British Journal of Mathematics & Computer Science
4(15): 2197-2216, 2014

SCIENCEDOMAIN *international*
www.sciencedomain.org

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

Zhiming Zhang[1*]

[1]*College of Mathematics and Computer Science, Hebei University, Baoding 071002, Hebei Province, P. R. China.*

Received: 26 March 2014
Accepted: 02 June 2014
Published: 11 June 2014

Original Research Article

Abstract

In this paper, we discuss the statistical learning theory based on intuitionistic fuzzy random sample. First of all, we introduce the definition of intuitionistic fuzzy number intuitionistic fuzzy random variables. Secondly, we give some properties of this concept. Thirdly, the definitions of intuitionistic fuzzy empirical risk functional, intuitionistic fuzzy expected risk functional, and intuitionistic fuzzy empirical risk minimization principle are presented. Finally, we prove the key theorem of intuitionistic fuzzy random sample and obtain the rate of uniform convergence of learning process based on the intuitionistic fuzzy random sample.

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), put forward in 1960s and completely founded by Vapnik et al. in 1990s. [1-3], has become an interesting and good law that supports the development of small samples statistical learning. The SLT has become the fastest growing discipline in machine learning in the late 1990's. 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 based on small data samples. Meanwhile, SLT gave rise to a new category of general learning algorithms, which we called the Support Vector Machine (SVM, for short). Currently, the SLT and SVM constitute interesting research avenues in machine learning [4-24].

The SLT covers four main parts [4]: (1) the learning process that minimizes the necessary and sufficient conditions for the consistency of the empirical risk, which is referred to as the key theorem of learning theory; (2) the scope of generalization; (3) the structural risk minimization principle; (4) the support vector machine (SVM) algorithm that implements the structure risk

**Corresponding author: zhimingzhang@ymail.com;*

minimization principle. The key theorem of learning theory and the boundary of convergence of the learning process are the two most important foundations for further research on related topics, such as the structural risk minimization and the SVM.

Although the SLT has reached a level of high maturity, there are still some problems to be solved, for example, the development of SLT and SVM is based on probability space and real-valued random sample (real-valued random variables). In the real world, there are many non-probability space (such as the Sugeno measure space [25], credibility measure space [26]) and non-real random samples (such as fuzzy sample [27], fuzzy random sample [28], and fuzzy complex random samples [29]). In order to solve these problems, the expansions of the SLT in non-probabilistic space and non-real random samples have become an urgent need to move forward. Some studies have been achieved along this line [29-35]. For example, Ha et al. [29-31] studied the key theorem and the generalized boundary on Sugeno measure probability space and quasi-probability measure space; Tian et al. [17] investigated the construction of statistical learning theory based on fuzzy random samples and fuzzy complex random samples; Lin et al. [15] constructed fuzzy support vector machine based on random sample.

Intuitionistic (IF, for short) set, originated by Atanassov [36-41], is an important concept used to cope with imperfect and/or imprecise information. It is an intuitively straightforward extension of Zadeh's fuzzy sets [42]: 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 [36-41]. 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., [43-50]. 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 [43,46,51,52]. Recently, IF set theory has been successfully applied in decision analysis and pattern recognition (see, e.g., [53-58]).

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 [28]. A fuzzy number is a fuzzy set $X : \mathbb{R} \to [0,1]$ 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) $\operatorname{supp} X = cl\{x \in \mathbb{R} : X(x) > 0\}$ 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

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

then it follows that X is a fuzzy number if and only if $(X)_1 \neq \varnothing$ 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 $X, Y \in F(\mathbb{R})$, $\forall r \in [0,1]$, $(X)_r = \left[(X)_r^-, (X)_r^+\right]$ and $(Y)_r = \left[(Y)_r^-, (Y)_r^+\right]$. Define

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

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 [28]. A fuzzy number valued function $\xi : \Omega \to F(\mathbb{R})$ 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 [28]. A fuzzy random variable $\xi(\omega) = \left\{\left[\xi_r^-(\omega), \xi_r^+(\omega)\right] \middle| 0 < r \leq 1\right\}$ 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 $\|\xi\| = \max\left\{\left|\xi_0^-\right|, \left|\xi_0^+\right|\right\}$. In this case, the expected value of ξ is defined in the following manner

