

IMPROVEMENT OF ACCURACY OF PARAMETRIC CLASSIFICATION IN THE SPACE OF $N \times 2$ FACTORS- ATTRIBUTES ON THE BASIS OF PRELIMINARY OBTAINED LINEAR DISCRIMINANT FUNCTION

Mourad Aouati

Chief Commissioner of Police

Central City Police Department of Constantine

Ali Mendjeli UV 01 Ilot 03 Bt H n°123

Constantine, Algeria

Abstract

A procedure for classifying objects in the space of $N \times 2$ factors-attributes that are incorrectly classified as a result of constructing a linear discriminant function is proposed. The classification accuracy is defined as the proportion of correctly classified objects that are incorrectly classified at the first stage of constructing a linear discriminant function. It is shown that, for improperly classified objects, the transition from use as the factors-attributes of their initial values to the use of the centers of gravity (COGs) of local clusters provides the possibility of improving the classification accuracy by 14 %. The procedure for constructing local clusters and the principle of forming a classifying rule are proposed, the latter being based on converting the equation of the dividing line to the normal form and determining the sign of the deviation magnitude of the COGs of local clusters from the dividing line.

Keywords: parametric classification, pattern recognition, object classification accuracy, cluster, linear discriminant function.

DOI: 10.21303/2461-4262.2017.00362

© Mourad Aouati

1. Introduction

The purpose of applying image recognition methods is to obtain classification rules that ensure the maximum accuracy of classification. As a rule, it is estimated by the number of incorrectly classified objects in relation to the total number of classified objects. A wide range of applications of these methods shows that one can't do without them when solving problems from completely different subject areas. The methods are used for image recognition [1–5], medical diagnostics [6, 7], technical diagnostics [8–12]. Principal differences in approaches for solving specific problems are related to the choice of methods and the feasibility of the chosen alternative. Nevertheless, when choosing any of the alternatives, the problem is reduced to obtaining a decisive rule allowing to assign the object to one of the classes. The probable error in the classification can lead to negative consequences:

- an error in the identification of the criminal leads to a newly committed crime because of the untimely detention of the criminal or the punishment of the innocent;
- an error in the medical diagnosis leads to an incorrect strategy of treatment and progression of the disease and severe complications;
- an error in identifying the fault and its causes in the technical devices leads to their failure.

This list can be continued, therefore, the relevance of research topics aimed at finding ways to improve the accuracy of the classification of objects, has no doubt. In this context, the key is the problem of creating new or modifying existing methods of recognition.

It should be noted that the use of parametric methods based on Bayesian statistics is favored by researchers who solve specific problems in industrial production [10–12]. As a classifying rule, linear discriminant functions are used, which, however, have the disadvantage that not always a linear hyperplane can qualitatively separate arbitrarily located clusters of classified objects.

Unlike the approach based on the construction of an analytical description of the classifying rule, some alternative variants do not at all suggest the construction of any mathematical description. Examples can be a linear learning machine or the nearest-neighbor classification [13].

Among the most common in recent years, it is necessary to identify classification methods based on the use of neural networks [14–16]. The key to using these methods is the neural network training quality. Thus, the paper [15] describes the results of applying one of the classical algo-

gorithms for development of a decision support subsystem in a neural network pattern recognition system. As effectiveness criteria, the authors suggest to use the level of subjectivism of expert decisions and the quality of expert assessments in the construction of training samples on the existing statistical data describing the objects. In this case, the first of the effectiveness criteria should be minimized, and the second one should be maximized. As a result of applying the proposed algorithm, the authors give the following data: the relative share of correct expert assessment increases by an average of 20 %, and the relative proportion of false ones decreases by 50 %.

It should be noted that the limitations imposed on the use of the proposed method are not presented in this article. Nevertheless, it is stated that the existing basic “classical” training models, for example, based on error correction, using memory, competitive training and the Boltzmann method, have a number of shortcomings. In particular, it is noted that when creating universal image recognition systems, it is impossible to manage one of the training models. This is especially evident in specific subject areas, where the construction of a universal qualitative recognition system is confronted with an almost unsolvable task of taking into account all the specific features inherent in these subject areas.

It is possible to use the Levenberg-Marquardt method [16], which, unlike the classical learning algorithms, uses Z-training followed by averaging the network error. The efficiency criterion is time of its operation of the algorithm, which should be minimized. However, the key presentation of the results is limited to the theoretical description of the algorithm.

All the mentioned studies assume the work with random variables with a known law of data distribution. Obvious complications in their use are due to the fact that the condition of knowledge of the laws of distribution is not always satisfied. At the same time, if the components of image vectors can't be measured with sufficient accuracy, the possibility of constructing effective recognition systems becomes questionable. The ways of solving the arising problems are considered, for example, in [17–19].

It is shown in [17] that the use of the principal component method (PCA) in combination with hierarchical clustering is an effective practical tool for the automated identification of the studied objects.

The approach described in [18] is based on a point estimate for fuzzy pattern recognition conditions. In this approach, the space is considered to consist of two fuzzy sets (“Truthful” and “Deceiver”). At the initial stage of the procedure each separate classified element is modeled as a fuzzy set based on the generation method. The purpose of this procedure is to eliminate the uncertainty in the assessment of the corresponding scores. At a later stage, new fuzzy conformity assessments are integrated with the fuzzy aggregation operator, after which the final decision on recognition is taken. The authors of [17, 18] note that the proposed method has high stability, but the limitations imposed on these methods by special conditions of subject areas are not named.

The approach to data processing for classification using the methods of computational intelligence can be found in the study [19], which proposed an adaptive algorithm for fuzzy clustering using an objective function of the form

$$E(w_j(k), c_j) = \sum_{k=1}^N \sum_{j=1}^m w_j^\beta(k) d^2(x(k), c_j) \rightarrow \min \quad (1)$$

under the following restrictions:

$$\sum_{j=1}^m w_j(k) = 1, \quad k=1, \dots, N, \quad 0 < \sum_{k=1}^N w_j(k) < N, \quad j=1, \dots, m. \quad (2)$$

$w_j(k) \in [0, 1]$ – the level of belonging of the vector $x(k)$ to the j -th cluster, c_j – centroid of the j -th cluster, $d^2(x(k), c_j)$ – distance between $x(k)$ and c_j in the accepted metric, β – a nonnegative parameter called the “fuzzifier” (if used $d^2(x(k), c_j)$ as a Euclidean Distance, is assumed to be 2).

The result of this algorithm is the formation of a fuzzy partition matrix, in which objects are divided into clusters (diagnoses), while the shape of clusters can vary from a hypersphere to a hyperellipsoid depending on the form of the original data. The deciding factor for deciding whether to assign an object to one of the classes is the choice of the distance between $x(k)$ and c_j :

$$d(x(k), c_j) = \sqrt{(x(k) - c_j)^T A_j (x(k) - c_j)}, \quad (3)$$

where A_j – the inverse fuzzy covariance matrix of each cluster.

