Mathematical Analysis of Fuzzy Classifiers

Frank Klawonn[1] and Erich–Peter Klement[2]

[1] Fachbereich Elektrotechnik und Informatik, Fachhochschule Ostfriesland, Constantiaplatz 4, D-26723 Emden, Germany
[2] Fuzzy Logic Laboratorium Linz
Johannes Kepler University
A–4040 Linz, Austria

Abstract. We examine the principle capabilities and limits of fuzzy classifiers that are based on a finite set of fuzzy if–then rules like they are used for fuzzy controllers, except that the conclusion of a rule specifies a discrete class instead of a (fuzzy) real output value. Our results show that in the two–dimensional case, for classification problems whose solutions can only be solved approximately by crisp classification rules, very simple fuzzy rules provide an exact solution. However, in the multi–dimensional case, even for linear separable problems, max–min rules are not sufficient.

1 Introduction

Fuzzy controllers are well examined as function approximators. Piecewise monotone functions of one variable can be exactly reproduced by a fuzzy controller [1, 9] and for the multi–dimensional case fuzzy controllers are known to be universal approximators [2, 6, 12]. Although a lot of approaches for automatically learning fuzzy classifiers are proposed in the literature (see for instance [3, 4, 5, 10, 11, 13]), they are usually evaluated only on an experimental basis. A theoretical analysis of the principal capabilities of fuzzy classifiers aiming at the assignment of discrete classes to input vectors, is still lacking. A natural question concerning fuzzy classification rules is, whether they have any advantage over crisp classification rules when in the end for an input vector a unique assignment to one class has to be made. We will provide a positive answer to this question in the sense that already in the two–dimensional case fuzzy classification rules can solve problems for which only approximate solutions can be constructed on the basis of crisp classification rules. This paper is devoted to the question what kind of classification problems are solvable in principle by fuzzy classifiers using if-then rules. We do not discuss techniques for actually constructing suitable if-then rules from data.

The paper is organized as follows. Section 2 provides the formal definition of the type of fuzzy classifiers we are examining. In Section 3 we demonstrate that in the two–dimensional case quite general classification problems can be solved, whereas for higher dimensional problems simple max–min rules must fail. As shown in Section 4 this can be amended by using other operations than max or min.

2 Formal Framework

Let us briefly introduce the formal framework we are considering. We consider fuzzy classification problems of the following form. There are p real variables $x_1, \ldots, x_p$ with underlying domains $X_i = [a_i, b_i]$, $a_i < b_i$. There is a finite set $\mathcal{C}$ of classes and a partial mapping

$$\text{class} : X_1 \times \ldots \times X_p \longrightarrow \mathcal{C}$$

that assigns classes to some, but not necessarily to all vectors $(x_1, \ldots, x_p) \in X_1 \times \ldots \times X_p$.

The aim is to find a fuzzy classifier that solves the classification problem. The fuzzy classifier is based on a finite set $\mathcal{R}$ of rules of the form $R \in \mathcal{R}$:

R: If x_1 is $\mu_R^{(1)}$ and $\ldots$ and x_p is $\mu_R^{(p)}$ then class is C_R.

$C_R \in \mathcal{C}$ is one of the classes. The $\mu_R^{(i)}$ are assumed to be fuzzy sets on X_i, i.e. $\mu_R^{(i)} : X_i \longrightarrow [0,1]$. In order to keep the notation simple, we incorporate the fuzzy sets $\mu_R^{(i)}$ directly in the rules. In real systems one would replace them by suitable linguistic values like *positive big*, *approximately zero*, etc. and associate the linguistic value with the corresponding fuzzy set.

