# Bayesian Classification of Interval-Type Information

**Piotr Kulczycki and Piotr A. Kowalski**

**Abstract** The subject of Bayes classification of imprecise multidimensional information of interval type by means of patterns defined through precise data (i.e. deterministic or sharp) is investigated here. To this aim the statistical kernel estimators methodology was applied, which avoids the pattern shape for the resulting algorithm. In addition, elements of pattern sets which have insignificant or negative influence on correctness of classification are eliminated. The concept for realizing the procedure is based on the sensitivity method, used in the domain of artificial neural networks. As a result of this procedure the number of correct classifications and—above all—calculation speed increased significantly. A further growth in quality of classification was achieved with an algorithm for the correction of classifier parameter values.

**Keywords** Data analysis · Classification · Interval information · Nonparametric methods · Kernel estimators · Reduction in pattern size · Classifier parameter correction · Neural networks

P. Kulczycki (✉) · P. A. Kowalski
Centre of Information Technology for Data Analysis Methods,
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland
e-mail: Piotr.Kulczycki@ibspan.waw.pl

P. A. Kowalski
e-mail: Piotr.A.Kowalski@ibspan.waw.pl

P. Kulczycki · P. A. Kowalski
Department of Automatic Control and Information Technology,
Cracow University of Technology, Krakow, Poland

L. T. Kóczy et al. (eds.), *Issues and Challenges of Intelligent Systems and Computational Intelligence*, Studies in Computational Intelligence 530,
DOI: 10.1007/978-3-319-03206-1_18, 

## 1 Introduction

The current dynamic development in computer technology offers a continuous increase in both capability and speed of contemporary calculational systems, thus allowing ever more frequent use of methods which up to now had only been applied to a relatively limited extent. One of these methods is the analysis of information which is imprecise in various—depending on a problem's conditioning—forms, for example uncertain (statistical methods [8]) or fuzzy (fuzzy logic [10]).

Lately many applications have noted an increase in the use of interval analysis. The basis for this concept is the assumption that the only available information on an investigated quantity is the fact that it fulfils the dependence $\underline{x} \leq x \leq \bar{x}$, and in consequence this quantity can be associated with the interval

$$[\underline{x}, \overline{x}]. \tag{1}$$

Interval analysis is a separate mathematical domain, with its own formal apparatus based on an axiom of the sets theory [20].

A fundamental application of interval analysis was to ensure the required precision of numerical calculations, through monitoring errors arising from rounding numbers [1], however as a result of its continuous development, this field is finding ever wider uses in engineering, econometrics and other related areas [9]. Its main advantage is the fact that by definition it models imprecision of a researched quantity, using the simplest possible formula. In many applications interval analysis shows to be absolutely sufficient, yet does not require many calculations (thus enabling its application in highly complex tasks) and is easy to identify and interpret, while also maintaining a formalism stemming from a convenient mathematical tool. Moreover, it can be noted that its concept is related to statistical interval estimation, and analysis of fuzzy numbers with rectangular membership functions.

Dynamic development is also currently taking place in information technologies in the area of data analysis and exploration [16]. This is due not only to an increase in the possibilities of the methodology used here, but above all to an increase in accessibility of its algorithms, up to now a domain only available to a relatively small group of specialists. Among the fundamental tasks of data analysis and exploration lies that of classification [7]. It consists of assigning a tested element to one of several previously selected groups. They are most often given by patterns, which are sets of elements representative for particular classes. This means that in many problems—including those where data containing imprecision is investigated—elements defining patterns are defined precisely (e.g. deterministic in probability approach, sharp for the case of fuzzy logic, or in relation to notation (1) fulfilling equality $\underline{x} = \overline{x}$).