The peculiarity of this approach is insensitivity to the ratio of the number of objects to the number of indicators characterizing these objects, as well as to the law of data distribution. However, nothing is said about the exact requirements that the domain of the initial data should satisfy.

The principles of clustering objects in the conditions of data fuzziness describing them are described in [20]. Their essence is as follows.

Let n experiments be done and the result of j -th of them is determined by the set of values of the coordinates \bar{F}_{jp} , $j = 1, 2, \dots, n$, $p = 1, 2, \dots, m$. It is assumed that the true values of the coordinates F_{jp} are fuzzy numbers with the corresponding membership functions, for example of (L – R) type [21–23]:

$$\mu(F_{jp}) = \begin{cases} L\left(\frac{\bar{F}_{jp} - F_{jp}}{\alpha_{jp}}\right), & F_{jp} \leq \bar{F}_{jp}, \\ R\left(\frac{F_{jp} - \bar{F}_{jp}}{\beta_{jp}}\right), & F_{jp} > \bar{F}_{jp}, \end{cases} \quad (4)$$

where \bar{F}_{jp} – the measured value of the p -th coordinate in the j -th experiment, which is modal for a fuzzy number F_{jp} , $j = 1, 2, \dots, n$, $p = 1, 2, \dots, m$, α_{jp} , β_{jp} – left and right fuzzy coefficients in the description (4).

For each point, the fuzzy distance to each of the attraction centers of the clusters and the corresponding membership function are calculated. Then the obtained membership functions are used to find a cluster that has the highest degree of preference relative to the point under consideration. A formal description of this procedure is as follows.

For the pair (the k -th cluster, the j -th point), the fuzzy value of the square of the distance from the attraction center of the cluster to the point is introduced

$$\rho_{kj}^2 = \sum_{p=1}^m (a_{kp} - F_{jp})^2, \quad k = 1, 2, \dots, 2^m.$$

It is necessary to obtain the fuzzy number ρ_{kj}^2 membership function and for this purpose let's use known relations for the results of operations on fuzzy numbers of (L – R) type [22, 23].

Let, $A_{LR} = \langle a_1, \alpha_1, \beta_1 \rangle$, $B_{LR} = \langle a_2, \alpha_2, \beta_2 \rangle$ – fuzzy numbers of (L – R) type.

Then the parameters of the fuzzy number $C_{LR} = A_{LR} + B_{LR} = \langle a, \alpha, \beta \rangle$ are calculated by the formulas:

$$a = a_1 + a_2, \quad \alpha = \alpha_1 + \alpha_2, \quad \beta = \beta_1 + \beta_2.$$

The parameters of the number $C_{LR} = A_{LR} - B_{LR} = \langle a, \alpha, \beta \rangle$ are calculated by the formulas:

$$a = a_1 - a_2, \quad \alpha = \alpha_1 + \beta_2, \quad \beta = \alpha_2 + \beta_1.$$

The parameters of the number $C_{LR} = A_{LR} \cdot B_{LR} = \langle a, \alpha, \beta \rangle$ are calculated by the formulas:

$$a = a_1 a_2, \quad \alpha = |a_1| \alpha_2 + |a_2| \alpha_1, \quad \beta = |a_1| \beta_2 + |a_2| \beta_1.$$

In accordance with this, the parameters of fuzzy numbers

$$\Delta_{kjp} = a_{kp} - F_{jp}, \quad \Delta_{kjp}^2 = (a_{kp} - F_{jp})^2, \quad \rho_{jk}^2 = \sum_{p=1}^q \Delta_{kjp}^2$$

are determined by the relations:

$$\begin{aligned}\bar{\Delta}_{kjp} &= \bar{a}_{kp} - \bar{x}_{jp}, \quad \alpha'_{kjp} = \alpha_{kp}^{(a)} + \beta_{jp}^{(x)}, \quad \beta'_{kjp} = \alpha_{jp}^{(x)} + \beta_{kp}^{(a)}; \\ \bar{\Delta}_{kjp}^2 &= (\bar{\Delta}_{kjp})^2, \quad \alpha_{kjp} = 2|\bar{\Delta}_{kjp}|(\alpha_{kp}^{(a)} + \beta_{jp}^{(x)}), \quad \beta_{kjp} = 2|\bar{\Delta}_{kjp}|(\alpha_{jp}^{(x)} + \beta_{kp}^{(a)}); \\ \bar{p}_{jk}^2 &= \sum_{p=1}^q (\bar{\Delta}_{kjp}), \quad \alpha_{kj} = \sum_{p=1}^q \alpha_{kjp}, \quad \beta_{kj} = \sum_{p=1}^q \beta_{kjp}.\end{aligned}$$

In this case, the membership function of the fuzzy value of the square of the distance from the k-th center to the j-th point has the form:

$$\mu(F_{jp}) = \begin{cases} L\left(\left(\bar{\rho}_{kj}^2\right) - \rho_{kj}^2\right) / \alpha_{kj}, & \rho_{kj}^2 \leq \left(\bar{\rho}_{kj}^2\right), \\ R\left(\rho_{kj}^2 - \left(\bar{\rho}_{kj}^2\right)\right) / \beta_{kj}, & \rho_{kj}^2 > \left(\bar{\rho}_{kj}^2\right). \end{cases} \quad (5)$$

In this problem, the above general relations are simplified, since the coordinates of the attraction centers of clusters are clear numbers and therefore

$$\bar{\Delta}_{kjp} = a_{kp} - \bar{F}_{jp}, \quad \alpha'_{kjp} = \beta_{jp}, \quad \beta'_{kjp} = \alpha_{jp}.$$

As a result of the implementation of the described procedure, for each of the points, the fuzzy number membership functions will be obtained, representing the “distances” to the centers of the corresponding clusters. These numbers now need to be compared among themselves, choosing one of them for which the degree of preference for the rest will be the least. This number will determine the cluster, “closest” to the considered point. The estimation of the degree of fuzzy number preference z_k before a fuzzy number z_l is carried out according to the formula:

$$\eta(\mu(z_k), \mu(z_l)) = \sup_{z_k > z_l} \min\{\mu(z_k), \mu(z_l)\}, \quad k, l \in \{1, 2, \dots, m\}. \quad (6)$$

Using (6), the choice of a fuzzy number with the least degree of preference in relation to other numbers of the set of difficulties does not cause. The cluster number k^* to which the next point will be attached is determined by the relation

$$k^* = \arg \min_k \min_l \{\eta(z_k, \mu(z_l))\}, \quad k, l \in \{1, 2, \dots, m\}. \quad (7)$$

The above relations (5)–(7) provide a solution to the problem of fuzzy clustering.