In Section 3, where we present our main results, we restrict ourselves to max–min rules, i.e., we evaluate the conjunction in the rules by the minimum and aggregate the results of the rules by the maximum. Therefore, we define

$$\mu_R(x_1, \ldots, x_p) \;=\; \min_{i \in \{1, \ldots, p\}} \left\{ \mu_R^{(i)}(x_i) \right\} \tag{1}$$

as the degree to which the premise of rule R is satisfied.

$$\mu_C^{(\mathcal{R})}(x_1, \ldots, x_p) \;=\; \max\{\mu_R(x_1, \ldots, x_p) \mid C_R = C\} \tag{2}$$

is the degree to which the vector $(x_1, \ldots, x_p)$ is assigned to class $C \in \mathcal{C}$. The defuzzification – the final assignment of a unique class to a given vector $(x_1, \ldots, x_p)$ – is carried out by the mapping

$$\mathcal{R}(x_1, \ldots, x_p) \;=\; \begin{cases} C & \text{if } \mu_C^{(\mathcal{R})}(x_1, \ldots, x_p) > \mu_D^{(\mathcal{R})}(x_1, \ldots, x_p) \\ & \text{for all } D \in \mathcal{C}, D \neq C \\ \textit{unknown} \notin \mathcal{C} & \text{otherwise.} \end{cases}$$

This means that we finally assign the class C to the vector $(x_1, \ldots, x_p)$ if the fuzzy rules assign the highest degree to class C for vector $(x_1, \ldots, x_p)$. If there are two or more classes that are assigned the maximal degree by the rules, then we refrain from a classification and indicate it by the symbol *unknown*. Note that we use the same letter $\mathcal{R}$ for the rule base and the induced classification mapping.

Finally,

$$\mathcal{R}^{-1}(C) \;=\; \{(x_1, \ldots, x_p) \mid \mathcal{R}(x_1, \ldots, x_p) = C\}$$

denotes the set of vectors that are assigned to class C by the rules (after defuzzification).

3 Max–Min Rules

Let us first take a look at crisp classification rules in the sense that the fuzzy sets $\mu_R^{(i)}$ are assumed to be characteristic functions of crisp sets, say intervals. Then it is obvious that in the two–dimensional case each rule assigns those inputs to the class appearing in the conclusion of the rule that are in the rectangle that is induced by the two intervals appearing as characteristic functions in the premise of the rule.

A classification problem with two classes that are separated by a hyperplane, i.e. a line in the two–dimensional case, is called linear separable. Obviously, a linear separable classification problem can be solved only approximately by crisp classification rules by approximating the separating line by a step function (see Figure 1).

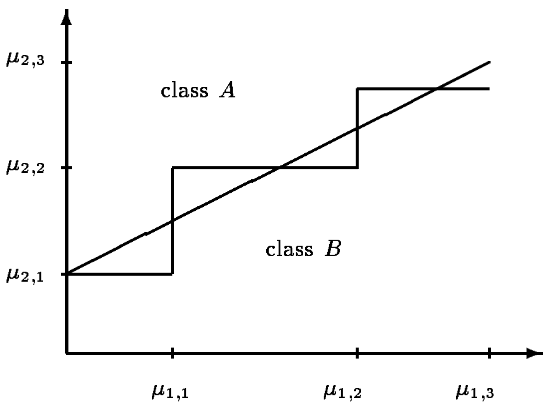

Fig. 1. Approximate solution of a linear separable classification problem by crisp classification rules

For fuzzy classification rules the situation is much better. The following lemma and its corollary show that in the two–dimensional case classification problems with two classes that are separated by a piecewise monotone function can be solved exactly using fuzzy classification rules.

Lemma 1. *Let $f : [a_1, b_1] \longrightarrow [a_2, b_2]$ $(a_i < b_i)$ be a monotone function. Then there is a finite set $\mathcal{R}$ of classification rules to classes P and N such that*

$$\mathcal{R}^{-1}(P) = \{(x, y) \in [a_1, b_1] \times [a_2, b_2] \mid f(x) > y\},$$
$$\mathcal{R}^{-1}(N) = \{(x, y) \in [a_1, b_1] \times [a_2, b_2] \mid f(x) < y\}.$$

Proof. Let us abbreviate $X = [a_1, b_1]$, $Y = [a_2, b_2]$.