This chapter offers a complete procedure for classification of imprecise information, defined as the interval vector

$$\begin{bmatrix} [\underline{x_1}, \overline{x_1}] \\ [\underline{x_2}, \overline{x_2}] \\ \vdots \\ [\underline{x_n}, \overline{x_n}] \end{bmatrix}, \tag{2}$$

where $\underline{x_k} \leq \overline{x_k}$ for $k = 1, 2, \ldots, n$, when the patterns of particular classes are given as sets of precise data (i.e. deterministic or sharp) elements, i.e. with $\underline{x_k} = \overline{x_k}$ $(k = 1, 2, \ldots, n)$. The classification concept is based on the Bayes approach, ensuring a minimum of potential losses occurring through classification errors. For a such formulated task the statistical kernel estimators methodology was employed, thereby freeing the above procedure from arbitrary assumptions regarding pattern forms—their identification becomes an integral part of the presented algorithm. A procedure was also developed for reducing the size of pattern sets by elements having negligible or negative influence on correctness of classification. Its concept is founded on the sensitivity method, used in the domain of artificial neural networks, although the intention is to increase the number of accurate classifications and—above all—calculation speed. Furthermore a method was designed to ensure additional improvements in classification results, obtained by correcting the values of classifier parameters. Basic investigations are presented in the chapter [17]. First research in this subject was described in the work [13].

## 2 Preliminaries

### *2.1 Statistical Kernel Estimators*

Kernel estimators belong to the group of nonparametric statistical methods. They allow calculation and clear illustration of characteristics of a random variable distribution, without knowledge of its membership of a given class.

Let $(\Omega, \Sigma, P)$ denote a probability space. Let also an $n$-dimensional random variable $X : \Omega \to \mathbb{R}^n$, with distribution density $f$, be given. Its kernel estimator $\hat{f} : \mathbb{R}^n \to [0, \infty)$ is calculated on the basis of an $m$-elements random sample $\{x_i\}_{i=1,2,\ldots,m}$ and defined—in the basic form—by the formula

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right), \tag{3}$$

where the positive coefficient $h$ is known as a smoothing parameter, while the measurable function $K : \mathbb{R}^n \to [0, \infty)$ symmetrical with respect to zero, having at this point weak global maximum and fulfilling the condition $\int_{\mathbb{R}^n} K(x)\mathrm{d}x = 1$ is termed a kernel. The choices of form for the kernel $K$ and value for the smoothing parameter

$h$ are most often made based on the criterion of minimization of integrated square error [15, 21, 23].

Thus, the form of the kernel $K$ has practically no influence on the statistical quality of estimation. This chapter applies the generalized (one-dimensional) Cauchy kernel

$$K(x) = \frac{2}{\pi(x^2+1)^2}, \tag{4}$$

in the multidimensional case defined using the product kernel concept

$$K(x) = K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \mathscr{K}(x_1) \cdot \mathscr{K}(x_2) \cdot \ \ldots \ \cdot \mathscr{K}(x_n)\ , \tag{5}$$

where $\mathscr{K}$ denotes here the one-dimensional kernel given by formula (4).

The value of the smoothing parameter $h$ can be calculated in practice with confirmed algorithms available in literature. The effective and convenient plug-in method [15, Sect. 3.1.5]; [23, Sect. 3.6.1] is recommended here. In the multidimensional case, regarding application of the product kernel in this chapter, the smoothing parameter will be naturally denoted as $h_1, h_2, \ldots, h_n$ respectively for subsequent coordinates, and can be obtained separately for each of them by the above suggested method.

In practice one employs additional procedures to generally increase the quality of the kernel estimator and fit its features to those of the considered reality. In this chapter the modification of the smoothing parameter [15, Sect. 3.1.6]; [21, Sect. 5.3.1] will be applied, thereby significantly improving the properties of the kernel estimator, particularly in areas where it assumes small values. In classification tasks this takes place especially near boundaries of specific classes, which makes this procedure particularly useful here.

Consider therefore nonnegative modifying coefficients

$$s_i = \left(\frac{\hat{f}_*(x_i)}{\bar{s}}\right)^{-c} \text{ for } i = 1, 2, \ldots, m, \tag{6}$$

where the constant $c \geq 0$ is called a modification intensity, $\hat{f}_*$ denotes the kernel estimator in its basic form (3), and $\bar{s}$—the geometrical mean of the quantities $\hat{f}_*(x_i)$ with $i = 1, 2, \ldots, m$. The final definition of estimator (3) with product kernel (5) then takes the form:

$$\hat{f}(x) = \frac{1}{m h_1 h_2 \ldots h_n} \sum_{i=1}^{m} \frac{1}{s_i^n} \mathscr{K}\left(\frac{x_1 - x_{i,1}}{h_1 s_i}\right) \mathscr{K}\left(\frac{x_2 - x_{i,2}}{h_2 s_i}\right) \ \ldots \ \mathscr{K}\left(\frac{x_n - x_{i,n}}{h_n s_i}\right), \tag{7}$$

