# How Good Are Fuzzy If-Then Classifiers?

Ludmila I. Kuncheva, *Member, IEEE*

***Abstract*—This paper gives some known theoretical results about fuzzy rule-based classifiers and offers a few new ones. The ability of Takagi–Sugeno–Kang (TSK) fuzzy classifiers to match exactly and to approximate classification boundaries is discussed. The lemma by Klawonn and Klement about the exact match of a classification boundary in $\Re^2$ is extended from monotonous to arbitrary functions. Equivalence between fuzzy rule-based and nonfuzzy classifiers (1-nn and Parzen) is outlined. We specify the conditions under which a class of fuzzy TSK classifiers turn into lookup tables. It is shown that if the rule base consists of all possible rules (all combinations of linguistic labels on the input features), the fuzzy TSK model is a lookup classifier with hyperbox cells, regardless of the type (shape) of the membership functions used. The question "why fuzzy?" is addressed in the light of these results.**

***Index Terms*—Fuzzy classifiers, fuzzy if-then systems (TSK), pattern recognition, theoretical result.**

## I. Introduction

*FUZZY rule-based classifiers* are a popular counterpart of fuzzy control systems. There are numerous studies discussing the practical design of such classifiers, among which are neuro-fuzzy models [11], [12], [16], fuzzy systems constructed using genetic algorithms [5], [6], [14], etc. While there is an abundance of theoretical results about fuzzy rule-based control systems, few publications explore rigorously architectural or theoretical aspects of fuzzy if-then classifiers [1], [7], [9], [17].

This paper has no ambition to survey the state of the art. Presented here are some of those results that are perhaps considered too intuitive and straightforward to be worth detailing. Yet, we think that summarizing and explicating them can help in understanding better fuzzy classifiers. Section II introduces the fuzzy classification formalism. *Exact match* of classification boundaries is discussed in Section III. The lemma by Klawonn and Klement [7] is extended from monotonous to arbitrary functions. Section IV discusses *approximation* with fuzzy classifiers by redressing the fuzzy TSK classifier as a fuzzy multi-input single-output (MISO) system and referring to the Stone–Weierstrass theorem of universal approximation. Shown there also are two theorems relating fuzzy TSK models with nearest neighbor classifier (1-nn) and the Parzen classifier [9]. Section V contains an original result. We show that if the rule base consists of all possible rules (all combinations of linguistic labels on the input features), then the fuzzy TSK classifier with any fuzzy $t$-norm, extended as the conjunction in the antecedent part of the rules, is a lookup classifier with hyperbox cells, regardless of the type (shape) of the membership functions used. The conclusion section addresses the question "why fuzzy?" in the light of the above results.

Manuscript received May 10, 1999; revised September 29, 1999 and March 12, 2000. This paper was recommended by Associate Editor L. O. Hall.

The author is with the School of Informatics, University of Wales, Bangor, Bangor, U.K. (e-mail: l.i.kuncheva@bangor.ac.uk).

Publisher Item Identifier S 1083-4419(00)06710-8.

## II. Fuzzy Classification

Let $\Omega = \{\omega_1, \ldots, \omega_c\}$ be a set of class labels, e.g., $\{victory, draw, defeat\}$ or $\{wolf, fox, bear\}$. Let $\mathbf{x} = [x_1, \ldots, x_n]^T \in \Re^n$ be a vector describing an object which for the example above can be a particular football match or a paw-print image. Each component of $\mathbf{x}$ expresses the value of a *feature*, such as length, temperature, number of prickle per cm$^2$, concentration of cadmium, etc. A *classifier* is any mapping

$$D : \Re^n \to \Omega. \tag{1}$$

We consider the canonical model of the classifier [3] as a black box at the input of which we submit $\mathbf{x}$, and at the output obtain the values of $c$ discriminant functions $g_1(\mathbf{x}), \ldots, g_c(\mathbf{x})$, expressing the support for the respective classes. The *maximum membership rule* assigns $\mathbf{x}$ to the class with the highest support.

In fuzzy systems (fuzzy classifiers in this number), typically, the features are associated with linguistic labels, e.g., *high, normal, early*. These values are represented as fuzzy sets on the feature axes (profit, blood pressure, arrival time). Let $K_j$ be the number of linguistic labels for the $j$th feature and let $A_{j,i}$ denote the $i$th fuzzy set on axis $x_j, i = 1, \ldots, K_j, j = 1, \ldots, n$. Fuzzy systems are meant to be a transparent model implementing logical reasoning, presumably understandable to the end-user of the system. A class of such systems employ if-then rules and an inference mechanism which, ideally, should correspond to the expert knowledge and decision making process for a given problem. A fuzzy if-then classifier uses rules of the type (called in the sequel *the general type*)

$$R_k\text{: IF } x_1 \text{ is } A_{1,i(1,k)} \text{ AND } \ldots \text{ AND } x_n \text{ is } A_{n,i(n,k)}$$
$$\text{THEN } g_{k,1} = z_{k,1} \text{ AND } \ldots \text{ AND } g_{k,c} = z_{k,c}$$

where $g_{k,i}$ is the discriminant function $g_i$ associated with rule $R_k$. The subscript $i(j,k)$ is an *input* index function showing which linguistic label is used for feature $x_j$ in rule $R_k$. The values $z_{k,j} \in \Re$ can be interpreted as "support" for class $\omega_j$ given by rule $R_k$ if the antecedent part is completely satisfied. If necessary, we can scale the support values in the interval $[0,1]$ by some order-preserving transformation $f : \Re \to (0,1)$, so that the discriminant functions are regarded as membership functions. For example, we can use $f(t) = 1/(1+\exp\{-t\})$. For $z_{k,j} \in [0,1]$, the vector $[z_{k,1}, \ldots, z_{k,c}]^T$ becomes a soft-class label defined over $\Omega$. Cordón *et al.* [1] distinguish between three types of fuzzy classification systems depending on the consequent.

1) Fuzzy rules with a class label in the consequent, e.g.,

$$R_k\text{: } \ldots \text{ THEN class is } \omega_{o(k)}$$

where $o(k)$ is an *output* indicator function giving the index of the class associated with rule $R_k$. In our general type model this translates to a $c$-dimensional binary output vector with 1 at $o(k)$ and 0, elsewhere.

2) Fuzzy rules with a class and a certainty degree in the consequent, e.g.,

$$R_k\text{: } \ldots \text{ THEN class is } \omega_{o(k)} \text{ with } z_{k,o(k)}.$$

This corresponds to $g_{k,1} = 0$ AND $\ldots$ AND $g_{k,o(k)} = z_{k,o(k)}, \ldots,$ AND $g_{k,c} = 0$.