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

Definition 2.4 [36]**.** Let U be a nonempty set. An intuitionistic fuzzy set A in U is an object having the form $A = \left\{\left\langle x, \mu_A(x), \nu_A(x)\right\rangle \middle| x \in U\right\}$, where $\mu_A : U \to [0,1]$ and $\nu_A : U \to [0,1]$ satisfy $0 \le \mu_A + \nu_A \le 1$ 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 [36-41,55,56]:

Let A and B be two IF sets, then

$A \subseteq B$ iff $\mu_A(x) \le \mu_B(x)$ and $\nu_A(x) \le \nu_B(x)$ for all $x \in U$,
$A \supseteq B$ iff $B \subseteq A$,
$A = B$ iff $A \subseteq B$ and $B \subseteq A$, i.e., $\mu_A(x) = \mu_B(x)$ and $\nu_A(x) = \nu_B(x)$ for any $x \in U$,
$A \cap B = \{\langle x, \mu_A(x) \wedge \mu_B(x), \nu_A(x) \vee \nu_B(x)\rangle | x \in U\}$,
$A \cup B = \{\langle x, \mu_A(x) \vee \mu_B(x), \nu_A(x) \wedge \nu_B(x)\rangle | x \in U\}$,
$\forall \alpha, \beta \in [0,1], \alpha + \beta \le 1,\ (\alpha, \beta) A = \{\langle x, \alpha \wedge \mu_A(x), \beta \vee \nu_A(x)\rangle | x \in U\}$.

Definition 2.5 [41]. Let A be an IF set and $\alpha, \beta \in [0,1]$ with $\alpha + \beta \le 1$, the $(\alpha, \beta)-$ level cut set of A, denoted by A_α^β, is defined as follows:

$$A_\alpha^\beta = \{x \in U \,|\, \mu_A(x) \ge \alpha, \nu_A(x) \le \beta\}$$

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

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

$$\mu_{\bigcup(\alpha,\beta)A_\alpha^\beta}(x) = \vee \mu_{(\alpha,\beta)A_\alpha^\beta}(x) = \vee\left(\alpha \wedge A_\alpha^\beta(x)\right)$$
$$= \left(\underset{0\le\alpha\le\mu_A(x)}{\vee}\left(\alpha \wedge A_\alpha^\beta(x)\right)\right) \vee \left(\underset{\mu_A(x)<\alpha\le 1}{\vee}\left(\alpha \wedge A_\alpha^\beta(x)\right)\right)$$
$$= \underset{0\le\alpha\le\mu_A(x)}{\vee} \alpha = \mu_A(x).$$

On the other hand, we have

$$\nu_{\bigcup(\alpha,\beta)A_\alpha^\beta}(x) = \wedge \nu_{(\alpha,\beta)A_\alpha^\beta}(x) = \wedge\left(\beta \vee \left(1 - A_\alpha^\beta(x)\right)\right)$$
$$= \left(\underset{0\le\beta\le\nu_A(x)}{\wedge}\left(\beta \vee \left(1 - A_\alpha^\beta(x)\right)\right)\right) \wedge \left(\underset{\nu_A(x)<\beta\le 1}{\wedge}\left(\beta \vee \left(1 - A_\alpha^\beta(x)\right)\right)\right)$$
$$= \underset{\nu_A(x)<\beta\le 1}{\wedge} \beta = \nu_A(x).$$

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

Definition 2.6 [41]. The IF set $A = \left\{\left\langle x, \mu_A(x), \nu_A(x)\right\rangle \middle| x \in \mathbb{R}\right\}$ 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) $\mu_A : \mathbb{R} \to [0,1]$ is an upper semi-continuous function and $\mu_A : \mathbb{R} \to [0,1]$ 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:

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

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

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

If $\rho = 0$, then

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

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

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

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

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

Definition 2.8 [41]. Let $A = \left\{\left\langle x, \mu_A(x), \nu_A(x)\right\rangle \middle| x \in U\right\}, B = \left\{\left\langle x, \mu_B(x), \nu_B(x)\right\rangle \middle| x \in U\right\} \in IF(\mathbb{R})$.