where the natural notations

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,n} \end{bmatrix} \text{ for } i = 1, 2, \ldots, m \tag{8}$$

are used, and together with formula (4) will be employed later in this chapter. The case $c = 0$ determines the lack of smoothing parameter modification, while together with an increase in the value $c$ its intensity grows. Corollaries resulting from the mean-square criterion primarily point to the value

$$c = 0.5 \, . \tag{9}$$

Statistical kernel estimators are dealt with in the monographs [15, 21, 23]. Information on the subject of their applications in standard classification tasks can be found in the books [2, 3, 18, 19].

## 2.2 Sensitivity Analysis of Neural Networks

When modeling multidimensional problems using artificial neural networks [22], particular components of an input vector most often are characterized by diverse significance of information, and in consequence influence variously the result of the data processing. In order to eliminate redundant—from the point of view of the investigated task—input vector components, a sensitivity analysis of the network with respect to particular learning data is often used. A basic factor for network reduction is sensitivity of the output function with regards to particular input data.

The essence of the sensitivity method [24] consists in defining—after the network learning phase—the influence of the particular inputs $u_i$ for $i = 1, 2, \ldots, m$ on the output value $y$, which is characterized by the real coefficients

$$S_i = \frac{\partial y(u_1, u_2, \ldots, u_m)}{\partial u_i} \quad \text{for} \quad i = 1, 2, \ldots, m. \tag{10}$$

Next, one aggregates the particular coefficients $S_i^{(p)}$ originating from successive iterations of the previous phase and corresponding to the sensitivity of subsequent learning data, with $p = 1, 2, \ldots, P$. The result is the final coefficient $\bar{S}_i$ given by the formula

$$\bar{S}_i = \sqrt{\frac{\sum_{p=1}^{P} (S_i^{(p)})^2}{P}} \quad \text{for} \quad i = 1, 2, \ldots, m. \tag{11}$$

After the sorting operation for the vector $\bar{S}_i$ according to decreasing values, an analysis of the relevance of particular components to the result of network operation is performed, and then the least important inputs are eliminated.

In the general case the above algorithm can be used repeatedly to achieve further reduction. However, during empirical testing of the classification method developed here, such action did not bring positive results and so was forsaken.

The application of the above method led to an increase in speed, as well as reduction of errors of learning and generalization, while at the same time reducing the input dimension of the artificial neural network by removing information of little significance or even elimination of data (input vector components) having unfavorable influence on the obtained result's correctness. Detailed considerations concerning the above procedure are found in the publications [4, 24].

## 3 An Algorithm for Interval Classification

### *3.1 One-Dimensional Case*

This section considers the one-dimensional case, i.e. when $n = 1$. Let therefore be given the quantity having undergone the classification procedure, for the case currently being considered, represented by the (one-dimensional) interval

$$[\underline{x}, \overline{x}] \,, \tag{12}$$

while $\underline{x} \leq \bar{x}$; if $\underline{x} = \bar{x}$ then the classic case is obtained where the quantity is precise (e.g. deterministic or sharp). Assume also that the real number sets (patterns):

$$x_1^1,\ x_2^1, \ldots, x_{m_1}^1 \tag{13}$$

$$x_1^2,\ x_2^2, \ldots, x_{m_2}^2 \tag{14}$$

$$\vdots$$

$$x_1^J,\ x_2^J, \ldots, x_{m_J}^J \tag{15}$$

represent subsequent $J$ marked classes. The upper index, introduced in the above notation, characterizes membership of an element to a given class. As stated before, the task of classification consists of deciding to which of these groups tested element (12) should be assigned.

Let now $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_J$ denote kernel estimators of probability distribution density, calculated successively based on sets (13)–(15) treated as random samples—a description of the methodology used for their construction is contained in Sect. 2.1.

In accordance with the classic Bayes approach [3], the classified element $\tilde{x} \in \mathbb{R}$ should then be given to the class for which the value

$$m_1 \hat{f}_1(\tilde{x}), m_2 \hat{f}_2(\tilde{x}), \ldots, m_J \hat{f}_J(\tilde{x}) \tag{16}$$

is the biggest. In the case of information of interval type, represented by element (12), one can infer that this element belongs to the class for which the expression

$$\frac{m_1}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_1(x)\mathrm{d}x, \frac{m_2}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_2(x)\mathrm{d}x, \ldots, \frac{m_J}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_J(x)\mathrm{d}x \tag{17}$$

is the greatest.