3) Fuzzy rules with certainty degrees for all classes in the consequent, e.g.,

$$R_k\text{: } \ldots \text{ THEN } g_{k,1} = z_{k,1} \text{ AND } \ldots \text{ AND } g_{k,c} = z_{k,c}, \text{ usually } z_{k,i} \in [0,1].$$

The TSK fuzzy classifier is characterized by

1) The rule-base consisting of $M$ if-then rules of the general type.
2) The conjunction (AND connective): $\mathcal{A}_t$.
   The *firing strength* of rule $R_k$ is

$$\tau_k(\mathbf{x}) = \mathcal{A}_t\left\{\mu_{1,i(1,k)}(x_1), \ldots, \mu_{n,i(n,k)}(x_n)\right\}. \quad (2)$$

3) The calculation of the output.

Four popular fuzzy classifier variants are detailed below. They are subsequently used for proving various fuzzy if-then classifier properties. In all four definitions, $k = 1, \ldots, M$ is the index for the rules, $i = 1, \ldots, c$ is the index for the classes, and $j = 1, \ldots, n$ is the index for the features.

The **TSK1 classifier** is characterized by
- $z_{k,i} \in \{0,1\}, \sum_{i=1}^{c} z_{k,i} = 1$; (crisp labels)
- $\mathcal{A}_t$ is minimum;
- The $i$th TSK1 output is

$$g_i^{\mathrm{TSK1}}(\mathbf{x}) = \max_{k=1}^{M}\{z_{k,i} \cdot \tau_k(\mathbf{x})\} = \max_{k=1}^{M}\left\{z_{k,i} \cdot \min_{j=1}^{n}\left\{\mu_{j,i(j,k)}(x_j)\right\}\right\}. \quad (3)$$

The **TSK2 classifier** is characterized by
- $z_{k,i} \in \Re$;
- $\mathcal{A}_t$ is product;
- The $i$th TSK2 output is

$$g_i^{\mathrm{TSK2}}(\mathbf{x}) = \frac{\sum_{k=1}^{M} z_{k,i} \cdot \tau_k(\mathbf{x})}{\sum_{k=1}^{M} \tau_k(\mathbf{x})} = \frac{\sum_{k=1}^{M} z_{k,i} \prod_{j=1}^{n} \mu_{j,i(j,k)}(x_j)}{\sum_{k=1}^{M} \prod_{j=1}^{n} \mu_{j,i(j,k)}(x_j)} \quad (4)$$

The most popular version of the TSK2 classifier uses $z_{k,i} \in [0,1]$.

The **TSK3 classifier** is characterized by
- $z_{k,i} \in \{0,1\}, \sum_{i=1}^{c} z_{k,i} = 1$; (crisp labels)
- $\mathcal{A}_t$ is product;
- The $i$th TSK3 output is

$$g_i^{\mathrm{TSK3}}(\mathbf{x}) = \max_{k=1}^{M}\{z_{k,i} \cdot \tau_k(\mathbf{x})\} = \max_{k=1}^{M}\left\{z_{k,i} \cdot \prod_{j=1}^{n}\left\{\mu_{j,i(j,k)}(x_j)\right\}\right\}. \quad (5)$$

The only difference between TSK1 and TSK3 is $\mathcal{A}_t$. Both of them are common fuzzy classifier designs. They assign $\mathbf{x}$ to the most supported class, irrespective of how many rules vote for each class.

The **TSK4 classifier** differs from TSK2 only by the type of its consequent constants.

The **TSK4 classifier** is characterized by
- $z_{k,i} \in \{0,1\}, \sum_{i=1}^{c} z_{k,i} = 1$; (crisp labels)
- $\mathcal{A}_t$ is product;
- The $i$th TSK4 output is

$$g_i^{\mathrm{TSK4}}(\mathbf{x}) = \frac{\sum_{k=1}^{M} z_{k,i} \cdot \tau_k(\mathbf{x})}{\sum_{k=1}^{M} \tau_k(\mathbf{x})} = \frac{\sum_{k, z_{k,i}=1} \prod_{j=1}^{n} \mu_{j,i(j,k)}(x_j)}{\sum_{k=1}^{M} \prod_{j=1}^{n} \mu_{j,i(j,k)}(x_j)} \quad (6)$$

## III. Exact Match of the Classification Boundaries

Consider first an example. Shown in Fig. 1 are two classes generated with equal prior probabilities from a normal and a uniform distribution[1]

$$p(\mathbf{x}\,|\,\omega_1) \sim N([4,2]^T, I); \quad \text{and}$$
$$p(\mathbf{x}\,|\,\omega_2) = \begin{cases} \frac{1}{64}, & \text{if } \mathbf{x} \in [-1,7]^2, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

The Bayes-optimal classification boundary between the classes ($\omega_1$, shown with solid dots and $\omega_2$, with open dots) in the region $[-1,7] \times [-1,7]$ is a circle centered at the vector of the expectation of $\omega_1$, $[4,2]^T$, and with radius 2.15. The TSK4 classifier that produces this classification boundary uses the following rule base:

$R$: IF $x_1$ is about 4 AND $x_2$ is about 2
THEN $g_1(\mathbf{x}) = 1; \quad g_2(\mathbf{x}) = 0.0$.

$R$: IF $x_1$ is any AND $x_2$ is any
THEN $g_1(\mathbf{x}) = 0; \quad g_2(\mathbf{x}) = 0.0982$.

The membership function of "any" is 1 for any value of $x_j$ and the membership functions for $A_1$ = "about 4" and $A_2$ = "about 2" are, respectively,

$$\mu_1(x_1) = \exp\left\{-\frac{(x_1-4)^2}{2}\right\}$$
$$\mu_2(x_2) = \exp\left\{-\frac{(x_2-2)^2}{2}\right\}. \quad (8)$$

[1]By $I$ we denote the identity matrix.

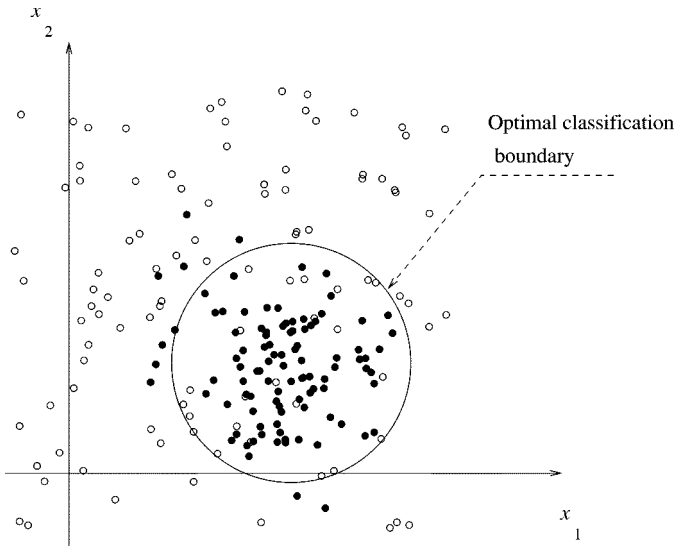


Fig. 1. Scatterplot of two classes with a normal and a uniform distribution in $\Re^2$. The circle depicts the Bayes-optimal classification boundary separating the two classes in $[-1,7]^2$.