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

$$d(A,B)=\sup_{\substack{\alpha,\beta\in[0,1]\\ \alpha+\beta\le 1}}\max\left\{d_H\left(A_\alpha,B_\alpha\right),d_H\left(A^\beta,B^\beta\right)\right\}$$
$$=\max\left\{d\left(\mu_A,\mu_B\right),d\left(\nu_A,\nu_B\right)\right\}$$
$$=\max\left\{d\left(\mu_A,\mu_B\right),d\left(\nu_A^c,\nu_B^c\right)\right\}$$

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

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

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

Definition 2.10. Let $(\Omega,\mathcal{A},P)$ be a probability measure space. An intuitionistic fuzzy number valued mapping $\xi:\Omega\to IF(\mathbb{R})$, $\omega\to\xi(\omega)=\left\{\left\langle x,\mu_{\xi(\omega)}(x),\nu_{\xi(\omega)}(x)\right\rangle \middle| x\in\mathbb{R}\right\}$ 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 $\mu_\xi:\omega\to\mu_{\xi(\omega)}$, $\nu_\xi:\omega\to\nu_{\xi(\omega)}$ and $\nu_\xi^c=1-\nu_\xi$.

Definition 2.11 Let $\xi(\omega)=\left\{\left\langle x,\mu_{\xi(\omega)}(x),\nu_{\xi(\omega)}(x)\right\rangle \middle| x\in\mathbb{R}\right\}$ be the intuitionistic fuzzy random variable defined on $(\Omega,\mathcal{A})$. We called

$$F(Z)=P\left\{\xi(\omega)\le Z\right\}=P\left\{\mu_{\xi(\omega)}\le\mu_Z,\nu_{\xi(\omega)}\ge\nu_Z\right\}$$

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

Definition 2.12. We call $\xi(\omega)=\left\{\left\langle x,\mu_{\xi(\omega)}(x),\nu_{\xi(\omega)}(x)\right\rangle \middle| x\in\mathbb{R}\right\}$ integrable if μ_ξ and ν_ξ^c are integrable. In this case, we define the mathematical expectation of ξ as the following manner

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

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

Proposition 2.6. Let ξ be an intuitionistic fuzzy random variable, then the following equalities

1. $\left\{E(\xi)\right\}_\alpha^\beta=E\left(\xi_\alpha^\beta\right)=E\left(\xi_\alpha\right)\cap E\left(\xi^\beta\right)$ for $0<r\le 1$;
2. $E(c\xi)=cE\xi$, whenever $c\in\mathbb{R}$;
3. $E\left(\xi_1\pm\xi_2\right)=E\xi_1\pm E\xi_2$

hold.

Definition 2.13. Assume that $\xi_n = \left\{ \left\langle x, \mu_{\xi_n}(x), \nu_{\xi_n}(x) \right\rangle \middle| x \in U \right\}, n = 1, 2, \cdots$ is a collection of intuitionistic fuzzy random variables.

(1) If the fuzzy random variables sequence $\mu_{\xi_n}, n = 1, 2, \cdots$ and $\nu^c_{\xi_n}, n = 1, 2, \cdots$ are respectively mutually independent, then $\xi_n, n = 1, 2, \cdots$ is called a collection of mutually independent intuitionistic fuzzy random variables.

(2) If $\mu_{\xi_n}, n = 1, 2, \cdots$ and $\nu^c_{\xi_n}, n = 1, 2, \cdots$ are two sequences of identically distributed fuzzy random vectors, then $\xi_n, n = 1, 2, \cdots$ is referred to as a sequence of identically distributed intuitionistic fuzzy random variables.

Definition 2.14. Let $\xi_n, n = 1, 2, \cdots$ be a sequence of intuitionistic fuzzy random variables and let ξ be an intuitionistic fuzzy random variable (or an intuitionistic fuzzy number). If $\forall \varepsilon > 0$,

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

then ξ_n converge in probability P to ξ, denoted by

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

Theorem 2.1. Suppose that $\xi_n = \left\{ \left\langle x, \mu_{\xi_n}(x), \nu_{\xi_n}(x) \right\rangle \middle| x \in U \right\}, n = 1, 2, \cdots$ is a sequence of intuitionistic fuzzy random variables and $\xi = \left\{ \left\langle x, \mu_{\xi}(x), \nu_{\xi}(x) \right\rangle \middle| x \in U \right\}$ is an intuitionistic fuzzy random variable. Then, $\xi_n \xrightarrow{P} \xi$ if and only if $\mu_{\xi_n} \xrightarrow{P} \mu_{\xi}$ and $\nu_{\xi_n} \xrightarrow{P} \nu_{\xi}$.

