{ifemp}(\alpha)\right)_{\geq}\right) > \varepsilon\right\} \xrightarrow[l \rightarrow \infty]{} 0$$

holds for any $\varepsilon > 0$.

Proof. Necessary: By Definition 3.5 for the set $\Lambda(C) = \left\{ \alpha : R_{if}(\alpha) \geq C \right\}$, we have

$$\inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha) \xrightarrow[l \rightarrow \infty]{P} \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha) \quad (3.1)$$

We denote by A the event of the form $\sup_{\alpha \in \Lambda} d\left(\left(R_{if}(\alpha), R_{ifemp}(\alpha)\right)_{\geq}\right) > \varepsilon$.

Suppose that A holds. Then there exists $\alpha^* \in \Lambda(C)$ such that $d\left(\left(R_{if}(\alpha^*), R_{ifemp}(\alpha^*)\right)_{\geq}\right) > \varepsilon$. We can

find $W_1 \leq a_k \leq W_2$, such that $\alpha^* \in \Lambda(a_k)$ and $d\left(\left(R_{if}(\alpha^*), a_k\right)_{\geq}, \frac{\varepsilon}{2}\right) < \frac{\varepsilon}{2}$. Then for these $\Lambda(a_k)$ the inequalities

$$d\left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha)\right)_{\geq}\right) < \frac{\varepsilon}{2} \text{ and } R_{ifemp}(\alpha^*) \geq \inf_{\alpha \in \Lambda(a_k)} R_{ifemp}(\alpha)$$

hold true.

Therefore

$$d\left(\left(\inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(a_k)} R_{ifemp}(\alpha)\right)_{\geq}\right) \geq d\left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(a_k)} R_{ifemp}(\alpha)\right)_{\geq}\right) - d\left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha)\right)_{\geq}\right) > \frac{\varepsilon}{2}.$$

According to (3.1), we have

$$P\left\{d\left(\left(\inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(a_k)} R_{ifemp}(\alpha)\right)_{\geq}\right) > \frac{\varepsilon}{2}\right\} \xrightarrow{l \rightarrow \infty} 0.$$

Denote by T_k the event $d\left(\left(\inf_{\alpha \in \Lambda(a_k)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(a_k)} R_{ifemp}(\alpha)\right)_{\geq}\right) > \frac{\varepsilon}{2}$, then $A \subseteq \bigcup_k T_k$,

$$P(A) \leq P\left(\bigcup_k T_k\right) \xrightarrow{l \rightarrow \infty} 0.$$

Hence

$$P\left\{\sup_{\alpha \in \Lambda} d\left(\left(R_{if}(\alpha), R_{ifemp}(\alpha)\right)_{\geq}\right) > \varepsilon\right\} \xrightarrow{l \rightarrow \infty} 0 \quad (3.2)$$

Sufficiency: We denote by A the event $\left\{\omega : d\left(\inf_{\alpha \in \Lambda(C)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha)\right) > \varepsilon\right\}$

$$A_1 = \left\{\omega : d\left(\left(\inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)\right)_{\geq}\right) > \varepsilon\right\}$$

$$A_2 = \left\{\omega : d\left(\left(\inf_{\alpha \in \Lambda(C)} R_{if}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha)\right)_{\geq}\right) > \varepsilon\right\}.$$

Then $A = A_1 \cup A_2$, and $P(A) \leq P(A_1) + P(A_2)$.

Suppose that the event A_1 occurs, we can find $Q(\xi, \alpha^*), \alpha^* \in \Lambda(C)$, such that

$$d\left(\left(R_{if}(\alpha^*), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)\right)_{\geq}\right) < \frac{\varepsilon}{2} \text{ and } R_{ifemp}(\alpha^*) \geq \inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha).$$

By $d\left(\left(\inf_{\alpha \in \Lambda(C)} R_{ifemp}(\alpha), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)\right)_{\geq}\right) > \varepsilon$, we have

$$d\left(\left(R_{ifemp}(\alpha^*), \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha)\right)_{\geq}\right) > \varepsilon.$$

Therefore

$$d\left(\left(R_{ifemp}(\alpha^*), R_{if}(\alpha^*)\right)_{\geq}\right) > \frac{\varepsilon}{2}.$$