The two firing strengths are ($\mathcal{A}_t$ is product for the TSK4 model)

$$\tau_1(\mathbf{x}) = \exp\left\{-\frac{(\mathbf{x}-[4,2]^T)^T(\mathbf{x}-[4,2]^T)}{2}\right\}$$
$$\tau_2(\mathbf{x}) = 1. \tag{9}$$

Then

$$g_1(\mathbf{x}) = \frac{1}{C}\exp\left\{-\frac{(\mathbf{x}-[4,2]^T)^T(\mathbf{x}-[4,2]^T)}{2}\right\} \tag{10}$$

$$g_2(\mathbf{x}) = \frac{1}{C}0.0982 \tag{11}$$

$C$ being the common denominator. According to the maximum membership rule, $\mathbf{x}$ will be assigned to class $\omega_1$ for $g_1(\mathbf{x}) > g_2(\mathbf{x})$, and to $\omega_2$, otherwise. To find the classification boundary, we solve $g_1(\mathbf{x}) = g_2(\mathbf{x})$, which leads to an equation of a circle with radius $R = 2.15$.

Fuzzy classifiers can match exactly a large class of classification boundaries in $\Re^2$. Here, we give the constructive proof of a lemma by Klawonn and Klement [7]. They consider a monotonic function $f(x)$ as the classification boundary between two classes in $\Re^2$. We extend the result for an arbitrary $f$ defined on some interval $[a_1, b_1] \subset \Re$.

*Lemma 1:* Let $a_1, b_1, a_2, b_2 \in \Re$ and $a_1 < b_1, a_2 < b_2$. Consider a two-class problem in $\Re^2$. Let $g(x,y) = 0$ be the equation of the classification boundary between the two classes $\omega_1$ and $\omega_2$. Assume that the function $g(x,y)$ can be represented as $g(x,y) = y - f(x)$ (and the boundary by the function $y = f(x)$, respectively) where $f(x)$ is defined on $[a_1, b_1]$ with a range $f([a_1, b_1]) \subseteq [a_2, b_2] \subset \Re$. There exists a **TSK1** classifier with two rules (a rule and its negation), which produces $g(x,y) = 0$ as the classification boundary.

To illustrate the proof, Fig. 2 depicts 300 points from two classes in $\Re^2$ and the classification boundary.

*Proof:* The classification boundary $g(x,y) = 0$ splits the region of interest $[a_1, b_1] \times [a_2, b_2] \subset \Re^2$ into two classification regions: $C_1$ (for $\omega_1$) and $C_2$ (for $\omega_2$). Without losing generality, assume that the points in $C_1$ get the positive value of $g$. We have to show that there exists a TSK1 classifier such that for any point $(x,y)$ in $\Re^2$ such that $g(x,y) > 0$, the classifier yields class $\omega_1$, and for any point $(x,y)$ such that $g(x,y) < 0$, the classifier yields class $\omega_2$. For points on the border ($g(x,y) = 0$), any class label is acceptable.

Consider a TSK1 classifier with the following membership functions (shown on the bottom two plots in Fig. 2 for the illustrative example):

$$\mu_{1,1}(x) = \frac{b_2 - f(x)}{b_2 - a_2}, \quad \text{and} \quad \mu_{1,2}(x) = 1 - \mu_{1,1}(x)$$

and

$$\mu_{2,1}(y) = \frac{y - a_2}{b_2 - a_2}, \quad \text{and} \quad \mu_{2,2}(y) = 1 - \mu_{2,1}(y)$$

and rule-base

$$R_1\text{: IF } x \text{ is } A_{1,1} \text{ AND } y \text{ is } A_{2,1}$$
$$\text{THEN } g_1 = 1 \text{ AND } g_2 = 0;$$
$$R_2\text{: IF } x \text{ is } A_{1,2} \text{ AND } y \text{ is } A_{2,2}$$
$$\text{THEN } g_1 = 0 \text{ AND } g_2 = 1.$$

Let $\mathbf{x} = [x,y]^T$ be an input vector. Using (3), the two discriminant functions are

$$g_1(x,y) = \min\{\mu_{1,1}(x), \mu_{2,1}(y)\} = \min\left\{\frac{b_2 - f(x)}{b_2 - a_2}, \frac{y - a_2}{b_2 - a_2}\right\} \tag{12}$$

and

$$g_2(x,y) = \min\{\mu_{1,2}(x), \mu_{2,2}(y)\} = \min\left\{1 - \frac{b_2 - f(x)}{b_2 - a_2}, 1 - \frac{y - a_2}{b_2 - a_2}\right\} = 1 - \max\left\{\frac{b_2 - f(x)}{b_2 - a_2}, \frac{y - a_2}{b_2 - a_2}\right\}. \tag{13}$$

According to the maximum membership rule, the TSK1 classifier assigns $\omega_1$ iff $g_1(x,y) \geq g_2(x,y)$, and class $\omega_2$, otherwise. To derive the class label, form

$$g_1(x,y) - g_2(x,y) = \min\{\mu_{1,1}(x), \mu_{2,1}(y)\} - 1 + \max\{\mu_{1,1}(x), \mu_{2,1}(y)\} = \mu_{1,1}(x) + \mu_{2,1}(y) - 1 \tag{14}$$

which leads to

$$g_1(x,y) - g_2(x,y) = \frac{y - f(x)}{(b_2 - a_2)}. \tag{15}$$

Assume that $g(x,y) > 0$, i.e., the input is from class $\omega_1$. By definition $b_2 > a_2$, and also for the input $[x,y]^T$, from $g(x,y) > 0$ follows that $y > f(x)$. Since $g_1(x,y) > g_2(x,y)$, TSK1 labels $[x,y]^T$ in class $\omega_1$. For points in class $\omega_2$ the numerator of (15) takes negative values, which completes the proof. ■

The problem with this lemma (and most of the theoretical results in general) is that it does not give you the tool to build the classifier. The proof is constructive, but notice that we *must know* the true discriminant function $g(x,y)$ to build the corresponding TSK1 classifier. If we knew $g(x,y)$, we might wish to use it directly. The value of this lemma is that it shows that TSK1 is versatile enough to fit any such classification boundary.