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

$$d\left(\mu_{\xi_n}, \mu_{\xi}\right) \le d\left(\xi_n, \xi\right),\ d\left(\nu_{\xi_n}, \nu_{\xi}\right) \le d\left(\xi_n, \xi\right),\ d\left(\xi_n, \xi\right) \le d\left(\mu_{\xi_n}, \mu_{\xi}\right) + d\left(\nu_{\xi_n}, \nu_{\xi}\right).$$

Hence,

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

The theorem is easily proven.

Theorem 2.2 (The strong law of large numbers of fuzzy random variables) [28]. Let $\left\{ \xi_n, n = 1, 2, \cdots \right\}$ be a sequence of independent and identically distributed fuzzy random variables with $E\|\xi_1\| < \infty$. Then we have

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

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

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

Proof. It follows from Theorem 2.2 and Proposition 2.3.

3 The Key Theorem of Learning Theory Based on Intuitionistic Fuzzy Random Samples

Let $\xi_j,\ j=1,2,\cdots,l$ be a sequence of independent and identically distributed intuitionistic fuzzy random samples whose distribution is given as $F(\xi)$.

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

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

Let the risk functional obtain its minimum at $Q(\xi,\alpha_0)$ and the empirical risk functional obtain its minimum at $Q(\xi,\alpha_l)$.

Definition 3.3. We minimize $R_{ifemp}(\alpha)$ to replace $R_{if}(\alpha)$ and refer to the function $Q(\xi,\alpha_l)$ as an approximation to the function $Q(\xi,\alpha_0)$. We call this principle the intuitionistic fuzzy empirical risk minimization principle.

Definition 3.4. If the following two sequences converge in probability to the same limit:

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

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

then 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 $F(\xi)$

Definition 3.5. 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

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

then 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 $F(\xi)$.

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

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

holds for any $\varepsilon>0$.

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

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

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 $d\left(\left(R_{if}(\alpha^*), a_k\right)_{\geq}\right)<\frac{\varepsilon}{2}$. Then for these $\Lambda(a_k)$ the inequalities

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

hold true.
Therefore

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

According to (3.1), we have

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

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 $A \subseteq \bigcup_k T_k$,

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

Hence

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

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

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

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

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

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

Then

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

Therefore

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

Hence

$$P\left(A\right) \leq P\left(A_1\right) + P\left(A_2\right) \xrightarrow[l\to\infty]{} 0 .$$

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

Corollary 3.1 [29]**.** Let the set of functions $Q\left(\xi,\alpha\right), \alpha \in \Lambda$, satisfies $X \leq R\left(\alpha\right) \leq Y$. Then the sufficient and necessary condition for the strict consistency of FERM principle is the convergence

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

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

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

hold true. Then the following two statements are equivalent:

1. For the given distribution function $F\left(\xi\right)$, the empirical risk minimization method is strictly consistent on the set of functions $Q\left(\xi,\alpha\right), \alpha \in \Lambda$.
2. For the given distribution function $F\left(\xi\right)$, the uniform one-sided convergence of the means to their mathematical expectation takes place over the set of functions $Q\left(\xi,\alpha\right), \alpha \in \Lambda$.