Thus, the procedure described in [20] is based on calculating the distance between the classified point and the attraction center of the cluster – the point refers to the cluster whose distance to the COG is smaller. This idea can also be used to improve the classification accuracy if the linear discriminant function constructed using the parametric method of classification does not provide high recognition rates. In this regard, the aim of this article is to develop a classification method for incorrectly classified objects after constructing a linear discriminant function.

To achieve this aim, it is necessary to prove the principles of selecting informative attributes of the space of factors-attributes and an algorithm that allows to form a classifying rule.

2. Materials and methods of research

As a research material, a test sample of data, including objects of two classes A and B, is selected and presented in [24] (**Fig. 1**). The following assumptions are accepted in the paper:

- let as a result of the normalization of the values of the variables x_i characterizing the position of objects in the space of dimensionality characteristics ($N \times 2$) are within the range $[-1; +1]$;
- let it be known that N_1 objects belong to class A and N_2 objects belong to class B, with $N_1 = N_2 = 35$;
- the range of values for both classes is arbitrary.

The problem consists in constructing of a straight line dividing in the space of attributes both classes by means of parametric classification methods.

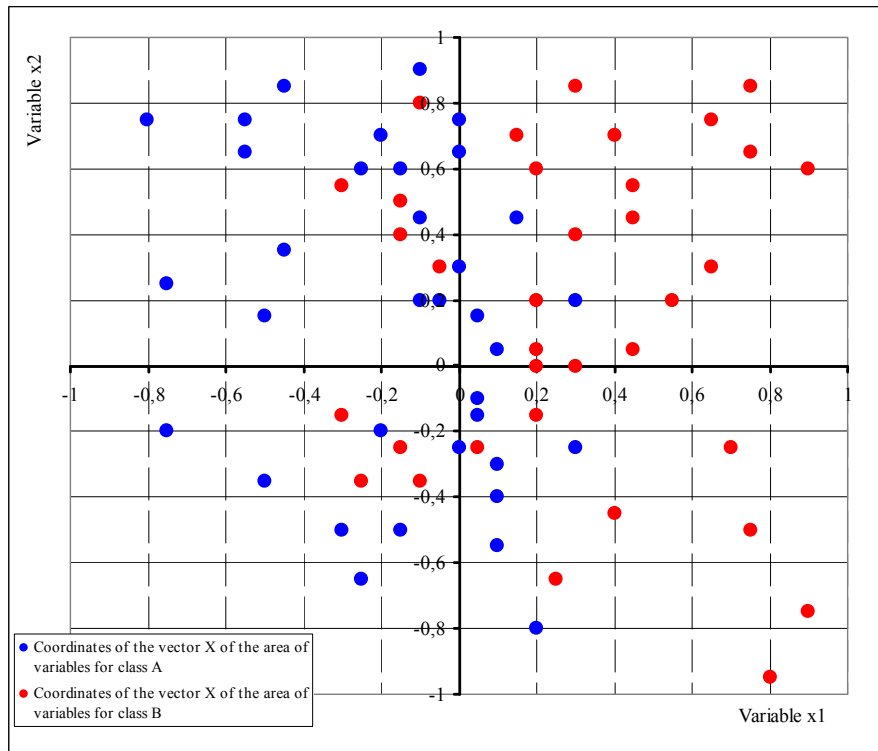


Fig. 1. Space of features for the test problem

The classifying rule using the likelihood ratio $l(x^j)$ is represented in the form [25]

$$L(x^j) = \frac{P(A)p_A(x^j)}{P(B)p_B(x^j)} = \frac{P(A)}{P(B)} l(x^j), \quad (8)$$

where $P(A)$, $P(B)$ – the a priori class probabilities, and the probability density $p_A(X)$ and $p_B(X)$ have the form:

$$\begin{aligned} p_A(X) &= d \exp\left(-\frac{1}{2}(x - m_A)^T \text{cov}^{-1}(x)(x - m_A)\right), \\ p_B(X) &= d \exp\left(-\frac{1}{2}(x - m_B)^T \text{cov}^{-1}(x)(x - m_B)\right). \end{aligned} \quad (9)$$

d – constant factor, m_A , m_B – the mathematical expectations of classes A and B respectively,

$$\text{cov}(x) = \frac{1}{N} X^T X - m^T m$$

is covariance matrix, $\text{cov}^{-1}(x)$ – inverse of the covariance matrix.

In this case the discriminant function has the form

$$y = f(x) = x' \text{cov}^{-1}(x)(m_A - m_B). \quad (10)$$

The threshold value of the discriminant function, which makes it possible to make a comparison for deciding whether an object belongs to a specific class, is calculated on the basis of equation

$$y_0 = \frac{1}{2}(m_A + m_B)' \text{cov}^{-1}(x)(m_A - m_B) - \ln \frac{P(A)}{P(B)}. \quad (11)$$

Let's suppose that the a priori probabilities of the classes are the same: $P(A)=P(B)=0,5$. The parameters of the distributions $p_A(x)$ and $p_B(x)$, calculated on the basis of the initial data of the test problem (**Fig. 1**), are:

$$m_A = \begin{pmatrix} -0,16143 \\ 0,135714 \end{pmatrix}, m_B = \begin{pmatrix} 0,29571 \\ 0,15429 \end{pmatrix},$$

$$\text{cov}_A(X) = \begin{pmatrix} 0,08644 & -0,0409 \\ -0,0409 & 0,22722 \end{pmatrix}, \text{cov}_B(X) = \begin{pmatrix} 0,12405 & -0,0085 \\ -0,0085 & 0,23991 \end{pmatrix}.$$

The classifying rule is constructed in the form of a linear discriminant function described by equation (10), in which the threshold value is determined analytically on the basis of (11). This discriminant function specifies the position of the dividing line in space ($N \times 2$) and the associated classification accuracy.

After implementation of the described procedures, a classification rule is constructed, the coefficients in the analytic description of which for the class A have the form:

$$x^j \in A \text{ if } -5,8225x_1 - 1,1292x_2 \geq -0,5547,$$

$$x^j \in B \text{ if } -5,8225x_1 - 1,1292x_2 < -0,5547. \quad (12)$$

The classifying rule, the coefficients in the analytical description of which are calculated for class B, has the form:

$$x^j \in A \text{ if } -3,6993x_1 - 0,2082x_2 \geq -0,2786,$$

$$x^j \in B \text{ if } -3,6993x_1 - 0,2082x_2 < -0,2786. \quad (13)$$

Fig. 2 shows the results of constructing the dividing lines described by the functions (12) and (13). The result, shown in **Fig. 2**, is obtained after eliminating the inaccuracy in calculating the coefficients in the description of the linear discriminant function, which presence can be defined as the error of the dividing line at the position $y=y_0$, which is proportional to the area of the shaded figure ABCD (**Fig. 3**) [24].