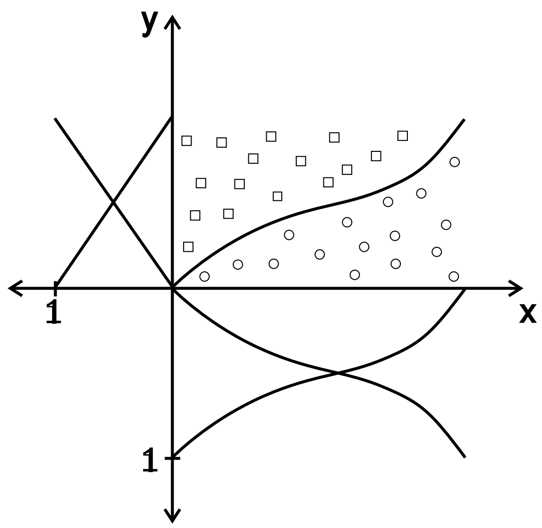

Fig. 2. The fuzzy sets for the classification rules

Define the fuzzy sets

$$\begin{aligned}
\mu_1 : X &\longrightarrow [0,1], \quad x \mapsto \tfrac{b_2 - f(x)}{b_2 - a_2},\\
\mu_2 : X &\longrightarrow [0,1], \quad x \mapsto \tfrac{f(x) - a_2}{b_2 - a_2} = 1 - \mu_1(x),\\
\nu_1 : Y &\longrightarrow [0,1], \quad y \mapsto \tfrac{y - a_2}{b_2 - a_2},\\
\nu_2 : Y &\longrightarrow [0,1], \quad y \mapsto \tfrac{b_2 - y}{b_2 - a_2} = 1 - \nu_1(y).
\end{aligned}$$

The fuzzy sets are illustrated in Figure 2. The rule base consists of the two rules:

R_1: If x is μ_1 and y is ν_1 then class is N.

R_2: If x is μ_2 and y is ν_2 then class is P.

It is easy to verify that these rules solve the classification problem. □

Note that the proof is based on a very similar technique as the proof for constructing a fuzzy controller for rebuilding a function with one argument [1]. It is obvious that we can extend the result of this lemma to piecewise monotone functions, simply by defining corresponding fuzzy sets on the intervals where the class separating function is monotone (see Figure 3) and defining corresponding rules for each of these intervals so that we have the following corollary.

Corollary 2. *Let* $f : [a_1, b_1] \longrightarrow [a_2, b_2]$ *(*$a_i < b_i$*) be a piecewise monotone function. Then there is a finite set* $\mathcal{R}$ *of classification rules to classes* P *and* N

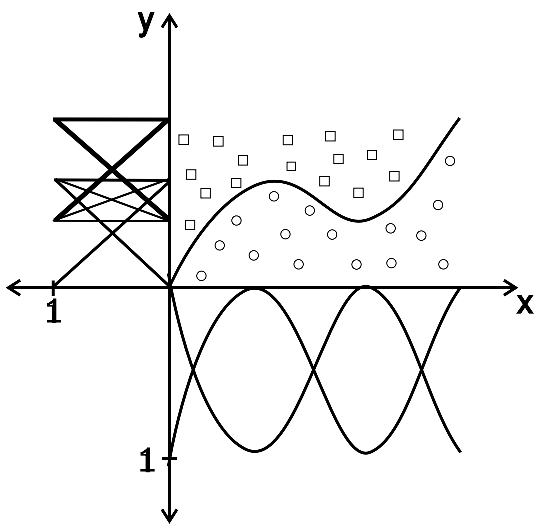

Fig. 3. The fuzzy sets for the classification rules of a piecewise monotone function

such that

$$\mathcal{R}^{-1}(P) = \{(x,y) \in [a_1,b_1] \times [a_2,b_2] \mid f(x) > y\},$$
$$\mathcal{R}^{-1}(N) = \{(x,y) \in [a_1,b_1] \times [a_2,b_2] \mid f(x) < y\}.$$

A direct consequence of Lemma 1 and its proof is that we can solve two–dimensional linear separable classification problems with only two fuzzy classification rules incorporating simple triangular membership functions so that we are in a much better situation than in the case of crisp classification rules. However, the result cannot be extended to more than two dimensions since the following theorem shows that even three–dimensional linear separable classification problems cannot be solved (evaluating the rules by the max–min schema).