4 Bounds of the Rate of Uniform Convergence of Intuitionistic Fuzzy Random Samples

In statistical learning theory, the important conclusions about the relationship between empirical risk and actual risk form the promotion of some boundary. They are essential to learning machine capacity analysis and the development of new learning algorithms. An important part of the rate of convergence of learning processes is the generalization bounds. In this section, consensus convergence of learning process based on intuitionistic fuzzy random samples is discussed. 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 $Q(\xi,\alpha_k)$, $k=1,2,\cdots,N$. At first, we introduce a basic theorem:

Theorem 4.1[28,29]**.** Let $\xi_j=\left\{\left[(\xi_j)_r^-,(\xi_j)_r^+\right]\middle|r\in[0,1]\right\}$, $j=1,2,\cdots,n$ be a sequence of fuzzy random variables, and $(\xi_j)_r^-,(\xi_j)_r^+$ 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 $\xi_j=\left\{\left\langle x,\mu_{\xi_j}(x),\nu_{\xi_j}(x)\right\rangle\middle|x\in U\right\}$, $j=1,2,\cdots,n$ be a sequence of intuitionistic fuzzy random variables, $(\mu_{\xi_j})_r^-,(\mu_{\xi_j})_r^+$ and $(\nu_{\xi_j})_r^-,(\nu_{\xi_j})_r^+$ 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

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

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

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

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

Proof.

$$P\left\{\sup_{1\leq k\leq N} d\left(\left(R_{if}\left(\alpha_k\right),R_{ifemp}\left(\alpha_k\right)\right)_{\geq}\right)>\varepsilon\right\}\leq\sum_{k=1}^{N}P\left\{d\left(\left(R_{if}\left(\alpha_k\right),R_{ifemp}\left(\alpha_k\right)\right)_{\geq}\right)>\varepsilon\right\}$$

$$\leq N\cdot\frac{8\left(M_1+M_2\right)}{l\varepsilon^2}$$

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

Therefore

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

Theorem 4.4. The inequality

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

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

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

$$d\left(\left(R_{if}\left(\alpha_l\right),R_{if}\left(\alpha_0\right)\right)_{\geq}\right)\leq d\left(\left(R_{if}\left(\alpha_l\right),R_{ifemp}\left(\alpha_l\right)\right)_{\geq}\right)+d\left(R_{ifemp}\left(\alpha_l\right),R_{if}\left(\alpha_0\right)\right)$$

$$\leq d\left(\left(R_{if}\left(\alpha_l\right),R_{ifemp}\left(\alpha_l\right)\right)_{\geq}\right)+d\left(\left(R_{ifemp}\left(\alpha_0\right),R_{if}\left(\alpha_0\right)\right)_{\geq}\right)$$

$$\leq 2\sqrt{\frac{8N\left(M_1+M_2\right)}{l\eta}}.$$

Let the following relationships hold

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

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

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

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

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

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[28,29]**.** Suppose that $Q(\xi,\alpha_k)$, $k = 1,2,\cdots,N$ is a set of functions and $Q(\xi,\alpha_k)_r = \left[\left(Q(\xi,\alpha_k)\right)_r^-, \left(Q(\xi,\alpha_k)\right)_r^+\right]$. If $\left(Q(\xi,\alpha_k)\right)_r^-$, $\left(Q(\xi,\alpha_k)\right)_r^+$ have the limited variance and the same upper bounds M , then

1) $R_f(\alpha_l) - R_{femp}(\alpha_l) \leq \sqrt{\frac{2M}{l\eta}}$ 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[3]**.** If $\{Q(\xi,\alpha_k), \alpha_k \in \Lambda, k = 1,2,\cdots,N\}$ 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 Conclusion

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.

Acknowledgements

The authors thank the anonymous referees for their valuable suggestions in improving this paper. This work is supported by the National Natural Science Foundation of China (Grant Nos. 61073121 and 61375075) and the Natural Science Foundation of Hebei Province of China (Grant Nos. F2012201020 and A2012201033).

Author's Contribution

'Zhiming Zhang' designed the study, performed the statistical analysis, wrote the protocol, and wrote the overall draft of the manuscript. The author read and approved the final manuscript.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Vapnik VN. The Nature of Statistical Learning Theory. Springer-Verlag, New York; 1995.

[2] Vapnik VN. Statistical Learning Theory. A Wiley-Interscience Publication, New York; 1998.

[3] Vapnik VN. An overview of statistical learning theory. IEEE Transactions on Neural Networks. 1999;10:988-999.

[4] Bian ZQ, Zhang XG. Pattern Recognition (in Chinese). Tsinghua University Press, Beijing; 1999.

[5] Brunato M, Battiti R. Statistical learning theory for location fingerprinting in wireless LANs. Computer Networks. 2005;47:825–845.