Fig. 2 shows that both dividing lines are close to each other, overlapping of classes A and B is observed, in which there is a fairly large number of objects from both classes. The classification accuracy, estimated by the probability of objects falling into the appropriate class using expressions (14) and (15), is 82.9 % for class A and 71.4 % for class B (using the classifying rule in the form (when using the classification rule in the form (12)).

$$P_A = \frac{n_A}{N_A} \cdot 100, \quad (14)$$

$$P_B = \frac{n_B}{N_B} \cdot 100, \quad (15)$$

where n_A, n_B – the number of correctly classified elements for classes A and B respectively, N_A, N_B – the total number of elements of classes A and B, respectively.

The results obtained using the classifying rule in the form (13) are approximately the same, due to the considerable proximity of the dividing lines.

The obtained result means that 6 objects from 35 are incorrectly classified for Class A, and 10 objects out of 35 are incorrectly classified for class B.

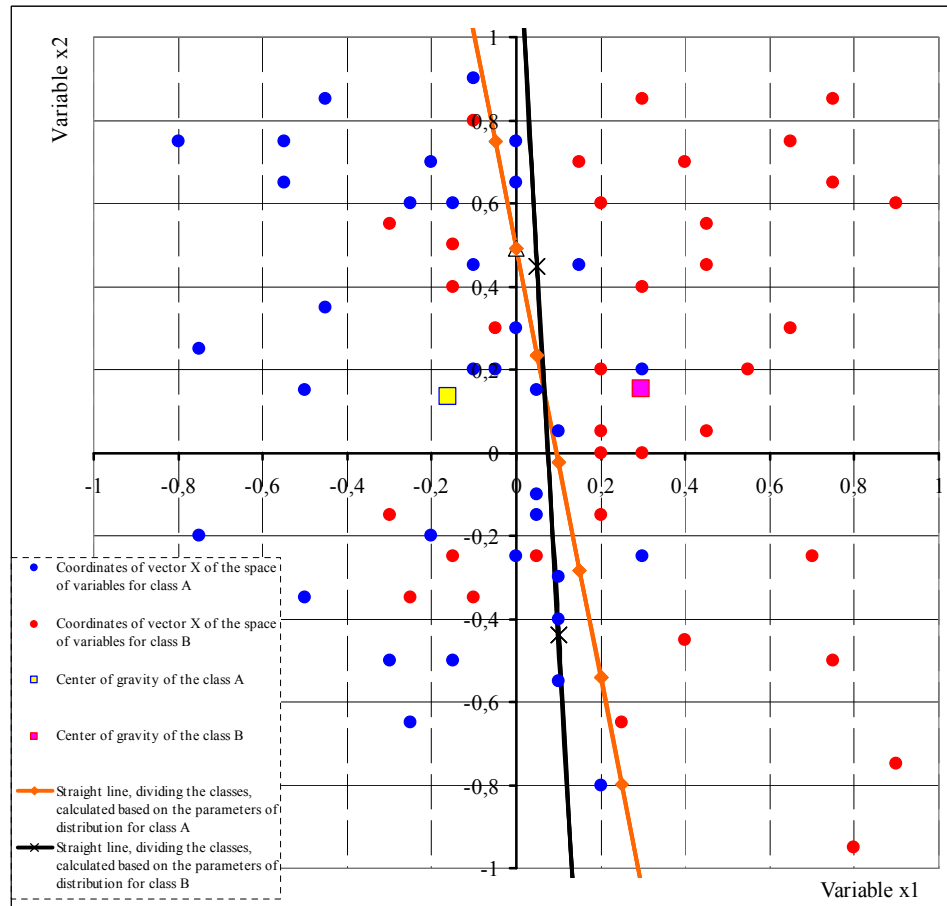


Fig. 2. Results of constructing the dividing straight lines in the space of dimensionality ($N \times 2$) for the test problem

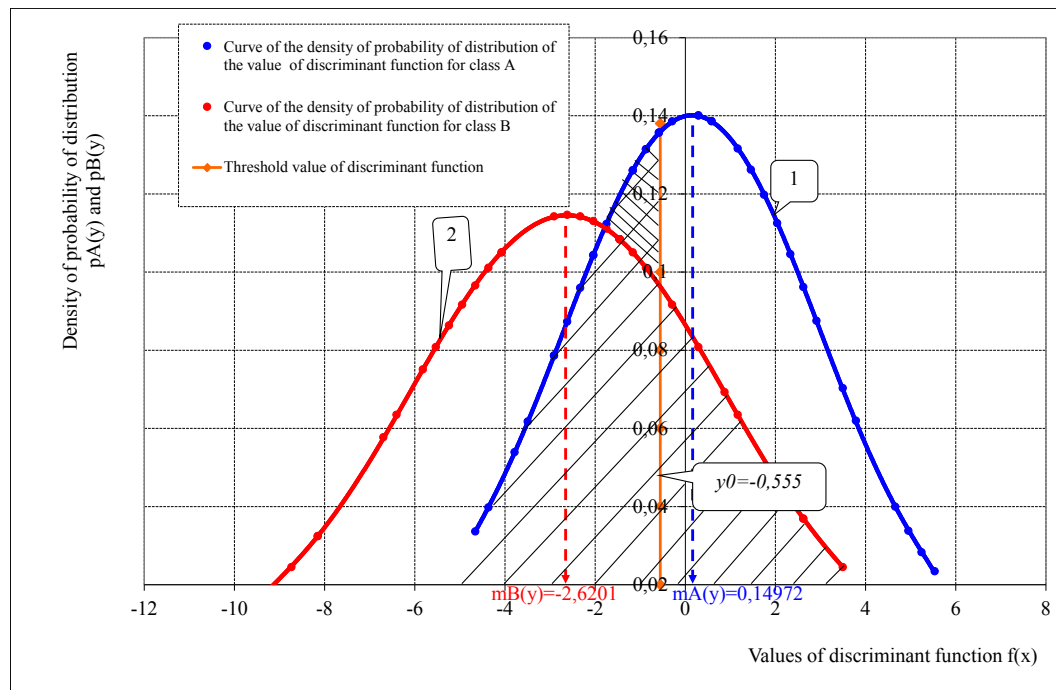


Fig. 3. Densities of probabilities of distribution $p_A(y)$ and $p_B(y)$ of the value of discriminant function for classes A and B: 1 – the curve $p_A(y)$, 2 – the curve $p_B(y)$

It is necessary to prove the principles of selecting informative attributes of the space of characteristic factors in such way as to maximize the proportion of correctly classified objects.