Theorem 3. *Let* $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$, $(x_1, x_2, x_3) \mapsto x_1 + x_2 + x_3$. *For fixed* $X_i = [a_i, b_i]$, $a_i < b_i$, *(*$i = 1,2,3$*), denote*

$$P = \{(x_1,x_2,x_3) \in X_1 \times X_2 \times X_3 \mid f(x_1,x_2,x_3) > 0\},$$

$$N = \{(x_1,x_2,x_3) \in X_1 \times X_2 \times X_3 \mid f(x_1,x_2,x_3) < 0\}.$$

For any choice of the intervals X_1, X_2, X_3 *and any finite set of classification rules* $\mathcal{R}$ *with classes* CP *and* CN*, at least one of the following conditions is not satisfied:*

(i) $P \neq \emptyset$ *and* $N \neq \emptyset$
(ii) The sets $\mu_R^{(i)}$ *(*$i \in \{1,2,3\}$, $R \in \mathcal{R}$*) are piecewise monotone and continuous.*
(iii) $\mathcal{R}(CP) = P$ *and* $\mathcal{R}(CN) = N$.

Proof. Assume the conditions (i), (ii), and (iii) could be satisfied simultaneously. Then we can choose X_1, X_2, X_3 and $\mathcal{R}$ in such a way that the cardinality of $\mathcal{R}$ is minimal, i.e., no matter how we choose X_1', X_2', X_3' (satisfying (i)), a rule base $\mathcal{R}'$ that guarantees for (ii) and (iii) will contain at least as many rules as $\mathcal{R}$.

The continuity of the fuzzy sets enforces the continuity of $\mu_{CP}^{(\mathcal{R})}$ and $\mu_{CN}^{(\mathcal{R})}$. Therefore, we have for all $(x_1, x_2, x_3) \in X_1 \times X_2 \times X_3$ satisfying $f(x_1, x_2, x_3) = 0$ that $\mu_{CP}^{(\mathcal{R})}(x_1, x_2, x_3) = \mu_{CN}^{(\mathcal{R})}(x_1, x_2, x_3)$ holds.

The set

$$Z = \{(x_1, x_2, x_3) \in X_1 \times X_2 \times X_3 \mid f(x_1, x_2, x_3) = 0\}$$

– the boundary between the classes P and N – is the intersection of the plane $x_1 + x_2 + x_3 = 0$ and the cube $X_1 \times X_2 \times X_3$. According to condition (i), part of the plane lies in the interior of the cube.

Assume, there is a point $(x_1, x_2, x_3) \in Z$ in the interior of $X_1 \times X_2 \times X_3$ for which the rules do not fire to the same degree, i.e., there is a rule $S \in \mathcal{R}$ s.t.

$$\mu_S(x_1, x_2, x_3) \ < \ \max_{R \in \mathcal{R}}\{\mu_R(x_1, x_2, x_3)\}.$$

Then there is an entourage $X_1' \times X_2' \times X_3' \subseteq X_1 \times X_2 \times X_3$ of (x_1, x_2, x_3) s.t. for all $(z_1, z_2, z_3) \in X_1' \times X_2' \times X_3'$

$$\mu_S(z_1, z_2, z_3) \ < \ \max_{R \in \mathcal{R}}\{\mu_R(z_1, z_2, z_3)\}$$

holds. But this means that the rule base $\mathcal{R} \backslash \{S\}$ satisfies (i), (ii), and (iii) on $X_1' \times X_2' \times X_3'$, which is a contradiction to the minimality of $\mathcal{R}$. Thus we have for all $(x_1, x_2, x_3) \in Z$

$$\mu_R(x_1, x_2, x_3) \ = \ \mu_S(x_1, x_2, x_3) \tag{3}$$

for all $R, S \in \mathcal{R}$.

Therefore, we can define the function

$$g : Z \longrightarrow [0, 1], \qquad (x_1, x_2, x_3) \mapsto \mu_R(x_1, x_2, x_3)$$

independent of the choice of $R \in \mathcal{R}$.