Unfortunately, this type of constructive proof cannot be extended beyond the two-dimensional case. Let $g(x_1, x_2, x_3) = 0$

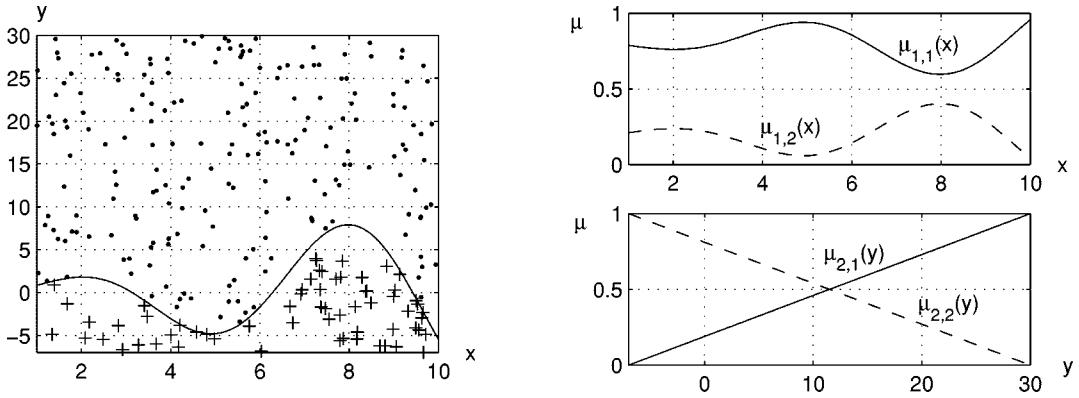


Fig. 2. Classification boundary approximated by TSK1 classifier. Points from $\omega_1$ are depicted with dots, and from class $\omega_2$ with pluses.

be a classification boundary in $\Re^3$. Klawonn and Klement [7] show that even if $g$ is a plane, the exact match by TSK1 (as in $\Re^2$) with a finite number of rules is impossible. They suggest other types of conjunction operations $\mathcal{A}_t$, e.g., product, and then a class of functions $g$ can be matched with a finite number of rules.

## IV. Universal Approximation with Fuzzy Classifiers

Many proofs exist that a type of fuzzy if-then systems can approximate to an arbitrary precision any continuous function $g : \Re^n \to \Re$ on a compact domain $U \subset \Re^n$ (universal approximation property) [10], [13], [18], [19], [21], [20], [22]. An extensive account of such methods is presented in [8]. We consider a **TSK2** model with Gaussians as the membership functions of the antecedents, centered at points $x_{j,k}$ with widths $\sigma_{j,k}$, $j = 1, \ldots, n, k = 1, \ldots, M$. In this model, each rule $R_k$ has its own "center" or *prototype* $\mathbf{x}_k$. Thus, each feature has $M$ (possibly different) fuzzy sets defined on its axis (one for each rule). The antecedent part of the rules for an input $\mathbf{x}$ can be interpreted as

$$\text{IF } \mathbf{x} \text{ is } like\ \mathbf{x}_k$$

with clauses

$$\text{IF } \ldots x_j \text{ is } like\ x_{j,k} \ldots$$

Hence, the fuzzy sets $A_{j,i(j,k)}$ are rule-specific and can be denoted simply by $A_{i,k}$. The membership functions for the antecedents are defined as

$$\mu_{j,k}(x_j) = a_{j,k} \exp\left\{-\frac{1}{2}\left(\frac{x_j - x_{j,k}}{\sigma_{j,k}}\right)^2\right\}, \quad 0 < a_{j,k} \leq 1. \quad (16)$$

Without losing generality consider a MISO TSK2 classifier approximating one discriminant function $g(\mathbf{x})$. Since there is only one output, denote by $z_k$ the constant in the consequent of rule $R_k$, $k = 1, \ldots, M$. The output of the TSK2 classifier is given by (4)

$$g(\mathbf{x}) = \frac{\sum_{k=1}^{M} z_k \prod_{j=1}^{n} \mu_{j,k}(x_j)}{\sum_{k=1}^{M} \prod_{j=1}^{n} \mu_{j,k}(x_j)}. \quad (17)$$

Let $\mathcal{G}$ be the class of all functions of the type (4), and let $d_\infty$ be a metric on $\mathcal{G}$:

$$d_\infty(g^1, g^2) = \sup_{\mathbf{x} \in \Re^n} \{|g^1(\mathbf{x}) - g^2(\mathbf{x})|\}, \quad g^1, g^2 \in \mathcal{G}. \quad (18)$$

The proof of the universal approximation of TSK2 classifier is based on the Stone–Weierstrass theorem (e.g., [15]).

*Stone–Weierstrass Theorem:* Let $Z$ be a set of real continuous functions on a compact set $U$, and let $C[U]$ be the set of all real continuous functions defined on $U$. If

1) $Z$ is an algebra, i.e., $Z$ is closed under addition, multiplication and scalar multiplication;
2) $Z$ separates points on $U$, i.e., for every $\mathbf{x}, \mathbf{y} \in U, \mathbf{x} \neq \mathbf{y}$, there exists a function $f \in Z$ such that $f(\mathbf{x}) \neq f(\mathbf{y})$;
3) $Z$ vanishes at no point on $U$, i.e., for each $\mathbf{x} \in U$ there exists $f \in Z$ such that $f(\mathbf{x}) \neq 0$;

then the uniform closure of $Z$ consists of all real functions on $U$, i.e., $(Z, d_\infty)$ is said to be dense in $(C[U], d_\infty)$.

By design, the set $\mathcal{G}$ of functions of interest is nonempty. Also, the requirement about the coefficients $a_{j,k}$ of the membership functions $(0 < a_{j,k} \leq 1)$ and the fact that Gaussians are used as the membership functions ensures that the denominator of (4) is always nonzero. It can be shown that $\mathcal{G}$ is an algebra, separates points of $U$, and does not vanish on any point of $U$ [18]. Then, for any continuous (discriminant) function $f(\mathbf{x})$ on a compact set $U \subset \Re^n$ and any $\epsilon > 0$, there exists a TSK2 classifier (with finite number of rules) with output $g(\mathbf{x})$ such that

$$\sup_{\mathbf{x} \in U} |f(\mathbf{x}) - g(\mathbf{x})| < \epsilon. \quad (19)$$

Similarly to the result about the exact match in the previous section, this theorem does not tell you how to build the classifier—it is only about the existence of it.

Next, typical TSK models can be fitted within some popular classifier designs. The following two theorems [9] establish this relationship.

Nearest neighbor classifier is an intuitive and simple nonparametric classification model [3]. According to it, the class label assigned to $\mathbf{x}$ is the label of the nearest object (we call it a *prototype*) from a set of pre-labeled prototypes (called the *reference set*). The term "nearest" implies a certain metric over $\Re^n$. The Euclidean metric is a typical choice.

*Theorem 1:* Let $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$ be a set of prototypes, $\mathbf{v}_i \in \Re^n, i = 1, \ldots, M$. A fuzzy **TSK3** classifier with $n$ inputs, $c$ outputs, $M$ rules and membership functions

$$\mu_{j,k}(x_j) = \exp\left\{-\frac{1}{2}(x_j - v_{j,k})^2\right\} \tag{20}$$

is equivalent to the nearest neighbor classifier (1-nn).

*Proof:* The $i$th output of the TSK3 classifier (corresponding to class $\omega_i$) is

$$g_i(\mathbf{x}) = \max_{k=1}^{M}\left\{z_{k,i} \cdot \prod_{j=1}^{n} \mu_{j,k}(x_j)\right\} = \max_{k=1}^{M}\left\{z_{k,i} \cdot \exp\left\{\frac{1}{2}(\mathbf{x} - \mathbf{v}_k)^T(\mathbf{x} - \mathbf{v}_k)\right\}\right\}. \tag{21}$$

Since $z_{k,i} \in \{0, 1\}$, the value of $g_i$ is determined by those elements of $\mathbf{V}$ whose corresponding rules have as the class label $\omega_i$. The terms corresponding to prototypes from other classes are multiplied by $z_{k,i} = 0$. The closest neighbor amongst the $\mathbf{v}_k$'s from $\omega_i$ will produce the highest value of the exponent in (21).