In virtue of the monotonicity of probability and the strong law of large numbers of intuitionistic fuzzy random variables, we obtain

$$P(A_1) \leq P\left\{d\left(\left(R_{ifemp}(\alpha^*), R_{if}(\alpha^*)\right)_{\geq}\right) > \frac{\varepsilon}{2}\right\} \xrightarrow{l \rightarrow \infty} 0.$$

On the other hand, if the event A_2 takes place, we can find $Q(\xi, \alpha^{**}), \alpha^{**} \in \Lambda(C)$, such that

$$d\left(\left(\inf_{\alpha \in \Lambda(C)} R_{if}(\alpha), R_{ifemp}(\alpha^{**})\right)_{\geq}\right) > \frac{\varepsilon}{2} \text{ and } R_{if}(\alpha^{**}) \geq \inf_{\alpha \in \Lambda(C)} R_{if}(\alpha).$$

Then

$$d\left(\left(R_{if}(\alpha^{**}), R_{ifemp}(\alpha^{**})\right)_{\geq}\right) > \frac{\varepsilon}{2}.$$

Therefore

$$\begin{aligned} P(A_2) &\leq P\left\{d\left(\left(R_{if}(\alpha^{**}), R_{ifemp}(\alpha^{**})\right)_{\geq}\right) > \frac{\varepsilon}{2}\right\} \\ &\leq P\left\{\sup_{\alpha \in \Lambda} d\left(\left(R_{if}(\alpha), R_{ifemp}(\alpha)\right)_{\geq}\right) > \frac{\varepsilon}{2}\right\} \xrightarrow{l \rightarrow \infty} 0. \end{aligned}$$

Hence

$$P(A) \leq P(A_1) + P(A_2) \xrightarrow{l \rightarrow \infty} 0.$$

The theorem has been proved.

The samples become fuzzy random samples when $\nu_{\xi} = 1 - \mu_{\xi}$. Then we have

Corollary 3.1 [26]. Let the set of functions $Q(\xi, \alpha), \alpha \in \Lambda$, satisfies $X \leq R(\alpha) \leq Y$. Then the sufficient and necessary condition for the strict consistency of FERM principle is the convergence

$$\lim_{l \rightarrow \infty} P\left\{\sup_{\alpha \in \Lambda} D\left(\left(R_f(\alpha), R_{femp}(\alpha)\right)_{\geq}\right) > \varepsilon\right\} = 0$$

is valid.

The samples become random samples in the usual sense when the samples are not represented as fuzzy sets, and let d be the subtraction of real numbers. Then we have

Corollary 3.2 [47]. Assume that there exist the constants a and A such that for all functions in the set $Q(\xi, \alpha), \alpha \in \Lambda$, and for a given distribution function $F(\xi)$, the inequalities

$$a \leq R(\alpha) \leq A, \alpha \in \Lambda$$

hold true. Then the following two statements are equivalent:

1. For the given distribution function $F(\xi)$, the empirical risk minimization method is strictly consistent on the set of functions $Q(\xi, \alpha), \alpha \in \Lambda$.

2. For the given distribution function $F(\xi)$, the uniform one-sided convergence of the means to their mathematical expectation takes place over the set of functions $Q(\xi, \alpha), \alpha \in \Lambda$.

4. Bounds of the rate of uniform convergence of learning process based on intuitionistic fuzzy random samples

In the SLT, the important conclusions about the relation between the empirical risk and practical risk come in the form of some bounds of generalization. They are essential in the analysis of capacity of the learning machines and when developing new learning algorithms. The bounds on the rates of uniform convergence of learning process are important components of the bounds of generalization. In this section, we discuss the bounds of the

rate of uniform convergence of learning process based on intuitionistic fuzzy random samples. We consider the model where a set of intuitionistic fuzzy measurable functions $Q(\xi, \alpha), \alpha \in \Lambda$ contains a finite number N of elements $Q(\xi, \alpha_k), k = 1, 2, \dots, N$. At first, we introduce a basic theorem:

Theorem 4.1[25,26]. Let $\xi_j = \left\{ \left[\left(\xi_j \right)_r^-, \left(\xi_j \right)_r^+ \right] \mid r \in [0, 1] \right\}, j = 1, 2, \dots, n$ be a sequence of fuzzy random variables, and $\left(\xi_j \right)_r^-, \left(\xi_j \right)_r^+$ have the limited variance and the same upper bounds M . Then

$$P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \xi_j, E \xi_j \right) \geq \varepsilon \right\} \leq \frac{2M}{n\varepsilon^2}.$$