Considering the limit transitions $\overline{x} \to \tilde{x}$ and $\underline{x} \to \tilde{x}$ for the fixed $\tilde{x} \in \mathbb{R}$, then due to the continuity of the function $K$ used here, given by formula (4), consequently implying the continuity of the kernel estimator $\hat{f}_j$, one obtains

$$\lim_{\substack{\underline{x} \to \tilde{x} \\ \overline{x} \to \tilde{x}}} \frac{1}{\overline{x} - \underline{x}} \int_{\underline{x}}^{\overline{x}} \hat{f}_j(x)\mathrm{d}x = \hat{f}_j(\tilde{x}) \quad \text{for } j = 1, 2, \ldots, J. \tag{18}$$

The expressions specified in formula (17) reduce therefore to the classic type (16).

In formula (17), the positive expression $1/(\overline{x} - \underline{x})$ can be omitted as irrelevant in an optimization problem, and so it is equivalent to

$$m_1 \int_{\underline{x}}^{\overline{x}} \hat{f}_1(x)\mathrm{d}x, m_2 \int_{\underline{x}}^{\overline{x}} \hat{f}_2(x)\mathrm{d}x, \ldots, m_J \int_{\underline{x}}^{\overline{x}} \hat{f}_J(x)\mathrm{d}x \ . \tag{19}$$

What is more, for any $j = 1, 2, \ldots, J$ one can note

$$\int_{\underline{x}}^{\overline{x}} \hat{f}(x)\mathrm{d}x = \hat{F}(\overline{x}) - \hat{F}(\underline{x}) \ , \tag{20}$$

where

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(y)\mathrm{d}y \ . \tag{21}$$

Taking into consideration dependence (21) substituting equalities defining kernel estimator (7) (for $n = 1$) and kernel (4) used, one can analytically calculate that

$$\hat{F}(x) = \sum_{i=1}^{m} \left[ \frac{(x^2 - 2xx_i + x_i^2 + h^2 s_i^2) \operatorname{arctg}\left(\frac{x-x_i}{s_i h}\right) + h s_i (x - x_i)}{x^2 - 2xx_i + x_i^2 + h^2 s_i^2} + \frac{\pi}{2} \right], \tag{22}$$

where again the positive constant $1/m\pi$ has been omitted. Finally it should be acknowledged that the considered element belongs to the class for which the corresponding expression in formula (19) is the greatest, whereby the integral appearing there for any $j = 1, 2, \ldots, J$ can be effectively calculated using dependences (20) and (22). The above completes the classification algorithm for the one-dimensional case.

## 3.2 The Multidimensional Case

The concept presented in the previous subsection can be naturally generalized for the multidimensional case, i.e. when $n > 1$. Thus, if information of interval type is represented by the interval vector

$$\begin{bmatrix} [\underline{x_1}, \overline{x_1}] \\ [\underline{x_2}, \overline{x_2}] \\ \vdots \\ [\underline{x_n}, \overline{x_n}] \end{bmatrix}, \tag{23}$$

and sets (13)–(15) contain the elements of the space $\mathbb{R}^n$, then one can infer that the considered element is assigned to the class with the greatest value for the expression

$$m_1 \int_E \hat{f}_1(x)\mathrm{d}x, m_2 \int_E \hat{f}_2(x)\mathrm{d}x, \ldots, m_J \int_E \hat{f}_J(x)\mathrm{d}x \,, \tag{24}$$

where $E = [\underline{x_1}, \overline{x_1}] \times [\underline{x_2}, \overline{x_2}] \times \cdots \times [\underline{x_n}, \overline{x_n}]$. It is slightly different, though, for the algorithm for calculating the integrals appearing above. However, thanks to the properties of the product kernel used here, for any fixed $j = 1, 2, \ldots, J$ and the kernel $K$, the following dependence is true:

$$\int_E K(x)\mathrm{d}x = [\mathscr{F}(\overline{x_1}) - \mathscr{F}(\underline{x_1})][\mathscr{F}(\overline{x_2}) - \mathscr{F}(\underline{x_2})] \ldots [\mathscr{F}(\overline{x_n}) - \mathscr{F}(\underline{x_n})] \,, \tag{25}$$

where $\mathscr{F}$ denotes the primitive of the function $\mathscr{K}$ introduced by definition (5). Taking into account the definition of the kernel estimator with product kernel (7) as well as the analytical form of the primitive function contained in formula (22), the above completes the procedure for classification of interval type information, for the multidimensional case also.