3. Research results

To select the informative attributes of the space of characteristic factors that allow to maximize the total fraction of properly classified objects by assigning incorrectly classified objects to the corresponding classes at the first stage - constructing a linear discriminant function - the following assumption can be used. An incorrectly classified object has a connection with its own cluster, but due to a number of reasons, a random character found itself in another cluster. If such connection is established, then it becomes possible to formalize the assignment of this object to its cluster. As such connection is determined, one can use the idea of “gravitation” of the object to its COG [20]. In this case, all the original object space must be “freed” from correctly classified objects and only incorrectly classified objects should be left in the test sample. The variant of data distribution after such transformation of the initial sample has the form shown in **Fig. 4**.

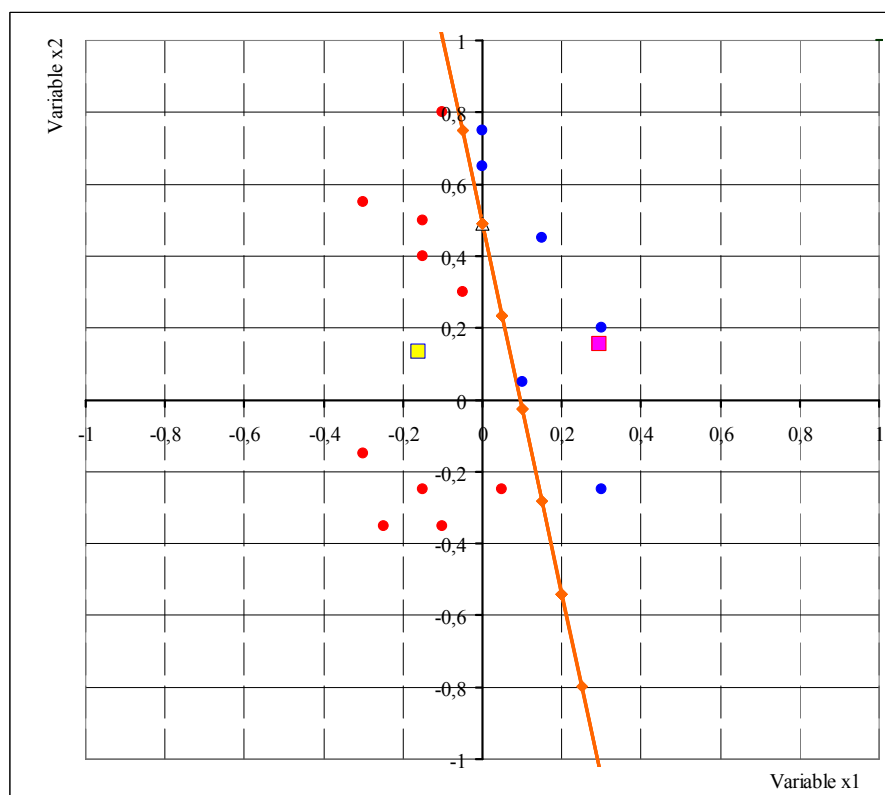


Fig. 4. Incorrectly classified objects of classes A ($n_A=6$) and B ($n_B=10$) after constructing a linear discriminant function in the form (12)

From this it follows that the position of the incorrectly classified object must be associated with two elements – the mathematical expectation of the corresponding class, which is its COG, and position of the dividing line. Such connection is represented in the form of the corresponding segments of straight lines representing the sides of a triangle with vertices: classifying point – COG – the intersection point of the normal that is drawn from the classified point to the dividing line. The COG of such “local” cluster made up of the vertices of this triangle is chosen as a new variable of the feature space. Consequently, it is from such COGs of local clusters that a new sample is formed to construct the classification rule. A demonstration of this idea is shown in **Fig. 5**.

From **Fig. 5** it can be seen that the implementation of the proposed idea provides the possibility of correct classification of previously incorrectly classified objects. Indeed, the COG of the

local cluster formed for an object with coordinates (0.3; -0.25) belongs to class A, but incorrectly assigned to class B (blue markers), is to the left of the dividing line. This means that an object with coordinates (0.3; -0.25) belongs to class A. An analogous conclusion can be made about an object with coordinates (-0.05, 0.3) belonging to class B, but incorrectly assigned to the class A (red markers) is to the right of the dividing line. This means that an object with coordinates (-0.05, 0.3) belongs to class B.

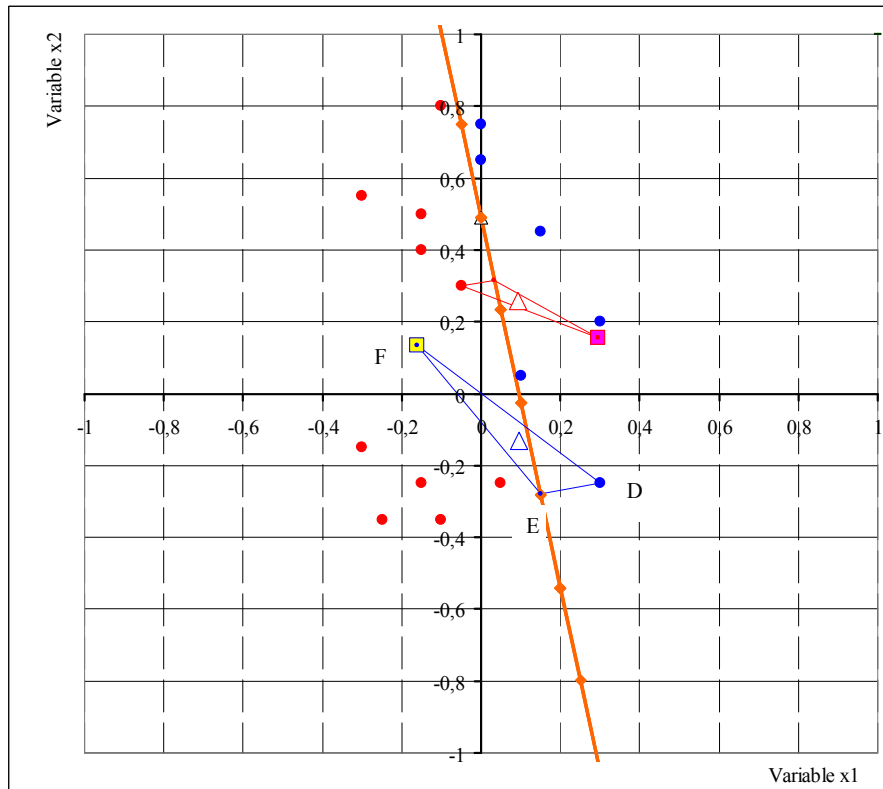


Fig. 5. Demonstration of the idea of constructing local clusters for formation of an informative data sample: Δ – COG of a local cluster DEF, \triangle – COG of a local cluster constructed for an incorrectly classified point of class B with coordinates (-0.05; 0.3)

Methods of analytical geometry are used to calculate the vertices of local clusters. Consider the i -th local cluster DEF, whose vertices have the coordinates $D(x_{1D}; x_{2D})$, $E(x_{1E}; x_{2E})$, $F(x_{1F}; x_{2F})$.