Theorem 4.2. Let $\xi_j = \left\{ \left\langle x, \mu_{\xi_j}(x), \nu_{\xi_j}(x) \right\rangle \mid x \in U \right\}, j = 1, 2, \dots, n$ be a sequence of intuitionistic fuzzy random variables, $\left(\mu_{\xi_j} \right)_r^-, \left(\mu_{\xi_j} \right)_r^+$ and $\left(\nu_{\xi_j} \right)_r^-, \left(\nu_{\xi_j} \right)_r^+$ have the limited variance and the same upper bounds M_1 and M_2 , respectively. Then

$$P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \xi_j, E \xi_j \right) \geq \varepsilon \right\} \leq \frac{8(M_1 + M_2)}{n\varepsilon^2}.$$

Proof. According to Theorem 4.1, we have

$$\begin{aligned} P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \xi_j, E \xi_j \right) \geq \varepsilon \right\} &= P \left\{ \max \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right), d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \right\} \geq \varepsilon \right\} \\ &\leq P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) + d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \varepsilon \right\} \\ &\leq P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \mu_{\xi_j}, E \mu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} + P \left\{ d \left(\frac{1}{n} \sum_{j=1}^n \nu_{\xi_j}, E \nu_{\xi_j} \right) \geq \frac{\varepsilon}{2} \right\} \\ &\leq \frac{8M_1}{n\varepsilon^2} + \frac{8M_2}{n\varepsilon^2} \\ &= \frac{8(M_1 + M_2)}{n\varepsilon^2}. \end{aligned}$$

Theorem 4.3. Suppose that $Q(\xi, \alpha_k), k = 1, 2, \dots, N$ is a set of the functions, $E[Q(\xi, \alpha)]$ exists and satisfies the conditions of Theorem 4.2. Then the following inequality

$$d \left(\left(R_{if}(\alpha_l), R_{ifemp}(\alpha_l) \right)_{\geq} \right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}}$$

holds with probability of at least $1 - \eta$.

Proof.

$$\begin{aligned} P \left\{ \sup_{1 \leq k \leq N} d \left(\left(R_{if}(\alpha_k), R_{ifemp}(\alpha_k) \right)_{\geq} \right) > \varepsilon \right\} &\leq \sum_{k=1}^N P \left\{ d \left(\left(R_{if}(\alpha_k), R_{ifemp}(\alpha_k) \right)_{\geq} \right) > \varepsilon \right\} \\ &\leq N \cdot \frac{8(M_1 + M_2)}{l\varepsilon^2} \end{aligned}$$

Let $N \cdot \frac{8(M_1 + M_2)}{l\varepsilon^2} = \eta$, we conclude that $\varepsilon = \sqrt{\frac{8N(M_1 + M_2)}{l\eta}}$.

Therefore

$$P \left\{ d \left(\left(R_{if}(\alpha_l), R_{ifemp}(\alpha_l) \right)_{\geq} \right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}} \right\} \geq 1 - \eta.$$

Theorem 4.4. The inequality

$$d \left(\left(R_{if}(\alpha_l), R_{if}(\alpha_0) \right)_{\geq} \right) \leq 2 \sqrt{\frac{2N(M_1 + M_2)}{l\eta}}$$

is satisfied with probability of at least $1 - 2\eta$.

Proof. Considering the properties of d and Theorem 4.3, we have

$$\begin{aligned} d \left(\left(R_{if}(\alpha_l), R_{if}(\alpha_0) \right)_{\geq} \right) &\leq d \left(\left(R_{if}(\alpha_l), R_{ifemp}(\alpha_l) \right)_{\geq} \right) + d \left(R_{ifemp}(\alpha_l), R_{if}(\alpha_0) \right) \\ &\leq d \left(\left(R_{if}(\alpha_l), R_{ifemp}(\alpha_l) \right)_{\geq} \right) + d \left(\left(R_{ifemp}(\alpha_0), R_{if}(\alpha_0) \right)_{\geq} \right) \\ &\leq 2 \sqrt{\frac{8N(M_1 + M_2)}{l\eta}}. \end{aligned}$$

Let the following relationships hold

$$A: d \left(\left(R_{if}(\alpha_l), R_{ifemp}(\alpha_l) \right)_{\geq} \right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}};$$

$$B: d \left(\left(R_{ifemp}(\alpha_0), R_{if}(\alpha_0) \right)_{\geq} \right) \leq \sqrt{\frac{8N(M_1 + M_2)}{l\eta}};$$

$$C: d \left(\left(R_{if}(\alpha_l), R_{if}(\alpha_0) \right)_{\geq} \right) \leq 2 \sqrt{\frac{8N(M_1 + M_2)}{l\eta}},$$

and $P(A) \geq 1 - \eta; P(B) \geq 1 - \eta$.