Comparing the $c$ discriminant functions, the winner in the maximum membership rule will be the function where the overall closest neighbor (prototype) has been found. This function will assign $\mathbf{x}$ to the class of its nearest neighbor. ■

Another classical nonparametric model with a great theoretical value is the Parzen classifier [3]. It has various layman-term interpretations—one of which is based on the notion of *potential*. Assume that we have again a set of $M$ labeled prototypes (in the Parzen model they are called *centers*) $\mathbf{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$, $\mathbf{c}_i \in \Re^n$. Each center generates a potential for the class it is labeled in. The potential has the highest intensity at the center and declines with the distance from the center. Each point in the feature space $\mathbf{x} \in \Re^n$ receives various potential levels: high intensities from the near by centers and small amounts from centers far away. We assume that the potential for each class at $\mathbf{x}$ (degree of membership of $\mathbf{x}$) is obtained as a superposition of the potentials generated by the centers from that class. Assuming a Gaussian model for potential distribution (Gaussian kernel $K_G$) at each center, the overall degree for class $\omega_i$ is [9]

$$g_i(\mathbf{x}) = \sum_{\mathbf{c} \in \omega_i} K_G\left(\frac{\mathbf{x} - \mathbf{c}}{h}\right), \quad i = 1, \ldots, c \tag{22}$$

where $h$ is a parameter of the kernel.[2] To derive the set of discriminant functions (22), we have assumed that the prior probabilities can be estimated by the proportion of elements from the respective classes, and have dropped from the discriminant functions all terms that do not depend on the class. The Gaussian kernel for statistically independent features $x_1, \ldots, x_n$ is

$$K_G\left(\frac{\mathbf{x} - \mathbf{c}}{h}\right) = \frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2h^2}(\mathbf{x} - \mathbf{c}_k)^T(\mathbf{x} - \mathbf{c}_k)\right\}. \tag{23}$$

It has been proven that the set of discriminant function (22) is asymptotically optimal [3]. This means that for an infinitely large set of independent identically distributed centers ($M \to \infty$) coming from the distribution of interest, the Parzen classifier will provide the minimal possible error rate (Bayes error rate). For this optimality to hold, the parameter $h$ has to satisfy $\lim_{M \to \infty} h(M) = 0$. It is worth mentioning that the optimality holds for a variety of kernels under mild conditions [4]. In practice, the Parzen classifier is often ignored because 1) using the whole labeled training set is cumbersome; 2) it is not always easy to select or extract a set of centers; and 3) the regularization parameter $h$ is difficult to guess or tune. The Parzen classifier, however, is usually very accurate. Its connection with radial basis networks has renewed the interest in it [2].

*Theorem 2:* Let $\mathbf{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_M\}$ be a set of centers, $\mathbf{c}_i \in \Re^n$, $i = 1, \ldots, M$. A fuzzy TSK4 classifier with $n$ inputs, $c$ outputs, $M$, rules and membership functions

$$\mu_{j,k}(x_j) = \exp\left\{-\frac{1}{2h^2}(x_j - c_{j,k})^2\right\} \tag{24}$$

is equivalent to the Parzen classifier.

*Proof:* The firing strength of rule $R_k$ is

$$\tau_k(\mathbf{x}) = \prod_{j=1}^{n} \mu_{j,k}(x_j) = \exp\left\{-\frac{1}{2h^2}(\mathbf{x} - \mathbf{c}_k)^T(\mathbf{x} - \mathbf{c}_k)\right\}. \tag{25}$$

$\tau_k(\mathbf{x})$ differs from the Gaussian kernel $K_G$ (23) by a constant which does not depend on the class label or the rule number $k$. Hence, substituting $K_G$ for $\tau$ and denoting by $C(\mathbf{x})$ the denominator (absorbing also the constant), the $i$th output of the TSK4 classifier is

$$g_i(\mathbf{x}) = C(\mathbf{x}) \sum_{k=1}^{M} z_{k,i} \cdot K_G\left(\frac{\mathbf{x} - \mathbf{c}_k}{h}\right). \tag{26}$$

Here, $z_{k,i}$ acts as an indicator function. By ignoring $C(\mathbf{x})$, which does not depend on $i$, we arrive at $c$ discriminant functions equivalent to those of the Parzen classifier. ■

Note that the universal approximation and the equivalences have been proven on fuzzy TSK classifiers that have a "prototype flavor." These are not the typical *transparent* architectures of fuzzy systems. In the models used for the proofs, the linguistic meaning of the fuzzy sets in the antecedents is obscured (prototype-based). Instead of having three or five linguistic labels for each variable, which is the comprehensible amount, there are $M$ labels per variable in this model.

## V. A Caveat: Fuzzy Classifiers Are Lookup Tables With Hyperbox Cells

Sometimes we may "overdo" the fuzzy classifier by designing a large rule base. Such fuzzy classifiers are expected to be more accurate but at the same time they become less transparent and approach the basic lookup table classifier. Then the whole point in introducing fuzzy semantics and inference becomes unclear.

We define the **lookup classifier** as an $n$-way table, with a finite number of cells, such that for any $\mathbf{x} \in \Re^n$, we can recover the class label from the table cell indexed by the intervals in which the components of $\mathbf{x}$ fall.

[2]For too small $h$'s, the discriminant functions become "prickly," and the generalization is not always good. For too large $h$'s, the discriminant functions are too smooth and can oversmooth useful classification boundaries. The difficulty is that the meaning of "small" and "large" is specific for each problem.

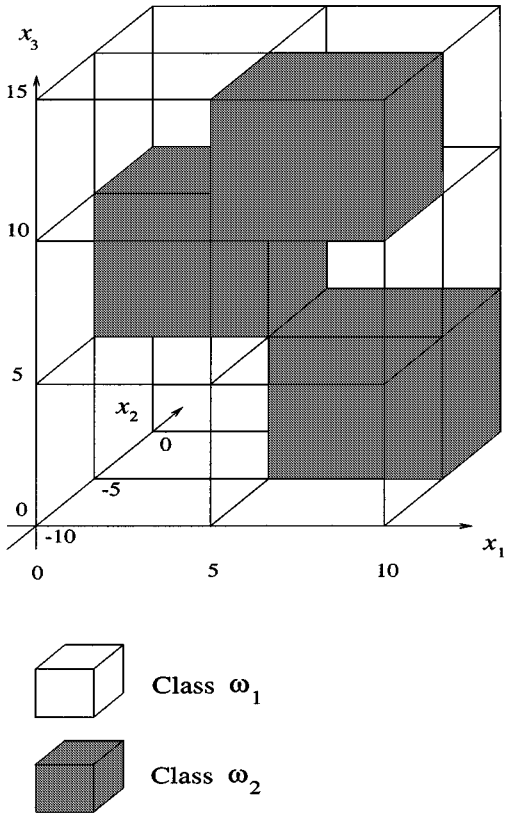


Fig. 3. Example of a lookup classifier.