The coordinates of the vertex $D(x_{1D}; x_{2D})$ are known – they are the coordinates of the incorrectly classified point B in the example $(x_{1D}; x_{2D}) = (0.3; -0.25)$.

Coordinates of the vertex F $(x_{1F}; x_{2F})$ are known – they are the coordinates of the COG of class A: $(x_{1F}; x_{2F}) = (m_A(x_1); m_A(x_2)) = (-0.16143; 0.13571)$.

The coordinates of the vertex $E(x_{1E}; x_{2E})$ are unknown, but they can be found as the coordinates of the intersection point of the dividing line and the normal drawn to the given line from D. For this it is necessary to obtain the equation of the normal DE, knowing the equation of the dividing line and presenting it as

$$-5,8225x_1 - 1,1292x_2 + 0,5547 = 0. \quad (16)$$

The transformation of the general form of the line $a_1x_1 + a_2x_2 + a_0 = 0$ to the form $x_2 = kx_1 + b$ allows to determine the coefficients k and b :

$$k = -\frac{a_1}{a_2}, \quad b = -\frac{a_0}{a_2}.$$

The equation of the normal passing through the point $D(x_{1D}; x_{2D})$ has the form

$$x_2 = x_{2D} + k_1(x_1 - x_{1D}), \quad (17)$$

where

$$k_1 = -\frac{1}{k}.$$

The coordinate of the point $E(x_{1E}; x_{2E})$ is determined by solving a system of linear equations of the form:

$$\begin{aligned} -5,8225x_1 - 1,1292x_2 + 0,5547 &= 0, \\ -1,1292x_1 + 5,8225x_2 + (-5,8225x_{2D} - (-1,1292x_{1D})) &= 0. \end{aligned} \quad (18)$$

It is easy to verify this by carrying out simple transformations of equations (16) and (17).

Calculation based on the system (18) allows obtaining the following value of the coordinates of the point E: $(x_{1E}; x_{2E}) = (0,149; -0,279)$.

Similarly, using Microsoft Excel tools, the vertices of local clusters for all incorrectly classified points of class A are calculated. **Fig. 6** shows incorrectly classified objects of class A and the calculated coordinates of the COGs of the corresponding local clusters.

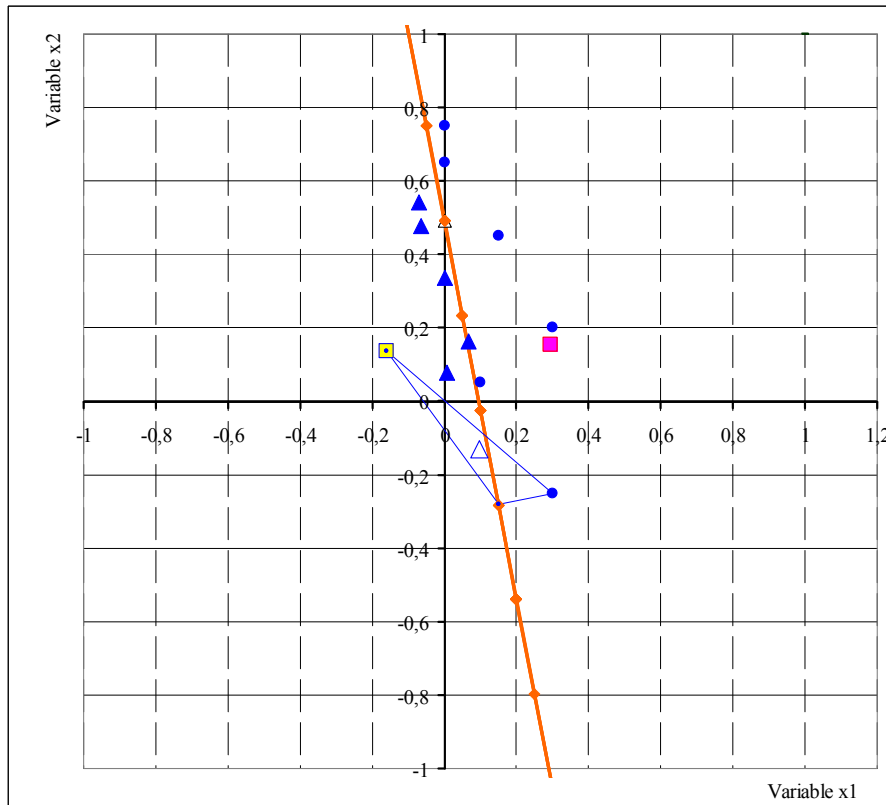


Fig. 6. Incorrectly classified objects of class A and calculated coordinates of the COGs of the corresponding local clusters

Fig. 7 shows the location of the COGs of local clusters for all six incorrectly classified objects of class A relative to the dividing line.

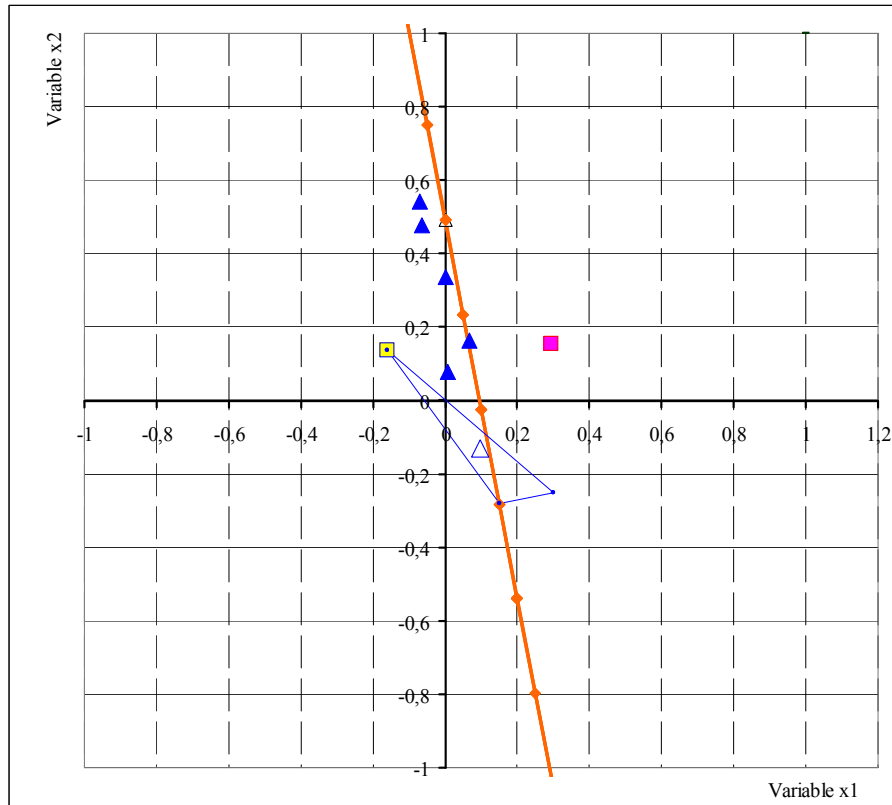


Fig. 7. Location of the COGs of local clusters for all six incorrectly classified objects of class A relative to the dividing line

As it follows from **Fig. 7**, five of the six objects can be assigned to their class A. One object with coordinates (0.3, 0.2) is in an undefined state - the COG of the corresponding local cluster has coordinates (0.067958, 0.163399) and practically falls on the dividing line. Taking into account the non-strict inequality of the classifying rule (12), it can be assigned to class A. So it is correctly classified. However, for a more accurate decision making, an accurate calculation of the distance from the COG of the corresponding local cluster to the dividing line is necessary. It is also necessary to determine the sign of the corresponding segment, located along the normal to the dividing surface.