### 3.3 Calculational Complexity of the Algorithm

From the point of view of calculational complexity, it is worth underlining the two-phased nature of the method presented in this chapter. The first stage contains the complex procedures for constructing the classifier, which are executed once at the beginning. The most time-consuming is the algorithm for calculating the smoothing parameter using the plug-in method of complexity $O(nm^2)$. This same complexity characterizes the calculations for the smoothing parameter modification procedure.

On the contrary, the procedure for calculating the values of the kernel estimator has the complexity $O(nm)$. Taking into account that the number of the operations above is equal to the number of assumed classes $J$, then the calculational complexity of the second phase is linear with respect to all three parameters: $n$, $m$ and $J$, where $m$ characterizes here the size of particular patterns. It implies a relatively short calculation time, which after earlier execution of the first phase, in most practical problems allows for the application of the investigated algorithm in real time, in an on-line regime.

## 4 Procedures for Increasing Classification Quality

### 4.1 Reducing Pattern Size

In practice, some elements of sets (13)–(15), constituting patterns of particular classes, may have insignificant or even negative—in the sense of classification correctness—influence on quality of obtained results. Their elimination should therefore imply a reduction in the number of erroneous assignations, as well as decreasing calculation time. To this aim the sensitivity method for learning data, used in artificial neural networks, described in Sect. 2.2, will be applied.

To meet requirements of this procedure, the definition of the kernel estimator will be generalized below with the introduction of nonnegative coefficients $w_1, w_2, \ldots, w_m$, adjusted by the condition

$$\sum_{i=1}^{m} w_i = m \ , \tag{26}$$

and mapped to particular elements of the random sample. The basic form of kernel estimator (3) then takes the form

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} w_i K\left(\frac{x - x_i}{h}\right) . \tag{27}$$

Formula (7) undergoes analogous generalization. The coefficient $w_i$ value may be interpreted as indicating the significance of the $i$-th element of the pattern to classification correctness. Note that if $w_i \equiv 1$, then dependence (27) is regressed to initial form (3).

In the method designed here, for the purpose of reduction of sets (13)–(15), separate neural networks are built for each investigated class. In order to ensure coherence of the notation below, let now the index $j = 1, 2, \ldots, J$ characterizing particular classes, be arbitrarily fixed.

The constructed network has three layers and is unidirectional, with $m$ inputs (corresponding to particular elements of a pattern), a hidden layer whose size is equal to the integral part of the number $\sqrt{m}$, and also one output neuron. This network is submitted to a learning process using a data set comprising of the values of particular kernels for subsequent pattern elements, while the given output constitutes the value of the kernel estimator calculated for the pattern element under consideration. Apart from the above topology, as a result of empirical research, the maximum number of epochs was assumed as 100, the maximum learning error 0.01, the learning speed 0.3, and the momentum coefficient as 0.1. On finishing the learning process, the thus obtained network undergoes sensitivity analysis on learning data, in accordance with the method presented in Sect. 2.2. The resulting coefficients $\bar{S}_i$ describing sensitivity, obtained on the basis of formula (11), constitute the fundament for calculating preliminary values

$$\tilde{w}_i = \left( 1 - \frac{\bar{S}_i}{\sum_{j=1}^{m} \bar{S}_j} \right), \tag{28}$$

after which they are adjusted to the form

$$w_i = m \frac{\tilde{w}_i}{\sum_{i=1}^{m} \tilde{w}_i} \tag{29}$$

to guarantee condition (26). It is worth noting that the form of formulas (10)–(11) accounting in practice for all coefficients $\bar{S}_i$ can not be equal to zero, which guarantees feasibility of the above operation. The formula for dependence (28) results from the fact that the network created here is the most sensitive to atypical and redundant elements, which—taking into account the form of kernel estimator (27)—implies a necessity to map the appropriately smaller values $\tilde{w}_i$, and in consequence $w_i$, to them. The coefficients (29) characterize—according to the idea presented during formulation of generalized form (27)—the significance of particular elements of the pattern, for classification procedure correctness.