The fuzzy sets $\mu_R^{(i)}$ are piecewise monotone, i.e., for each $\mu_R^{(i)}$ there are only finitely many points which have no entourage in which $\mu_R^{(i)}$ is monotone. Let z be such a point. For each z and $\mu_i^{(R)}$ we obtain a plane

$$H_{z, \mu_R^{(i)}} \ = \ \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_i = z\}.$$

Since $H_{z, \mu_R^{(i)}}$ is a plane parallel to two axes, whereas

$$\hat{Z} \ = \ \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 0\}$$

is not axes–parallel, their intersection is a line. Thus, if we consider the set

$$\hat{Z} \Big\backslash \left(\bigcup_{R,i,z} H_{z,\mu_R^{(i)}} \right),$$

we simply cut out a finite set of lines from the plane $\hat{Z}$. Therefore, and since $\hat{Z}$ has a non–empty intersection with the interior of the cube $X_1 \times X_2 \times X_3$,

$$\left(\hat{Z} \cap (X_1 \times X_2 \times X_3)\right) \Big\backslash \left(\bigcup_{R,i,z} H_{z,\mu_R^{(i)}} \right)$$

is non–empty. Choose an element (z_1, z_2, z_3) from this set which lies in the interior of $X_1 \times X_2 \times X_3$ and an entourage

$$X_1' \times X_2' \times X_3' \;=\; [z_1 - \delta, z_1 + \delta] \times [z_2 - \delta, z_2 + \delta] \times [z_3 - \delta, z_3 + \delta] \subseteq X_1 \times X_2 \times X_3$$

that does not contain any of the points of

$$\bigcup_{R,i,z} H_{z,\mu_R^{(i)}}.$$

We now prove that g is constant on $Z \cap (X_1' \times X_2' \times X_3')$. We can write any point

$$(x_1, x_2, x_3) \in Z \cap (X_1' \times X_2' \times X_3')$$

in the form $(z_1 + \varepsilon_1, z_2 + \varepsilon_2, z_3 + \varepsilon_3)$ with $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = 0$ and $|\varepsilon_i| \leq \delta$.

Let us consider the point

$$(z_1 + \varepsilon_1, z_2 + \varepsilon_2, z_3 + \varepsilon_3) \in Z \cap (X_1' \times X_2' \times X_3').$$

Since a variation of the value z_3 in (z_1, z_2, z_3) leads to a change in the classification (from *unknown* to CP or CN), there has to be a rule $R \in \mathcal{R}$ with

$$\mu_R(z_1, z_2, z_3) \;=\; \mu_R^{(3)}(z_3).$$

Otherwise a (sufficiently small) variation of z_3 would neither change the value $\mu_{CP}(z_1, z_2, z_3)$ nor $\mu_{CN}(z_1, z_2, z_3)$, resulting in the wrong classification *unknown*.

This implies

$$\begin{aligned} g(z_1 + \varepsilon_1, z_2 - \varepsilon_1, z_3) &= \mu_R(z_1 + \varepsilon_1, z_2 - \varepsilon_1, z_3) \\ &\leq \mu_R^{(3)}(z_3) \\ &= \mu_R(z_1, z_2, z_3) \\ &= g(z_1, z_2, z_3). \end{aligned}$$

The same argument as above guarantees the existence of a rule $S \in \mathcal{R}$ with

$$\mu_S(z_1 + \varepsilon_1, z_2 - \varepsilon_1, z_3) \;=\; \mu_S^{(1)}(z_1 + \varepsilon_1).$$

Thus we obtain

$$\begin{aligned} g(z_1+\varepsilon_1, z_2+\varepsilon_2, z_3+\varepsilon_3) &= \mu_S(z_1+\varepsilon_1, z_2+\varepsilon_2, z_3+\varepsilon_3) \\ &\le \mu_S^{(1)}(z_1+\varepsilon_1) \\ &= \mu_S(z_1+\varepsilon_1, z_2-\varepsilon_1, z_3) \\ &= \mu_R(z_1+\varepsilon_1, z_2-\varepsilon_1, z_3) \\ &\le g(z_1, z_2, z_3). \end{aligned}$$

By exchanging the roles of $(z_1+\varepsilon_1, z_2-\varepsilon_1, z_3)$ and (z_1, z_2, z_3), we can prove that

$$g(z_1, z_2, z_3) \ \le \ g(z_1+\varepsilon_1, z_2+\varepsilon_2, z_3+\varepsilon_3)$$

also holds so that g has to be constant on $Z \cap (X_1' \times X_2' \times X_3')$, say $g(x_1, x_2, x_3) = \alpha$ for all $(x_1, x_2, x_3) \in Z \cap (X_1' \times X_2' \times X_3')$.