[6] Castro JL, Flores-Hidalgo LD, Mantas CJ, Puche JM. Extraction of fuzzy rules from support vector machines. Fuzzy Sets and Systems. 2007;158-205:7-2077.

[7] Cawley GC, Talbot NLC. Improved sparse least-squares support vector machines. Neurocomputing. 2002;48(1-4):1025-1031.

[8] Celikyilmaz A, Turksen IB. Fuzzy functions with support vector machines. Information Sciences. 2007;177:5163-5177.

[9] Evgeniou T, Poggio T, Pontil M, et al. Regularization and statistical learning theory for data analysis. Computational Statistics & Data Analysis. 2002;38:421-432.

[10] Huang CL, Chen MC, Wang CJ. Credit scoring with a data mining approach based on support vector machines. Expert System with Applications. 2007;33:847-856.

[11] Hwang C, Hong DH, Seok KH. Support vector interval regression machine for crisp input and output data. Fuzzy Sets and Systems. 2006;157:1114-1125.

[12] Jin B, Tang YC, Zhang YQ. Support vector machines with genetic fuzzy feature transformation for biomedical data classification. Information Sciences. 2007;177:476-489.

[13] Kikuchi T, Abe S. Comparison between error correcting output codes and fuzzy support vector machines. Pattern Recognition Letters. 2005;26:1937-1945.

[14] Lin CC, Chen SH, Truong TK, Chang Y. Auto classification and categorization based on wavelets and support vector machine. IEEE Transactions on Speech and Audio Processing. 2005;13(5):644-651.

[15] Lin CF, Wang SD. Fuzzy support vector machines. IEEE Transactions on Neural Networks. 2002;13:464-471.

[16] Liu YH, Chen Y. T. Face recognition using total margin-based adaptive fuzzy support vector machines. IEEE Transactions on Neural Networks. 2007;18:178-192.

[17] Tian DZ, Zhang ZM, Ha MH. The key theorem of complex statistical learning theory, Dynamics of Continuous. Discrete and Impulsive Systems, Series A: Mathematical Analysis. 2007;14(S1):46-50.

[18] Tsai CF. Training support vector machines based on stacked generalization for image classification. Neurocomputing. 2005;64:497-503.

[19] Tsong JJ. Hybrid approach of selecting hyper-parameters of support vector machine for regression. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 2006;36:699-709.

[20] Wang YQ, Wang SY, Lai KK. A new fuzzy support vector machine to evaluate credit risk. IEEE Transactions on Fuzzy Systems. 2005;13:820-931.

[21] Wang ZY, George JK. Fuzzy Measure Theory. Plenum Press, New York; 1992.

[22] Wing WYN, Dorado A, Yeung DS, Pedrycz W, Izquierdo E. Image classification with the use of radial basis function neural networks and the minimization of the localized generalization error. Pattern Recognition. 2007;40:19 -32.

[23] Xu ZS. Intuitionistic preference relations and their application in group decision making. Information Sciences. 2007;177:2363–2379.

[24] Zhang CY, Fu HY. Similarity measures on three kinds of fuzzy sets. Pattern Recognition Letters. 2006;27:1307–1317.

[25] Wechsler H, Duric Z, Li FY, Cherkassky V. Motion estimation using statistical learning theory. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004;26:466-478.

[26] Liu BD. Uncertainty Theory: An Introduction to its Axiomatic Foundations. Springer-Verlag, Berlin; 2004.

[27] Zhan YQ, Shen DG. Design efficient support vector machine for fast classification. Pattern Recognition. 2005;38(1):157-161.

[28] Kim YK. A strong law of large numbers for fuzzy random variables. Fuzzy Sets and Systems. 2000;111:319-323.

[29] Ha MH, Tian J. The theoretical foundations of statistical learning theory based on fuzzy number samples. Information Sciences. 2008;178:3240-3246.

[30] Ha MH, Bai YC, Wang P, 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. 2006;1:143-172.

[31] Ha MH, Feng ZF, Song SJ, Gao LQ. The key theorem and the bounds on the rate of uniform convergence of statistical learning theory on quasi-probability spaces. Chinese Journal of Computers. 2008;31:476-485.