Empirical research confirmed the natural assumption that the pattern set should be relieved of those elements for which $w_i < 1$. (Note that, thanks to adjustments made by formula (29), the mean value of coefficients $w_i$ equals 1.) Decreasing of a such assumed threshold value resulted in a significant drop in the degree of a pattern size reduction, while in vicinity of the value 1 the influence on classification quality was practically unnoticeable, however considerable diminishing implied a sizable

rise in number of errors. On the other hand, an increase in this value caused a sharp fall in classification quality, due to a loss of valuable and non-redundant information included in the pattern.

### *4.2 Correcting the Smoothing Parameter and Modification Intensity Values*

Subject literature often presents the opinion that the classic universal methods of calculating the smoothing parameter value— most often based on a quadratic criterion— are not satisfying for the classification task. For example, in the article [6] experimental research conducted on two classes was presented showing the significant difference between the value of this parameter when calculated by minimizing integrated square error, and when obtained by minimizing the number of misclassifications. However the latter method is difficult in practical use for the multidimensional case, due to an extraordinarily long calculation time—a problem which becomes more important the greater the number of classes. Available literature does not propose a definitive solution for such a task.

This chapter suggest introducing $n + 1$ multiplicative correcting coefficients for the values of the parameter defining the intensity of modification procedure $c$ and smoothing parameters for particular coordinates $h_1, h_2, \ldots, h_n$, with respect to optimal ones calculated using the integrated square error criterion. Denote them as $b_0 \geq 0, b_1, b_2, \ldots, b_n > 0$, respectively. It is worth noticing that $b_0 = b_1 = \cdots = b_n = 1$ means in practice no correction. Next through a comprehensive search using a grid with a relatively large discretization value, one finds the most advantageous points regarding minimal incorrect classification sense. The final phase is a static optimization procedure in the $(n+1)$-dimensional space, where the initial conditions are the points chosen above, while the performance index is given as

$$J(b_0, b_1, \ldots, b_n) = \#\{incorrect \quad classifications\}\,, \tag{30}$$

when # denotes the power (size) of a set. The value of the above functional for a fixed argument is calculated with the help of the classic leave-one-out method. This is an integer—to find the minimum a modified Hook-Jeeves algorithm [11] was applied.

Following experimental research it was assumed that the grid used for primary searches has intersections at the points $0.25, 0.5, \ldots, 1.75$ for every coordinate. For such intersections the value of functional (30) is calculated, after which the obtained results are sorted, and the 5 best become subsequent initial conditions for the Hook-Jeeves method, where the value of the initial step is taken as 0.2. After finishing every one of the above 5 "runs" of this method, the functional (30) value for the end point is calculated, and finally among them is shown the one with the smallest value.

## 5 Final Remarks and Conclusions

This chapter presents the complete Bayes algorithm—thereby ensuring minimum potential losses—for the classification of multidimensional imprecise information of interval type, where patterns of particular classes are given on the basis of sets of precisely defined elements, with no limits to the number of classes. In addition two optional procedures are provided, which improve and enhance the quality of classification: a reduction in pattern size and a correction of the classifier parameter values.

Numerical testing wholly confirmed the positive features of the method worked out. It was carried out with the use of pseudorandom and benchmark data. In particular, the results show that the classifying algorithm can be used successfully for inseparable classes of complex multimodal patterns as well as for those consisting of incoherent subsets at alternate locations. This is thanks to the application of the statistical kernel estimators methodology, which makes the above procedure independent of the shapes of patterns—their identification is an integral part of the presented algorithm. As shown by numerical verification, the algorithm has beneficial features in the multidimensional case too. The results also compared positively to those obtained by applying support vectors machines as well as by the two natural methods.

In particular, during numerical testing it was established that, after applying the procedure for reducing pattern sets, presented in Sect. 4.1, the number of wrong classifications was lowered by approximately 15 %, while the size of patterns was reduced by approximately 40 %. The conjunction of these results is particularly worthy of attention: while appropriately reducing pattern sizes, which does imply a significant increase in calculation speed, the classification quality is also importantly improved. In addition, the procedure of correcting the smoothing parameter and intensity of its modification, presented in Sect. 4.2, conducted after the reduction of patterns caused a further decrease in the number of classification errors to approximately 14 %.