For example, consider the lookup classifier in $\Re^3$ shown in Fig. 3. The classifier will label $\mathbf{x} = [6, -7, 11]$ in $\omega_2$ because $x_1 \in [5, 10]$, $x_2 \in [-5, -10]$, and $x_3 \in [10, 15]$. (Of course, the cells need not be of equal size.)

*Lemma 2:* Consider a TSK1 fuzzy classifier. Let $\Delta_j$ be the coincidence set of all membership functions defined on $x_j$, i.e., the set of all $x_j$'s such that one or more membership functions have the same value. We assume that for any $j = 1, \ldots, n$, $\Delta_j$ is a finite union of closed bounded intervals, possibly degenerate.[3]

Let $B_{j,l}$ be the subset of $\Re$ where $\mu_{j,l}(x_j)$ is the maximal amongst the $K_j$ membership functions defined on $x_j$, i.e.,

$$B_{j,l} = \{x_j, \mid \mu_{j,l}(x_j) > \mu_{j,t}(x_j), t = 1, \ldots, K_j, t \neq l\}. \tag{27}$$

Fig. 4 illustrates the regions $B_{j,l}$.

Let $\mathbf{B}_k \subset \Re^n$ be a hyperbox formulated as

$$\mathbf{B}_k = B_{1,i(1,k)} \times \cdots \times B_{n,i(n,k)} \tag{28}$$

and let $R_k$ be the corresponding rule in the rule-base. Then, the firing strength $\tau_k(\mathbf{x})$ satisfies

$$\tau_k(\mathbf{x}) > \tau_s(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{B}_k, \quad s = 1, \ldots, M, \quad s \neq k. \tag{29}$$

*Proof:* Let $\mathbf{x}$ be a point in $\mathbf{B}_k$. Then, for all $j = 1, \ldots, n$

$$\mu_{j,i(j,k)}(x_j) > \mu_{j,i(j,s)}(x_j), \quad \forall s = 1 \ldots, M, \quad s \neq k. \tag{30}$$

[3]Degenerate intervals account for point intersection of membership functions.

Then

$$\tau_k(\mathbf{x}) = \min_{j=1}^{n} \{\mu_{j,i(j,k)}(x_j)\} > \min_{j=1}^{n} \{\mu_{j,i(j,s)}(x_j)\} = \tau_s(\mathbf{x}), \quad \forall s = 1, \ldots, M, \quad s \neq k \tag{31}$$

which completes the proof. ■

This lemma shows that each rule $R_k$ defines a hyperbox in $\Re^n$ for any point of which the firing strength $\tau_k$ dominates the firing strengths of all other rules.

*Theorem 3:* Consider a TSK1 classifier. Assume that the coincidence sets $\Delta_j$, $j = 1, \ldots, n$ are finite unions of closed bounded intervals, possibly degenerate. If the TSK1 classifier contains all possible rules in its rule-base (all combinations of linguistic labels of the inputs), then it is a lookup classifier with rectangular cells, regardless of the shape of the membership functions used.

*Proof:* Let $\mathcal{B}$ denote the feature space covered by the fuzzy sets defined over the feature axes, i.e., for any $\mathbf{x} \in \mathcal{B}$ and for any $j = 1, \ldots, n$, there exists at least one membership function $\mu_{j,t(j)}$ such that $\mu_{j,t(j)} > 0$. (If for some $j$ all degrees of membership are zero, then $\mathbf{x}$ cannot be described by any linguistic term on $x_j$, and is therefore outside the region of interest). That is,

$$\mathcal{B} = \operatorname{supp}\left(\cup_{i=1}^{K_1} A_{1,i}(x_1) \times \cdots \times \cup_{i=1}^{K_n} A_{n,i}(x_n)\right) \subset \Re^n. \tag{32}$$

Let $\mu_{j,s(j)}(x_j)$, $j = 1, \ldots, n$, be the $n$ largest nonzero degrees of membership for some $\mathbf{x} \in \mathcal{B}$. Then, the rule with the highest firing strength for $\mathbf{x}$ has the antecedent

$$\text{IF } x_1 \text{ is } A_{1,s(1)} \text{ AND } \ldots \text{ AND } x_n \text{ is } A_{n,s(n)}.$$

Whatever the class label, this rule is in the rule base by definition, and therefore each $\mathbf{x}$ in $\mathcal{B}$ belongs in a box. Alternatively, $\mathbf{x}$ belongs in the boundary region $\Delta_B \subset \Re^n$ defined as

$$\mathbf{x} \in \Delta_B \Leftrightarrow \text{one or more components of } \mathbf{x} \text{ are in } \Delta_j. \tag{33}$$

Then

$$\mathcal{B} = \cup_{k=1}^{M} \mathbf{B}_k \cup \Delta_B. \tag{34}$$

The region $\mathcal{B}$ in Fig. 4 is composed of all boxes $\mathbf{B}_k$ and their borders.

Let $\mathbf{x} \in \mathbf{B}_k$. From the lemma, the firing strength $\tau_k(\mathbf{x})$ dominates the firing strengths of all other rules in $\mathbf{B}_k$ and the TSK1 outputs are

$$g_{o(k)}(\mathbf{x}) = \tau_k(\mathbf{x}) \tag{35}$$

and

$$g_i(\mathbf{x}) = \max_{s=1}^{M} \{z_{s,i} \cdot \tau_s(\mathbf{x})\} < \tau_k(\mathbf{x}), \quad i = 1, \ldots, c, \quad i \neq o(k). \tag{36}$$

Using the maximum membership rule, the class label assigned to $\mathbf{x}$ (an arbitrary point in box $\mathbf{B}_k$) is $\omega_{o(k)}$. For the points in $\Delta_B$, the classification decision can be made for any of the bordering boxes.

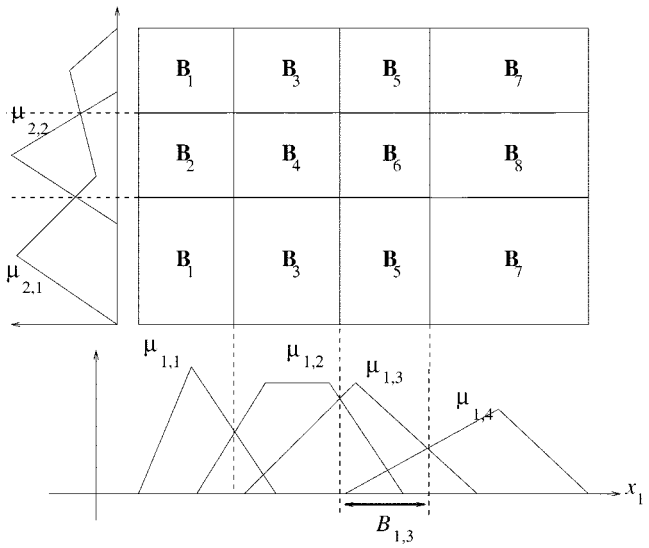


Fig. 4. Lookup classifier regions equivalent to the TSK1 regions in $\Re^2$ with all rules in the rule-base.