Assume there exist $i \in \{1, 2, 3\}$, $R \in \mathcal{R}$, and $|\varepsilon| < \delta$ s.t.

$$\mu_R^{(i)}(z_i+\varepsilon) \ < \ \alpha.$$

Without loss of generality let $i = 1$. This leads to the contradiction

$$\alpha \ = \ \mu_R(z_1+\varepsilon, z_2-\varepsilon, z_3) \ \le \ \mu_R^{(1)}(z_1+\varepsilon) \ < \ \alpha.$$

Thus we have for all $i \in \{1, 2, 3\}$, for all $x \in X_i'$, and for all $R \in \mathcal{R}$

$$\mu_R^{(i)}(x) \ \ge \ \alpha. \tag{4}$$

Since a variation of the value z_1 in (z_1, z_2, z_3) leads to a change in classification, there must be an $\varepsilon > 0$ and a rule $R \in \mathcal{R}$ s.t.

$$\mu_R(z_1+\varepsilon, z_2, z_3) \ \neq \ \mu_R(z_1, z_2, z_3) \ = \ \alpha.$$

By inequality (4) we obtain

$$\begin{aligned} \min\{\mu_R^{(1)}(z_1+\varepsilon), \mu_R^{(2)}(z_2), \mu_R^{(3)}(z_3)\} &= \mu_R(z_1+\varepsilon, z_2, z_3) \\ &> \alpha \\ &= \mu_R(z_1, z_2, z_3) \\ &= \min\{\mu_R^{(1)}(z_1), \mu_R^{(2)}(z_2), \mu_R^{(3)}(z_3)\}. \end{aligned}$$

Thus we have

$$\mu_R(z_1, z_2, z_3) \ = \ \mu_R^{(1)}(z_1) \ < \ \min\{\mu_R^{(2)}(z_2), \mu_R^{(3)}(z_3)\}.$$

Taking the monotonicity of $\mu_R^{(1)}$ into account, we derive that $\mu_R^{(1)}$ has to be increasing. The continuity of the fuzzy sets $\mu_R^{(i)}$ guarantees the existence of $\tilde{\varepsilon} > 0$ s.t.

$$\begin{aligned} \alpha &= \mu_R^{(1)}(z_1) \\ &< \mu_R^{(1)}(z_1+\tilde{\varepsilon}) \\ &\le \min\{\mu_R^{(2)}(z_2-\tilde{\varepsilon}, \mu_R^{(3)}(z_3)\} \end{aligned}$$

which leads to the final contradiction

$$\begin{aligned}\alpha &= \mu_R(z_1+\tilde{\varepsilon}, z_2-\tilde{\varepsilon}, z_3)\\ &= \mu_R^{(1)}(z_1+\tilde{\varepsilon})\\ &> \alpha.\end{aligned}$$

□

4 Other t–Norms and t–Conorms

The fact that linear separable higher dimensional classification problems cannot be solved with fuzzy classification rules can be amended by replacing the maximum by another t–conorm (an associative, commutative, monotone increasing binary operation with unit 0 on the unit interval, see for instance [7]), namely the bounded sum, or by replacing the minimum by another t–norm (an associative, commutative, monotone increasing binary operation with unit 1 on the unit interval), namely the Łukasiewicz–t–norm. The following two theorems show that it is sufficient to replace either the minimum or the maximum by a suitable t–norm, respectively t–conorm. The function f appearing in these theorems describes an arbitrary hyperplane that separates the two classes to be distinguished by the classifiers.