The task of classifying interval information based on precise data can be interpreted illustratively with the example where the patterns present actual, precisely measured quantities, while intervals being classified represent uncertainties and imprecision in plans, estimations or difficult measurements to make. In particular, pattern sets may consists of very accurate measurements, in which errors are practically ignored, while the classified interval constitutes a measurement taken from another, much less accurate apparatus or carried out in much worse conditions. Another example of the application of this kind of classification is the possibility of treating precise data as actual information from the past, e.g. temperature or currency exchange rates, while the classified element represents a prognosis which by nature is limited in precision.

In particular, the method investigated here can be applied for purposes of the diagnosis process [5, 12, 14]. Namely, let interval (12) or interval vector (23) represent a quantity or $n$ quantities, respectively, whose values attest to the current or—in the case of fault prognosis—predicted technical state of a supervised device. Because of measurement errors and natural fluctuations, the interval form can be

justified in many practical tasks. Let also sets (13)–(15) constitute patterns of particular types of possible faults. The classification procedure presented in this chapter allows for precise diagnostic readings to be obtained, with regard to interval character of investigated quantities.

## References


1. Alefeld, G., Hercberger, J.: Introduction to Interval Computations. Academic Press, New York (1986)
2. Devroye, L., Gyorfi, L., Lugosi, G.: A Probabilistic Theory of Pattern Recognition. Springer, New York (1996)
3. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (2001)
4. Engelbrecht, A.P., Cloete, I., Zurada, J.: Determining the significance of input parameters using sensitivity analysis. In: International Workshop on Artificial Neural Networks, Torremolinos (Spain), 7–9 June 1995. LNCS 930, pp. 382–388
5. Gertler, J.: Fault Detection and Diagnosis in Engineering Systems. Dekker, New York (1998)
6. Ghosh, A.K., Chaudhuri, P., Sengupta, D.: Classification Using Kernel Density Estimation: Multiscale Analysis and Visualization. Technometrics, **48**, 120–132 (2006)
7. Hand, D.J.: Construction and Assessment of Classification Rules. Wiley, Chichester (1997)
8. Hryniewicz, O.: Reliability Sampling; in Encyclopedia of Statistics in Quality and Reliability. Wiley, Chichester (2008)
9. Jaulin, L., Kieffer, M., Didrit, O., Walter, E.: Applied Interval Analysis. Springer, Berlin (2001)
10. Kacprzyk, J.: Multistage Fuzzy Control: A Model-Based Approach to Control and Decision-Making. Wiley, Chichester (1997)
11. Kelley, C.T.: Iterative Methods for Optimization. SIAM, Philadelphia (1999)
12. Korbicz, J., Koscielny, J.M., Kowalczuk, Z., Cholewa, W., (eds.): Fault Diagnosis. Models, Artificial Intelligence Applications. Springer, Berlin (2004)
13. Kowalski, P.A.: Klasyfikacja bayesowska informacji niedokladnej typu przedzialowego. Ph.D. thesis, Systems Research Institute, Polish Academy of Sciences (2009)
14. Kulczycki, P.: Wykrywanie uszkodzen w systemach zautomatyzowanych metodami statystycznymi. Alfa, Warsaw (1998)
15. Kulczycki, P.: Estymatory jadrowe w analizie systemowej. WNT, Warsaw (2005)
16. Kulczycki, P., Hryniewicz, O., Kacprzyk, J. (eds.): Techniki informacyjne w badaniach systemowych. WNT, Warsaw (2007)
17. Kulczycki, P., Kowalski, P.A.: Bayes classification of imprecise information of interval type. Control Cybern. **40**, 101–123 (2011)
18. Ledl, T.: Kernel density estimation, theory and application in discriminant analysis. Austrian J. Stat. **33**, 267–279 (2004)
19. McLachlan, G.J.: Discriminant Analysis and Statistical Pattern Recognition. Wiley, Hoboken (2004)
20. Moore, R.E.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)
21. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
22. Tadeusiewicz, R., Ogiela, M.R.: Medical Image Understanding Technology. Springer, Berlin (2004)
23. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman and Hall, London (1995)
24. Zurada, J.: Introduction to Artificial Neural Systems. West Publishing, St. Paul (1992)