1. INTRODUCTION

The classification problem lies in the assignment an object to one of preassigned classes. It is implemented by analyzing the attributes of the classified object. Various engineering, management, economic, political, medical, sport, and other problems are reduced to classification.

Fuzzy classifiers, namely, those using fuzzy sets during functioning or learning [1], have recently become more and more popular. The application of fuzzy sets for classification problems is presented in [2] for the first time. At present, classifiers based on logic inference in terms of production rules, the antecedents of which contain the fuzzy terms “low,” “average,” “high,” and so on, are most popular. Each rule describes an area of factor space, wherein the objects belong to one class. Since the borders of these areas are fuzzy, one object can belong to several classes but with different degrees.

The main advantages of fuzzy classifiers are caused by the following factors:

• The logic inference over the fuzzy rule base is transparent. It is clear for the customers, among which are doctors, economists, politicians, and other specialists with low cybernetic engineering background.

• The classification models are compact. Only a few linguistic rules are required to describe the complicated dividing surfaces.

• Generating a base of linguistic rules is commonly simple for an expert.

• The logic inference can be implemented not only for numerical but for categorical and fuzzy values of input features as well. In this case, only the fuzzification procedure is modified in the logic inference algorithm [3], while the classification model remains constant.

The aforementioned advantages allow fuzzy decision-making models to be successful rivals for classifiers based on Bayesian rules, the nearest neighbor method, support vector machines, neural networks, and other data induction processing methods.

To increase the correctness, the fuzzy classifier is learned by experimental data. There are two approaches to the learning of the fuzzy classifier. The first one is based on the structural identification of the “inputs—output” relationship with fuzzy rules. It consists in the generation of a base of rules from the candidate-list [4], the selection of linguistic hedges, including “very” and “more or less” for the terms of rule antecedents [5], etc. Here, the learning is reduced to solving the discrete optimization problem. The second approach is based on the parametrical identification of the “inputs—output” relationship with the fuzzy rules. During learning, the rule semantics remains constant, and the membership functions of fuzzy terms and weight factors of rules are modified [1, 6, 7]. The learning is reduced to solving the optimization problem with continuous controllable variables.
The authors of the present paper consider parametrical identification, during which the classifier’s parameters are iteratively changed to provide the minimum distance between the experimental data and the fuzzy inference results. There are several methods to define such distance, which is called a learning criterion. The purpose of the article is to reveal the criterion for which learning provides the best correctness of the fuzzy classifier. Cases with equal and different costs of various errors are studied. The last one assumes that the cost matrix is known.

2. FUZZY CLASSIFIER

Let us denote by \( \mathbf{X} = (x_1, x_2, \ldots, x_n) \) the vector of informative features (attributes) of the classification object and by \( l_1, l_2, \ldots, l_m \) the decision classes. Then, the representation \( \mathbf{X} = (x_1, x_2, \ldots, x_n) \rightarrow y \in \{l_1, l_2, \ldots, l_m\} \), implemented by fuzzy rules, is said to be the fuzzy classifier. Based on [1, 4, 6, 7], the fuzzy rule base of this representation can be written as follows:

\[
\text{If } (x_1 = \tilde{a}_{ij} \text{ and } x_2 = \tilde{a}_{2j} \text{ and } \ldots \text{ and } x_n = \tilde{a}_{nj} \text{ with the weight } w_j), \text{ then } y = d_j, \quad j = \overline{1, k},
\]

where \( k \) is the number of rules;

\( d_j \in \{l_1, l_2, \ldots, l_m\} \) is the value of consequent of the \( j \)-th rule;

\( w_j \in [0, 1] \) is the weight factor specified the reliability of the \( j \)-th rule, \( j = \overline{1, k} \);

\( \tilde{a}_{ij} \) is a fuzzy term that is the evaluating attribute \( x_i \) in the \( j \)-th rule, \( i = \overline{1, n}, j = \overline{1, k} \).

The classification of the current object given by the attribute vector \( \mathbf{X}^* = (x_1^*, x_2^*, \ldots, x_n^*) \) is implemented as follows. At first, the degree of fulfillment of the \( j \)-th rule from base (1) is calculated:

\[
\mu_j(\mathbf{X}^*) = w_j \cdot \left( \mu_j(x_1^*) \land \mu_j(x_2^*) \land \ldots \land \mu_j(x_n^*) \right), \quad j = \overline{1, k},
\]

where \( \mu_j(x_i^*) \) is the membership degree of \( x_i^* \) to the fuzzy term \( \tilde{a}_{ij} \); \( \land \) is the \( t \)-norm, which is generally realized by minimum operation or product.

The membership degree of the input vector \( \mathbf{X}^* \) to classes \( l_1, l_2, \ldots, l_m \) is estimated as follows:

\[
\mu_i(y^*) = \text{agg} \left( \mu_j(\mathbf{X}^*) \right), \quad s = \overline{1, m},
\]

where \( \text{agg} \) is the aggregation of the results of fuzzy inference by the rules with the same consequents. The aggregation is realized by the maximum operation over the membership degrees that corresponds to the logic inference scheme with a single winner rule [8].

The fuzzy solution of the classification problem is the fuzzy set

\[
\tilde{y}^* = \left( \frac{\mu_1(y^*)}{l_1}, \frac{\mu_2(y^*)}{l_2}, \ldots, \frac{\mu_m(y^*)}{l_m} \right).
\]

The result of fuzzy inference is selected to be the core of the fuzzy set (4), namely, the class with the maximum membership degree:

\[
y^* = \arg \max_{s=\overline{1, m}} (\mu_i(y^*)).
\]

The core of the fuzzy set (4) can include several elements. The object then concurrently belongs to several classes with equal degrees, the value of which is \( \max_{s=\overline{1, m}} (\mu_i(y^*)) \). Let us apply the voting-based scheme [8]. According to this scheme, the sum of degrees (2) of fulfillment of the corresponding rules is calculated for each class. The class with the maximum sum is selected as the decision.

3. LEARNING CRITERIA FOR THE FUZZY CLASSIFIER WITHOUT REGARD TO THE COST MATRIX

It is assumed that the learning set from \( M \) pairs “inputs—output” is known:

\[
(\mathbf{X}_r, y_r), \quad r = \overline{1, M},
\]

where \( \mathbf{X}_r = (x_{r1}, x_{r2}, \ldots, x_{rn}) \) and \( y_r \in \{l_1, l_2, \ldots, l_m\} \) are the \( r \)-th element of the learning set and the output variable of this element, respectively.