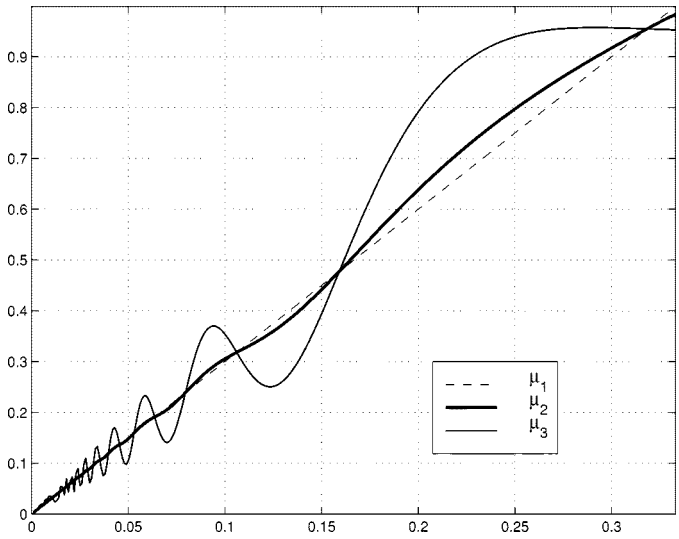


Fig. 5. Counterexample showing membership functions ($\mu_1$ and $\mu_2$ or $\mu_1$ and $\mu_3$) generating infinitely many boxes.

Taking together (34) and the result that each hyperbox produces one and only one class label, the TSK1 classifier is shown to be a lookup table regardless of the shape of the membership functions used. ■

Three points need further comments.

1) There can be more than one hyperbox per rule. This number is determined by how many times the membership functions of the antecedent clauses dominate the remaining membership functions on the respective features axes. In the example in Fig. 4, $\mu_{2,1}(x_2)$ is higher than $\mu_{2,2}(x_2)$ on two compact sets, thereby defining two boxes when used with any of the fuzzy sets on $x_1$.
2) Points inside the hyperboxes are covered by *one and only one* rule, and points on the borders, by more than one rule.
3) Equation (31) holds for any aggregation based on extended $t$-norms due to the monotonicity property of $t$-norms. Therefore, Theorem 3 also holds for the TSK3 classifier where $\mathcal{A}_t$ is product instead of minimum.

*Corollary 3.1:* TSK1 and TSK3 fuzzy if-then classifiers are universal approximators.

The idea of the proof: An integral-based norm $d_{\text{int}}$ can be defined, such that by using lookup tables (fuzzy classifiers in this capacity) we can approximate any integrable function and, thereby, any classification boundary $f(\mathbf{x}) = 0, \mathbf{x} \in \Re^n$ on a compact $U \subset \Re^n$ with an arbitrary precision.

The requirement about coincidence sets $\Delta_j$, $j = 1, \ldots, n$, merely ensures that the fuzzy classifier has a finite number of regions. Most of the classifiers do, e.g., with triangular, trapezoidal, or Gaussian membership functions. Can there be fuzzy classifiers with an infinite number of regions? Yes, as the following counterexample suggests. We borrow the example from [17] to show that there exists a setup where the TSK1 classifier has infinitely many regions, and therefore does not meet the definition of a lookup table.[4] Let $x \in [0, (1/3)]$ be the feature interval, and let the following two membership functions be defined on it:

$$\mu_1(x) = 3x \tag{37}$$

and

$$\mu_2(x) = 3x - x^2 \cdot \sin\left(\frac{1}{x}\right). \tag{38}$$

Assume $\mu_2(0) = 0$. When $\mu_2(x)$ approaches zero from the right, it oscillates around $\mu_1(x)$ with a frequency approaching infinity at zero. Therefore, since the two membership functions intersect infinitely many times, there should be infinitely many alternating regions where one of the functions dominates the other. Fig. 5 shows $\mu_1$ and $\mu_2$. For clarity, we also plotted

$$\mu_3(x) = 3x - x \cdot \sin\left(\frac{1}{x}\right) \tag{39}$$

with a higher amplitude of the oscillations.

To avoid situations like this, Von Schmidt and Klawonn [17] require that the membership functions have a local one-sided Taylor expansion everywhere. This means that for each point on the feature axis $x_j$, we can expand the membership function to the left and to the right in a vicinity of the point. The authors point out that most widely used membership functions satisfy this. In the condition, parts of Lemma 2 and Theorem 3 we do not specify what functions we use, but define $\Delta_j$'s to restrict their relationship. This is less specific than the assumption in [17] (so functions such as $\mu_2$ are still allowed provided there is no $\mu_1$ on the same feature axis). However, our assumption is more difficult to verify.

Another important consequence of Theorem 3 is that no improvement on the accuracy can be gained by altering the membership functions from one type to another as long as they intersect for the same value of the argument. It is not even helpful to switch to the product instead of minimum as $\mathcal{A}_t$. Thus, a fuzzy classifier with Gaussian membership functions and product as $\mathcal{A}_t$, and a classifier with triangular membership functions and

[4] I wish to thank reviewer D. for bringing this example to my attention.

TABLE I
FUZZY IF-THEN CLASSIFIER MODELS AND THE MAIN RESULTS

| Classifier model | TSK1 | TSK2 | TSK3 | TSK4 |
|---|---|---|---|---|
| $\mathcal{A}_t$ | minimum | product | product | product |
| Consequent | binary | real-valued | binary | binary |
| Universal approximation | look-up table Corollary 3.1 | Stone-Weierstrass Theorem[18] | look-up table Corollary 3.1 | Bayes-optimality |
| Prototype-related | No | Yes | Yes | Yes |
| **Main result** | Lemma 1 (Approximation of an arbitrary classification boundary in $\Re^2$) | Universal approximation | Theorem 1 (Equivalent to 1-nn) | Theorem 2 (Equivalent to Parzen classifier) |
| | Theorem 3 (Reduces to a look-up table) | | Theorem 3 (Reduces to a look-up table) | |

minimum as $\mathcal{A}_t$ can give exactly the same result on $\mathcal{B} \subset \Re^n$. Both fuzzy classifiers are identical to a lookup classifier on $\mathcal{B}$.

If we use a lookup table, however, $\mathcal{B}$ is restricted to the area of the table only. Fuzzy classifiers can smooth the edges between regions covered by the rules and the rest of the feature space. If the membership functions do not vanish anywhere on the feature space (e.g., Gaussians), there will be no blank spots on the feature space, i.e., the fuzzy system will be able to infer a class label for any point in $\Re^n$.