We can conclude that C must hold true in virtue of A and B true, then $P(C) \geq P(AB)$.

Therefore

$$P(C) \geq P(AB) = 1 - P((AB)^c) \geq 1 - P(A^c) - P(B^c) \geq 1 - 2\eta.$$

The theorem is proved.

When the samples are the fuzzy random samples, let $R_f(\alpha_l) = d \left(\int Q(\xi, \alpha) dF(\xi), I_{\{0\}} \right)$ and $R_{femp}(\alpha_l) = d \left(\frac{1}{n} \sum_{k=1}^l Q(\xi_k, \alpha), I_{\{0\}} \right)$. Then we have

Corollary 4.1[25,26]. Suppose that $Q(\xi, \alpha_k)$, $k = 1, 2, \dots, N$ is a set of functions

and $Q(\xi, \alpha_k)_r = \left[(Q(\xi, \alpha_k))_r^-, (Q(\xi, \alpha_k))_r^+ \right]$. If $(Q(\xi, \alpha_k))_r^-, (Q(\xi, \alpha_k))_r^+$ have the limited variance and the same upper bounds M , then

1) $R_f(\alpha_l) - R_{f_{emp}}(\alpha_l) \leq \sqrt{\frac{2M}{l\eta}}$ holds true with probability of at least $1 - \eta$.

2) $R_f(\alpha_l) - R_f(\alpha_0) \leq 2\sqrt{\frac{2M}{l\eta}}$ is satisfied with probability of at least $1 - 2\eta$.

Let d denote the subtraction of real numbers. We use Hoeffding's inequalities when the samples are the random samples in the usual sense. We have

Corollary 4.2[47]. If $\{Q(\xi, \alpha_k), \alpha_k \in \Lambda, k = 1, 2, \dots, N\}$ is bounded, i.e., $A \leq Q(\xi, \alpha_k) \leq B$, then

1) $R(\alpha_l) - R_{emp}(\alpha_l) \leq (B - A)\sqrt{\frac{\ln N - \ln \eta}{2l}}$ is valid with probability of at least $1 - \eta$.

2) $R(\alpha_l) - R(\alpha_0) \leq B\sqrt{\frac{\ln N - \ln \eta}{2l}} + (B - A)\sqrt{\frac{-\ln \eta}{2l}}$ holds true with probability of at least $1 - 2\eta$.

5. Conclusions

This paper discusses the intuitionistic fuzzy numbers and intuitionistic fuzzy random variables. We have proved the strong law of large numbers of intuitionistic fuzzy random variables and showed some related theorems and useful properties. Furthermore, based on intuitionistic fuzzy random samples, we propose the principle of intuitionistic fuzzy empirical risk minimization of learning theory, prove the key theorem of learning theory based on intuitionistic fuzzy random samples, and discuss the bounds of the rate of uniform convergence of learning process. Altogether these findings have laid the foundation for further research in statistical learning theory involving intuitionistic fuzzy random samples. Further investigations might focus on such fundamental issues as intuitionistic fuzzy structural risk minimization and address the applied aspects such as support vector machines based on intuitionistic fuzzy random samples, etc.

References

- [1] K. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets and Systems 20 (1986) 87–96.
- [2] K. Atanassov, More on intuitionistic fuzzy sets, Fuzzy Sets and Systems 33 (1989) 37–45.
- [3] Atanassov, New operations defined over the intuitionistic fuzzy sets, Fuzzy Sets and Systems 61 (1994) 137–142.
- [4] K. Atanassov, Remarks on the intuitionistic fuzzy sets – III, Fuzzy Sets and Systems 75 (1995) 401–402.
- [5] K. Atanassov, Intuitionistic Fuzzy Sets: Theory and Applications, Physica-Verlag, Heidelberg, 1999.
- [6] Zhaoqi Bian and Xuegong Zhang, Pattern Recognition (in Chinese), Tsinghua University Press, Beijing, 1999.
- [7] M. Brunato, R. Battiti, Statistical learning theory for location fingerprinting in wireless LANs. Computer Networks 47 (2005) 825–845.
- [8] P. Burillo, H. Bustince, Construction theorems for intuitionistic fuzzy sets, Fuzzy Sets and Systems 84