The results of similar calculations of the COGs of local clusters constructed for incorrectly classified objects of class B are shown in **Fig. 8, 9**.

To determine the distance from the i -th COG to the dividing line, it is necessary to reduce its equation to the normal form by introducing a normalizing factor

$$\mu = \pm \frac{1}{\sqrt{a_1^2 + a_2^2}}. \quad (19)$$

Taking into account that the free term of equation (16) has a positive sign, the “-” sign should be taken before the right-hand side of equation (19). Multiplying (19) and (16) and substituting the coordinates of i -th COG, the distances d_i from the i -th COG to the dividing line are calculated. In this case, the sign in front of d_i makes it possible to determine the location of i -th COG relative to the dividing line, thus forming a classification rule:

$$\begin{aligned} x^j &\in A \text{ if } d_i \leq 0, \\ x^j &\in B \text{ if } d_i > 0. \end{aligned} \quad (20)$$

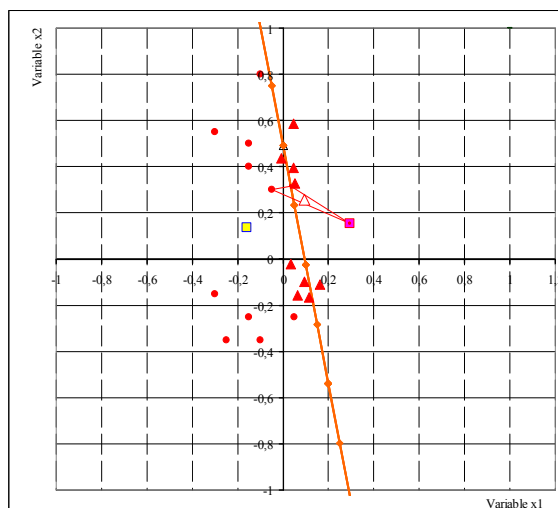


Fig. 8. Incorrectly classified objects of class B and calculated coordinates of the COGs of the corresponding local clusters

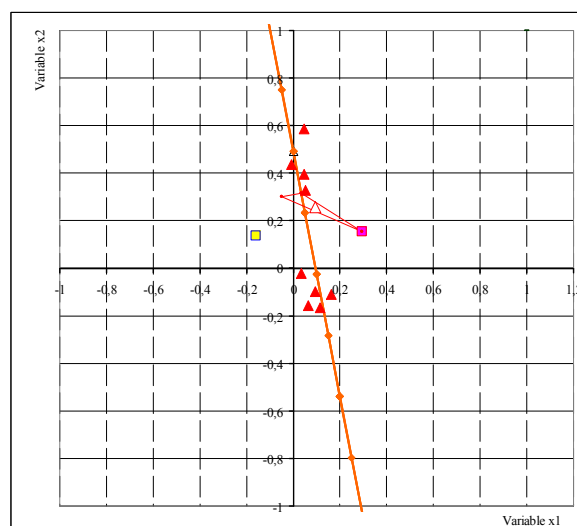


Fig. 9. Location of the COGs of local clusters for all ten incorrectly classified objects of class B relative to the dividing line

The results of calculations of distances d_i are given in **Table 1**.

Table 1

The results of applying the classifying rule (20)

d_i for classifying objects of class A *	d_i for classifying objects of class B*
	d_i
-0,05896	-0,01905
-0,06531	-0,06348
-0,07066	-0,05981
-0,02892	0,026859
-0,02425	0,020512
0,004304	-0,02074
	0,062259
	-0,01072
	0,04689
	0,044708

Note: * – incorrectly classified objects after applying the proposed procedure are in bold

4. Discussion of research results

As can be seen from the results obtained for incorrectly classified objects of class B (**Fig. 9**), the proposed procedure allows to solve the problem of correct classification of incorrectly classified after constructing the linear discriminant function of objects by only 50 %. The same is confirmed by the results of applying the proposed classification rule (20) – from **Table 1** it can be seen that 5 of 10 previously incorrectly classified objects are again classified incorrectly. Much better is the result for Class A – only 1 of the 6 objects is again classified incorrectly.

In terms of the effectiveness of the proposed classification procedure as a two-stage process in which a linear discriminant function is constructed at the first stage, and a classifying rule of the form (20) is applied on the second stage, then it is obvious. Indeed, for class A, the classification accuracy calculated using formula (14) increases from 82.9 % to 97.1 %, and for class B, the classification accuracy calculated using formula (15) increases from 71.4 % up to 85.7 %. Such result can be considered good, thus confirming the effectiveness of the proposed procedure.

Despite the obvious disadvantages associated with a number of assumptions inherent in all parametric classification methods [24], the proposed procedure can be used in practical recognition problems for feature factor spaces of $N \times 2$ dimension. Moreover, it has the potential for increasing accuracy, due to an increase in the accuracy of the location of the dividing line. The latter can be done by using the methods of experiment planning to construct a linear discriminant function. In particular, it is about the selection of objects of both classes having coordinates corresponding to the vertices of the plan of the full factorial design inscribed in the actual area of the spaces of factors-attributes of dimension $N \times 2$.

5. Conclusions

1. A procedure is proposed for selecting informative attributes of the space of characteristic factors that allow to maximize the total fraction of correctly classified objects by assigning incorrectly classified objects to the corresponding classes after the first stage of parametric classification – construction of a linear discriminant function. Its essence consists in establishing the connection of an incorrectly classified object precisely with its class. It is suggested, as such connection, to consider the position of an incorrectly classified object with respect to the mathematical expectation of the corresponding class, which is its COG, and the position of the dividing line, which has the form of a linear discriminant function. It is shown that in order to realize this idea it is necessary to develop local clusters that are a triangle with vertices: the classified point is the COG is the point of intersection of the normal drawn from the classified point to the dividing line. The COGs of such local clusters are chosen as new variables of feature space. Consequently, it is from such COGs of local clusters that a new sample is formed to construct the classification rule.