## VI. CONCLUSIONS

Table I summarizes the results brought about in this paper. The universal approximation, the equivalence with prototype-based statistical designs, and the lookup table isomorphism pull fuzzy classifiers out of their initial philosophical context. In the beginning, interpretability was perceived to be the most essential bonus of fuzzy classifiers. Now, little attention is paid to that, and accuracy renders the main concern.

How can we achieve good interpretability? One way is by using a small number of rules. The lemma by Klawonn and Klement [7] shows that a small number of rules can suffice. However, interpretability of the membership functions associated with these rules is not straightforward. A small number of rules usually implies specific and irregular shape of the membership functions (needed to achieve a good accuracy), which cannot be associated with linguistic labels on the feature axes. On the other hand, if we adopt simple models such as triangular or trapezoidal functions, we might need many of them per axis and the inevitable exponent explosion of rules to achieve high accuracy. Again, interpretability will not benefit.

Nonfuzzy (e.g., statistical) designs do not share this drawback because their functioning is *not supposed* to be interpretable. Therefore, the curse of dimensionality is not as acute for nonfuzzy classifiers as it is for fuzzy ones. The question is to what extent are fuzzy classifiers useful as *fuzzy*, and at which point do they turn into black boxes? As soon as interpretability is dismissed as a requirement to the system, fuzzy classifiers fall in the pool of numerous other designs, which are judged by their performance. Such designs include statistical classifiers and neural networks, to whom fuzzy classifiers are hardly the best rivals. Practice has shown so far that trying to reach the accuracy of a good nonfuzzy model by a fuzzy one is likely to require more time and resources than for building up the initial nonfuzzy classifier. (Usually the resulting fuzzy model is not transparent enough for the end user to verify and appreciate.) The results in this paper are not explicitly linked with the number of rules or interpretability of the classifiers (we assume that the rules are specified in advance): they apply to both small interpretable fuzzy designs to huge and opaque ones. We have shown that fuzzy classifiers are (theoretically!) able to achieve *any* accuracy. Fuzzy models are even richer than the few designs presented in this paper. For example, there are many fuzzy operators that can be plugged in the TSK classifier discussed here. If we forget about interpretability, the competition is between training algorithms for fuzzy and nonfuzzy designs. Statistical pattern recognition has established elegant theoretical models—many of which work with small data sets, too. Fuzzy pattern recognition has produced dozens of sophisticated *heuristic* training algorithms, yet, developing good (probably geometrically-driven) training algorithms is still an open problem.

Finally, picture a situation when you collaborate with an end-user who is prepared to accept a less accurate system but one which is fully comprehensible to them. Nonfuzzy classifiers will not be a good choice for that, so the fuzzy toolkit should be kept on the shelf.

## ACKNOWLEDGMENT

The author would like to thank the four anonymous reviewers for their valuable comments and suggestions, and Prof. T. Porter and Prof. R. Brown, School of Informatics, University of Wales, Bangor, U.K., for the very helpful discussions.

## REFERENCES

[1] O. Cordón, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *Int. J. Approx. Reason.*, vol. 20, no. 1, pp. 21–45, 1999.

[2] R. L. Coultrip and R. H. Granger, "Sparse random networks with LTP learning rules approximate Bayes classifiers via Parzen's method," *Neural Networks*, vol. 7, pp. 463–476, 1994.

[3] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[4] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Orlando, FL: Academic, 1972.

[5] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms," *Fuzzy Sets Syst.*, vol. 65, pp. 237–253, 1994.

[6] ——, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 260–270, 1995.

[7] F. Klawon and P. E. Klement, "Mathematical analysis of fuzzy classifiers," in *Lecture Notes in Computer Science*, 1997, vol. 1280, pp. 359–370.

[8] V. L. Kreinovich, C. G. Mouzouris, and H. T. Nguyen, "Fuzzy rule based modeling as a universal approximation tool," in *Fuzzy Systems: Modeling and Control*, H. T. Nguyen and M. Sugeno, Eds. Boston, MA: Kluwer, 1998, pp. 135–196.

[9] L. I. Kuncheva, "On the equivalence between fuzzy and statistical classifiers," *Int. J. Uncertainty, Fuzziness, Knowl.-Based Syst.*, vol. 4.3, pp. 245–253, 1996.

[10] Z.-H. Mao, Y.-D. Li, and X.-F. Zhang, "Approximation capability of fuzzy systems using translations and dilations of one fixed function as membership functions," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 3, pp. 468–473, 1997.
[11] F. Masulli, F. Casalino, and F. Vannucci, "Bayesian properties and performances of adaptive fuzzy systems in pattern recognition problems," in *ICANN'94*, Sorrento, Italy, 1994, pp. 189–192.
[12] D. Nauck and R. Kruse, "A neuro-fuzzy method to learn fuzzy classification rules from data," *Fuzzy Sets Syst.*, vol. 89, pp. 277–288, 1997.
[13] R. Rovatti, "Fuzzy piecewise multilinear and liecewise linear systems as universal approximators," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 2, pp. 235–249, 1998.
[14] M. Russo, "FuGeNeSys—A fuzzy genetic neural system for fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 3, pp. 373–388, 1998.
[15] G. F. Simmons, *Introduction to Topology and Modern Analysis*, Tokyo, Japan: McGraw-Hill, 1963.
[16] C.-T. Sun and J.-S. Jang, "A neuro-fuzzy classifier and its applications," in *2nd IEEE Int. Conf. Fuzzy Syst.*, San Francisco, CA, 1993, pp. 94–98.
[17] B. von Schmidt and F. Klawonn, "Fuzzy max-min classifiers decide locally on the basis of two attributes," *Mathware Soft Comput.*, 1999.
[18] L.-X. Wang, "Fuzzy systems are universal approximators," *Proc. IEEE Int. Conf. Fuzzy Systems*, pp. 1163–1170, 1992.
[19] L. X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation and orthogonal least squares learning," *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 807–814, 1992.
[20] H. Ying, "General SISO Takagi-Sugeno fuzzy systems with linear rue consequent are universal approximators," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 4, pp. 582–587, 1998.
[21] ——, "Sufficient conditions on uniform approximation of multivariate functions by general Takagi-Sugeno fuzzy systems with linear rule consequent," *IEEE Trans. Syst., Man, Cybern. A.*, vol. 28, pp. 515–520, July 1998.
[22] X.-J. Zeng and M. G. Singh, "Approximation properties of fuzzy systems generated by the min inference," *IEEE Trans. Syst., Man, Cybern. B*, vol. 26, pp. 187–193, Feb. 1996.

**Ludmila I. Kuncheva** (M'99) received the M.Sc. degree from the Technical University, Sofia, Bulgaria, in 1982, and the Ph.D. degree from the Bulgarian Academy of Sciences, Sofia, in 1987.

Until 1997, she worked at the Central Laboratory of Biomedical Engineering, Bulgarian Academy of Sciences, as a Senior Research Associate. She is currently a Lecturer at the School of Mathematics, University of Wales, Bangor, U.K. Her interests include pattern recognition, neural networks, fuzzy classifiers, prototype classifiers, and multiple classifier systems.