- (1996) 271–281.
- [9] H. Bustince, P. Burillo, Structures on intuitionistic fuzzy relations, *Fuzzy Sets and Systems* 78 (1996) 293–303.
 - [10] J. L. Castro, L. D. Flores-Hidalgo, C. J. Mantas, J. M. Puche, Extraction of fuzzy rules from support vector machines, *Fuzzy Sets and Systems* 158 (2007) 2057-2077.
 - [11] G. C. Cawley and N. L. C. Talbot, Improved sparse least-squares support vector machines, *Neurocomputing*, 48 (1-4) (2002) 1025-1031.
 - [12] A. Celikyilmaz, I. B. Turksen, Fuzzy functions with support vector machines, *Information Sciences* 177 (2007) 5163-5177.
 - [13] M. D. Cock, C. Cornelis, E.E. Kerre, Intuitionistic fuzzy relational images, *Studies in Computational Intelligence* 2 (2005) 129 – 145.
 - [14] D. Coker, An introduction to intuitionistic fuzzy topological spaces, *Fuzzy Sets and Systems* 88 (1997) 81–89.
 - [15] C. Cornelis, G. Deschrijver, E.E. Kerre, Implication in intuitionistic fuzzy and interval-valued fuzzy set theory: construction, classification, application, *International Journal of Approximate Reasoning* 35 (2004) 55 – 95.
 - [16] S.K. De, R. Biswas, A.R. Roy, Some operations on intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 114 (2000) 477–484.
 - [17] S.K. De, R. Biswas, A.R. Roy, Some operations on intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 114 (2000) 477 – 484.
 - [18] M. Demirci, Axiomatic theory of intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 110 (2000) 253–266.
 - [19] G. Deschrijver, E.E. Kerre, On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision, *Information Sciences* 177 (2007) 1860–1866.
 - [20] T. Evgeniou, T. Poggio, M. Pontil, et al, Regularization and statistical learning theory for data analysis, *Computational Statistics & Data Analysis* 38 (2002) 421-432.
 - [21] M. H. Ha, Y. C. Bai, P. Wang, et al, The key theorem and the bounds on the rate of uniform convergence of statistical learning theory on a credibility space, *Advances in Fuzzy Sets and Systems* 1 (2006) 143-172.
 - [22] M. H. Ha, Z. F. Feng, S. J. Song, L. Q. Gao, The key theorem and the bounds on the rate of uniform convergence of statistical learning theory on quasi-probability spaces, *Chinese Journal of Computers* 31 (2008) 476-485.
 - [23] M. H. Ha, Y. Li, J. Li, D. Z. Tian, The key theorem and the bounds on the rate of uniform convergence of learning theory on Sugeno measure space, *Science in China: Series F* 49 (2006) 1-14.
 - [24] M. H. Ha, W. Pedrycz, L. F. Zheng, The theoretical fundamentals of learning theory based on intuitionistic fuzzy random samples, *Fuzzy Sets and Systems* 178 (2009) 3240-3246.
 - [25] Minghu Ha, Wenguang Tang and Yunchao Bai. The key theorem of learning theory about fuzzy samples, *Proceeding of the Second International Conferences on Machine Learning and Cybernetics*, Xian, 2-5 November 2003, 1213-1216. 3-7.
 - [26] M. H. Ha, J. Tian, The theoretical foundations of statistical learning theory based on fuzzy number samples, *Information Sciences* 178 (2008) 3240-3246.
 - [27] M. H. Ha, J. Tian, C. L. Zhou, The key theorem and bounds on the rate of convergence of statistical learning theory on the fuzzy samples, *Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis* 14(S1) (2007) 72-76.
 - [28] C. L. Huang, M.C. Chen, C.J. Wang, Credit scoring with a data mining approach based on support vector machines, *Expert System with Applications* 33 (2007) 847-856.
 - [29] C. Hwang, D. H. Hong, K. H. Seok, Support vector interval regression machine for crisp input and output data, *Fuzzy Sets and Systems* 157 (2006) 1114-1125.
 - [30] B. Jin, Y. C. Tang, Y. Q. Zhang, Support vector machines with genetic fuzzy feature transformation for biomedical data classification, *Information Sciences* 177 (2007) 476-489.