[32] Ha MH, Li Y, Li J, Tian DZ. The key theorem and the bounds on the rate of uniform convergence of learning theory on Sugeno measure space. Science in China: Series F. 2006;49:1-14.

[33] Ha MH, Pedrycz W, Zheng LF. The theoretical fundamentals of learning theory based on intuitionistic fuzzy random samples. Fuzzy Sets and Systems. 2009;178:3240-3246.

[34] Ha MH, Tang WG, Bai YC. The key theorem of learning theory about fuzzy samples. Proceeding of the Second International Conferences on Machine Learning and Cybernetics, Xian. 2003;1213-1216:3-7.

[35] Ha MH, Tian J, Zhou CL. 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. 2007;14(1):72-76.

[36] Atanassov K. Intuitionistic fuzzy sets. Fuzzy Sets and Systems. 1986;20:87–96.

[37] Atanassov K. More on intuitionistic fuzzy sets. Fuzzy Sets and Systems. 1989;33:37–45.

[38] Atanassov K. New operations defined over the intuitionistic fuzzy sets. Fuzzy Sets and Systems. 1994;61:137–142.

[39] Atanassov K. Remarks on the intuitionistic fuzzy sets – III. Fuzzy Sets and Systems. 1995;75:401–402.

[40] Atanassov K. Intuitionistic Fuzzy Sets: Theory and Applications. Physica-Verlag, Heidelberg; 1999.

[41] Atanassov K. Remark on intuitionistic fuzzy numbers. Notes on Intuitionistic Fuzzy Sets. 2007;13(3):29-32.

[42] Zadeh LA. Fuzzy sets. Information and Control. 1965;8:338–353.

[43] Burillo P, Bustince H. Construction theorems for intuitionistic fuzzy sets. Fuzzy Sets and Systems. 1996;84:271–281.

[44] Bustince H, Burillo P. Structures on intuitionistic fuzzy relations. Fuzzy Sets and Systems. 1996;78:293–303.

[45] Cock MD, Cornelis C, Kerre EE. Intuitionistic fuzzy relational images. Studies in Computational Intelligence. 2005;2:129–145.

[46] Coker D. An introduction to intuitionistic fuzzy topological spaces. Fuzzy Sets and Systems. 1997;88:81–89.

[47] Cornelis C, Deschrijver G, Kerre EE. Implication in intuitionistic fuzzy and interval-valued fuzzy set theory: Construction, classification, application. International Journal of Approximate Reasoning. 2004;35:55–95.

[48] Li D, Shan F, Cheng C. On properties of four IFS operators. Fuzzy Sets and Systems. 2005;154:151–155.

[49] Liang ZZ, Shi PF. Similarity measures on intuitionistic fuzzy sets. Pattern Recognition Letters. 2003;24:2687–2693.

[50] Atanassov, K. Intuitionistic fuzzy sets, VII ITKR's Session, Sofia, (Deposed in Central Sci. -Techn. Library of Bulg Acad of Sci, 1697/84) (in Bulg.); 1983.

[51] Demirci M. Axiomatic theory of intuitionistic fuzzy sets. Fuzzy Sets and Systems. 2000;110:253–266.

[52] Su C, Yang CH. Feature selection for the SVM: An application to hypertension diagnosis. Expert Systems with Applications. 2008;34:754-763.

[53] Deschrijver G, Kerre EE. On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision. Information Sciences. 2007;177:1860–1866.

[54] Lin L, Yuan XH, Xia ZQ. Multicriteria fuzzy decision-making methods based on intuitionistic fuzzy sets. Journal of Computer and System Sciences. 2007;73:84–88.

[55] De SK, Biswas R, Roy AR. Some operations on intuitionistic fuzzy sets. Fuzzy Sets and Systems. 2000;114:477–484.

[56] Mondal TK, Samanta SK. Topology of interval-valued intuitionistic fuzzy sets. Fuzzy Sets and Systems. 2001;119:483–494.

[57] Tomonori K, Shigeo A. Comparison between error correcting output codes and fuzzy support vector machines. Pattern Recognition Letters. 2005;26(12):1937-1945.

[58] Vidyasagar M. Statistical learning theory and randomized algorithms for control. IEEE Control Systems Magazine. 1998;12:69-85.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=550&id=6&aid=4877