Theorem 4. *Let* $f : [a_1, b_1] \times \ldots \times [a_p, b_p] \longrightarrow \mathbb{R}$, $(x_1, \ldots, x_p) \mapsto c + \sum_{i=1}^{p} c_i x_i$ $(a_i < b_i)$. *Then there is a finite set* $\mathcal{R}$ *of classification rules to classes* P *and* N *such that*

$$\mathcal{R}^{-1}(P) = \{(x_1, \ldots, x_p) \in [a_1, b_1] \times \ldots \times [a_p, b_p] \mid f(x_1, \ldots, x_p) > 0\}, \quad (5)$$

$$\mathcal{R}^{-1}(N) = \{(x_1, \ldots, x_p) \in [a_1, b_1] \times \ldots \times [a_p, b_p] \mid f(x_1, \ldots, x_p) < 0\} \quad (6)$$

when the minimum in (1) *is replaced by an arbitrary t–norm and the maximum in* (2) *is replaced by the bounded sum, i.e.*

$$\mu_C^{(\mathcal{R})}(x_1, \ldots, x_p) \;=\; \min\left\{ \sum_{R \in \mathcal{R}: C_R = C} \mu_R(x_1, \ldots, x_p), 1 \right\}.$$

Proof. Without loss of generality let $c \geq 0$. (Otherwise consider the function $-f$ and exchange the rules for the classes P and N.) Without loss of generality, let

$$\alpha \;=\; \max\left\{ c, \max_{i \in \{1,\ldots,p\}} \left\{ \sup_{x \in [a_i, b_i]} \{|c_i x|\} \right\} \right\} \;<\; \frac{1}{2p}.$$

Otherwise (5) and (6) could be defined equivalently by

$$\mathcal{R}^{-1}(P) = \{(x_1,\ldots,x_p) \in [a_1,b_1]\times\ldots\times[a_p,b_p] \mid \frac{1}{2p\alpha}f(x_1,\ldots,x_p) > 0\}$$

$$\mathcal{R}^{-1}(N) = \{(x_1,\ldots,x_p) \in [a_1,b_1]\times\ldots\times[a_p,b_p] \mid \frac{1}{2p\alpha}f(x_1,\ldots,x_p) < 0\}.$$

Define $\mathcal{R} = \{R, R_1, \ldots, R_p\}$ where

$$\begin{aligned} \mu_R^{(i)} &= \tfrac{1}{2} + c && (i = 1,\ldots,p) \\ \mu_{R_i}^{(i)} &= c_i x_i + \tfrac{1}{2p} && \\ \mu_{R_i}^{(j)} &= 1 && \text{for } j \neq i \\ C_R &= N && \\ C_{R_i} &= P && (i = 1,\ldots,p). \end{aligned}$$

It is easy to verify that these rules solve the classification problem. □

Theorem 5. *Let* $f : [a_1,b_1]\times\ldots\times[a_p,b_p] \longrightarrow \mathbb{R}$, $(x_1,\ldots,x_p) \mapsto c + \sum_{i=1}^{p} c_i x_i$ $(a_i < b_i)$. *Then there is a finite set* $\mathcal{R}$ *of classification rules to classes* P *and* N *such that*

$$\mathcal{R}^{-1}(P) = \{(x_1,\ldots,x_p) \in [a_1,b_1]\times\ldots\times[a_p,b_p] \mid f(x_1,\ldots,x_p) > 0\},$$

$$\mathcal{R}^{-1}(N) = \{(x_1,\ldots,x_p) \in [a_1,b_1]\times\ldots\times[a_p,b_p] \mid f(x_1,\ldots,x_p) < 0\}$$

when the minimum in (1) *is replaced by the Łukasiewicz t–norm, i.e.*

$$\mu_R(x_1,\ldots,x_p) \;=\; \max\left\{1 - p + \sum_{i=1}^{p} \mu_R^{(i)}(x_i), 0\right\} \;=\; \bigotimes_{i=1}^{p} \mu_R^{(i)}(x_i) \qquad (7)$$

and the maximum in (2) *is replaced by an arbitrary t–conorm.*

Proof. With the same argument as in the proof of Theorem 4 we may assume without loss of generality

$$\max\left\{|c|, \max_{i\in\{1,\ldots,p\}}\left\{\sup_{x\in[a_i,b_i]}\{|c_i x|\}\right\}\right\} \;<\; \varepsilon = \frac{1}{8p}.$$