- [31] T. Kikuchi, S. Abe, Comparison between error correcting output codes and fuzzy support vector machines, *Pattern Recognition Letters* 26 (2005) 1937-1945.
- [32] Y. K. Kim, A strong law of large numbers for fuzzy random variables, *Fuzzy Sets and Systems* 111 (2000) 319-323.
- [33] D. Li, F. Shan, C. Cheng, On properties of four IFS operators, *Fuzzy Sets and Systems* 154 (2005) 151 – 155.
- [34] Z. Z. Liang, P. F. Shi, Similarity measures on intuitionistic fuzzy sets, *Pattern Recognition Letters* 24 (2003) 2687 – 2693.
- [35] C. C. Lin, S. H. Chen, T. K. Truong and Y. Chang, Auto classification and categorization based on wavelets and support vector machine, *IEEE Transactions on Speech and Audio Processing*, 13 (5) (2005), 644-651.
- [36] C. F. Lin, S. D. Wang, Fuzzy support vector machines, *IEEE Transactions on Neural Networks* 13 (2002) 464-471.
- [37] L. Lin, X. H. Yuan, Z. Q. Xia, Multicriteria fuzzy decision-making methods based on intuitionistic fuzzy sets, *Journal of Computer and System Sciences* 73 (2007) 84 – 88.
- [38] B. D. Liu. *Uncertainty Theory: An Introduction to its Axiomatic Foundations*, Springer-Verlag, Berlin, 2004.
- [39] Y. H. Liu, Y. T. Chen, Face recognition using total margin-based adaptive fuzzy support vector machines, *IEEE Transactions on Neural Networks* 18 (2007) 178-192.
- [40] T. K. Mondal, S. K. Samanta, Topology of interval-valued intuitionistic fuzzy sets, *Fuzzy Sets and Systems* 119 (2001) 483–494.
- [41] C. Su, C. H. Yang, Feature selection for the SVM: an application to hypertension diagnosis, *Expert Systems with Applications* 34 (2008) 754-763.
- [42] D. Z. Tian, Z. M. Zhang, M. H. Ha, The key theorem of complex statistical learning theory, *Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis* 14(S1) (2007) 46-50.
- [43] K. Tomonori and A. Shigeo, Comparison between error correcting output codes and fuzzy support vector machines, *Pattern Recognition Letters*, 26 (12) (2005), 1937-1945.
- [44] C. F. Tsai, Training support vector machines based on stacked generalization for image classification, *Neurocomputing*, 64 (2005), 497-503.
- [45] J. J. Tsong, Hybrid approach of selecting hyper-parameters of support vector machine for regression, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 36 (2006) 699-709.
- [46] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- [47] V. N. Vapnik, *Statistical Learning Theory*, A Wiley-Interscience Publication, New York, 1998.
- [48] V. N. Vapnik, An overview of statistical learning theory, *IEEE Transactions on Neural Networks* 10 (1999) 988-999.
- [49] M. Vidyasagar, Statistical learning theory and randomized algorithms for control, *IEEE Control Systems Magazine* 12 (1998) 69-85.
- [50] Y. Q. Wang, S. Y. Wang, K. K. Lai, A new fuzzy support vector machine to evaluate credit risk, *IEEE Transactions on Fuzzy Systems* 13 (2005) 820-931.
- [51] Z. Y. Wang, J. K. George, *Fuzzy Measure Theory*, Plenum Press, New York, 1992.
- [52] H. Wechsler, Z. Duric, F. Y. Li, V. Cherkassky, Motion estimation using statistical learning theory, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (2004) 466-478.
- [53] W. Y. N. Wing, A. Dorado, D. S. Yeung, W. Pedrycz and E. Izquierdo, Image classification with the use of radial basis function neural networks and the minimization of the localized generalization error, *Pattern Recognition*, 40 (2007), 19 -32.
- [54] Z.S. Xu, Intuitionistic preference relations and their application in group decision making, *Information Sciences* 177 (2007) 2363 – 2379.
- [55] L.A. Zadeh, Fuzzy sets, *Information and Control* 8 (1965) 338–353.
- [56] Yiqiang Zhan and Dinggang Shen, Design efficient support vector machine for fast classification, *Pattern*

Recognition, 38 (1) (2005), 157-161.

- [57] C. Y. Zhang, H. Y. Fu, Similarity measures on three kinds of fuzzy sets, Pattern Recognition Letters 27 (2006) 1307 - 1317.