2. A classifying rule is proposed that allows one to classify an object incorrectly classified after constructing a linear discriminant function to its class. It is based on the determination of the position of the COGs of local clusters relative to the dividing line (distance d_i). This rule is as follows: $x^j \in A$ if $d_i \leq 0$, $x^j \in B$ if $d_i > 0$.

3. It is shown that for incorrectly classified objects the transition from use as the factors-attributes of their initial values to the use of the COGs of local clusters provides the possibility of increasing the classification accuracy by 14 %.

References

- [1] Nemrodov, D., Niemeier, M., Mok, J. N. Y., Nestor, A. (2016). The time course of individual face recognition: A pattern analysis of ERP signals. *NeuroImage*, 132, 469–476. doi: 10.1016/j.neuroimage.2016.03.006
- [2] Liu, L., Fieguth, P., Zhao, G., Pietikäinen, M., Hu, D. (2016). Extended local binary patterns for face recognition. *Information Sciences*, 358-359, 56–72. doi: 10.1016/j.ins.2016.04.021
- [3] Kushnirenko, O. A., Cadchenko, A. V., Troyanskiy, A. V. (2013). Correlation algorithm of recognition of character information with low-complexity of technical implementation. *Eastern-European Journal of Enterprise Technologies*, 3 (9 (63)), 42–45. Available at: <http://journals.uran.ua/eejet/article/view/14848/12650>

- [4] Kasian, K., Bratchykov, V., Shkarupylo, V. (2015). Development of modified method for text recognition in standardized picture. *Eastern-European Journal of Enterprise Technologies*, 3 (2 (75)), 11–17. doi: 10.15587/1729-4061.2015.43047
- [5] Anokhin, M. (2014). Moving objects recognition method by their video images. *Eastern-European Journal of Enterprise Technologies*, 4 (9 (70)), 33–37. doi: 10.15587/1729-4061.2014.26275
- [6] Voloshyn, N. (2012). Identification of iris sign in automated systems of iridology based on petri nets. *Technology Audit and Production Reserves*, 4(1(6)), 5–6. doi: 10.15587/2312-8372.2012.4767
- [7] Alfaro-Ponce, M., Argüelles, A., Chairez, I. (2016). Pattern recognition for electroencephalographic signals based on continuous neural networks. *Neural Networks*, 79, 88–96. doi: 10.1016/j.neunet.2016.03.004
- [8] Nesterenko, S. A., Stanovskiy, A. O., Oborotova, O. O. (2015). The wireless computer networks state recognition over the three-dimensional field of directions. *Technology audit and production reserves*, 6 (2 (26)), 28–35. doi: 10.15587/2312-8372.2015.56825
- [9] Demin, D. (2010). Priniatie reshenii v protsesse upravleniia elektroplavkoi s uchetom faktorov nestabil'nosti tehnologicheskogo protsesssa. *Bulletin of NTU «KhPI». Series: New desicions of modern technologies*, 17, 67–72.
- [10] Demin, D., Bozhko, A., Zraichenko, A., Nekrasov, A. (2006). Identifikatsiia chuguna dlia opredele-niia ratsional'nyh rezhimov legirovaniia. *Eastern-European Journal of Enterprise Technologies*, 4 (1 (22)), 29–32.
- [11] Vasenko, Y. (2012). Technology for improved wear iron. *Technology Audit and Production Re-serves*, 1 (1 (3)), 17–21. doi: 10.15587/2312-8372.2012.4870
- [12] Ponomarenko, O. I., Tenev, N. S. (2013). Computer modeling of crystallization processes as a reserve of improving the quality of pistons of ICE. *Technology Audit and Production Reserves*, 6(2(14)), 36–40. doi: 10.15587/2312-8372.2013.19529
- [13] Stuper, A. J., Brugger, W. E., Jurs, P. C. (1979). *Computer Assisted Studies of Chemical Struc-ture and Biological Function*. Published by John Wiley & Sons.
- [14] Arsirii, E., Manikaeva, O., Vasilevskaja, O. (2015). Development of the decision support sub-system in the systems of neural network pattern recognition by statistical information. *Eastern-European Journal of Enterprise Technologies*, 6 (4 (78)), 4–12. doi: 10.15587/1729-4061.2015.56429
- [15] Frazee-Frazenko, A. (2012). Algorithm of study neural network for image recognition. *Technolo-gy Audit and Production Reserves*, 4 (1 (6)), 33–34. doi: 10.15587/2312-8372.2012.4781
- [16] Fertsev, A. A. (2012). Uskorenie obucheniya neyronnoy seti dlya raspoznavaniya izobrazheniy s pomoshchyu tehnologii NVidia Cuda. *Vestnik Samarskogo gosudarstvennogo tehnikeskogo universiteta*, 1 (26), 183–191.
- [17] Unglert, K., Radić, V., Jellinek, A. M. (2016). Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra. *Journal of Volcanology and Geothermal Research*, 320, 58–74. doi: 10.1016/j.jvolgeores.2016.04.014
- [18] Fakhar, K., El Aroussi, M., Saidi, M. N., Aboutajdine, D. (2016). Fuzzy pattern recognition-based approach to biometric score fusion problem. *Fuzzy Sets and Systems*. doi: 10.1016/j.fss.2016.05.005
- [19] Perova, I. G. (2014). Adaptive treatment of these medico-biological researches by methods of computational intelligence. *Eastern-European Journal of Enterprise Technologies*, 1 (4 (67)), 24–28. doi: 10.15587/1729-4061.2014.21202
- [20] Demin, D. (2013). Mathematical modeling in the problem of selecting opti-mal control of obtain-ing alloys for machine parts in un-certainty conditions. *Problems Of Mechanical Engineering*, 16 (6), 15–23. Available: <http://journals.urau.ua/jme/article/view/21309>
- [21] Seraya, O. V., Demin, D. A. Linear regression analysis of a small sample of fuzzy input data (2012) *Journal of Automation and Information Sciences*, 44 (7), 34–48.
- [22] Djubua, D. (1990). *Teorija vozmozhnostej. Prilozhenie k predstavleniju znanij v informatike*. Moscow: Radio i svjaz', 286.
- [23] Raskin, L. G., Seraja, O. V. (2008). *Nechetkaja matematika*. Kharkiv: Parus, 352.
- [24] Aouati, M. (2016). Localization of vectors–patterns in the problems of parametric classification with the purpose of increasing its accuracy. *Eastern-European Journal of Enterprise Technologies*, 4 (4 (82)), 10–20. doi: 10.15587/1729-4061.2016.76171
- [25] Hartmann, K., Lezki, E., Schafer, V. (1974). *Statistische Versuchsplanung und auswertung in der Stoffwirtschaft*. Leipzig.