Define

$$\delta \;=\; \frac{p - \frac{1}{4}}{p}.$$

Define $\mathcal{R} = \{R_P, R_N\}$ where

$$\mu_{R_P}^{(i)}(x) = \delta + c_i(x - a_i)$$

$$\mu_{R_N}^{(1)}(x) = -c + \frac{3}{4} - \sum_{i=1}^{p} c_i a_i$$

$$\mu_{R_N}^{(i)}(x) = 1 \qquad \text{for } (i \neq 1)$$

$$C_{R_P} = P$$

$$C_{R_N} = N$$

It is easy to verify that these rules solve the classification problem. □

5 Conclusions

Our analysis of fuzzy if–then classification rules shows that their principal capabilities are superior to simple crisp classification rules. Thus it is worthwhile to design efficient learning algorithms for such systems. However, one has to take into account the limitations of such classifiers, as they are described in Theorem 3. One possibility is to use other operations than simply max and min. An alternative is the design of hierarchical fuzzy classifiers on the basis of max–min rules. Since we have shown that quite general classification problems can be solved with max–min rules in the two–dimensional case, a hierarchical fuzzy classifier consisting of cascaded rules where each single rule is restricted to two variables might be a promising approach.

References

1. Bauer, P., Klement, E.P., Leikermoser, A., Moser, B.: Interpolation and approximation of real input–output functions using fuzzy rule bases. In: [8], 245–254
2. Castro, J.L., Trillas, E., Cubillo, S.: On consequence in approximate reasoning. Journal of Applied Non-Classical Logics **4** (1994), 91–103
3. Genther, H., Glesner, M.: Automatic generation of a fuzzy classification system using fuzzy clustering methods. Proc. ACM Symposium on Applied Computing (SAC'94), Phoenix (1994), 180–183
4. Ishibuchi, H.: A fuzzy classifier system that generates linguistic rules for pattern classification problems. In: Fuzzy Logic, Neural Networks, and Evolutionary Computation, Springer, Berlin (1996), 35–54
5. Klawonn, F., Kruse, R.: Derivation of fuzzy classification rules from multidimensional data. In: Lasker, G.E., Liu, X. (eds.): Advances in Intelligent Data Analysis. The International Institute for Advanced Studies in Systems Research and Cybernetics, Windsor, Ontario (1995), 90–94

6. Kosko, B.: Fuzzy systems as universal approximators. Proc. IEEE International Conference on Fuzzy Systems 1992, San Diego (1992), 1153–1162
7. Kruse, R., Gebhardt, J., Klawonn, F.: Foundations of fuzzy systems. Wiley, Chichester (1994)
8. Kruse, R., Gebhardt, J., Palm, R. (eds.): Fuzzy systems in computer science. Vieweg, Braunschweig (1994)
9. Lee, J., Chae, S.: Analysis on function duplicating capabilities of fuzzy controllers. Fuzzy Sets and Systems **56** (1993), 127–143
10. Meyer Gramann, K.D.: Fuzzy classification: An overview. In: [8], 277–294
11. Nauck D., Kruse, R.: NEFCLASS – A neuro–fuzzy approach for the classification of data. In: George, K.M, Carrol, J.H., Deaton, E., Oppenheim, D., Hightower, J. (eds.): Applied Computing 1995: Proc. of the 1995 ACM Symposium on Applied Computing. ACM Press, New York (1995), 461–465
12. Wang, L.X.: Fuzzy systems are universal approximators. Proc. IEEE International Conference on Fuzzy Systems 1992, San Diego (1992), 1163–1169
13. Weber, R.: Fuzzy–ID3: A class of methods for automatic knowledge acquisition. In: Proc. 2nd International Conference on Fuzzy Logic and Neural Networks, Iizuka (1992), 265–268