Journal of Global Optimization (2006) 35: 215–234
DOI 10.1007/s10898-005-3836-2

The Method of Virtual Experts in Mathematical Diagnostics*

VERONIKA V. DEMYANOVA and VLADIMIR F. DEMYANOV
Applied Mathematics Department, St. Petersburg State University, Staryi Peterhof, 198504, St. Petersburg, Russia (e-mail: avd@ad9503.spb.edu)

(Accepted 6 October 2005)

Abstract. Mathematical Diagnostics (MD) deals with identification problems arising in different practical areas. Some of these problems can be described by mathematical models where it is required to identify points belonging to two or more sets of points. Most of the existing tools provide some identification rule (a classifier) by means of which a given point is assigned (attributed) to one of the given sets. Each classifier can be viewed as a virtual expert. If there exist several classifiers (experts), the problem of evaluation of experts' conclusions arises. In the paper for the case of supervised classification the method of virtual experts (the VE-method) is described. Based on this method, a generalized VE method is proposed where each of the classifiers can be chosen from a given family of classifiers. As a result, a new optimization problem with a discontinuous functional is stated. Examples illustrating the proposed approach are provided.

Key words: classifier, data mining, identification, identification rule, mathematical diagnostics, separation, virtual experts method

1. Introduction

Many problems of practical importance (such as, e.g., Data Mining problems, Classification and Identification problems, Pattern Recognition, Medical and Engineering Diagnostics, Assignment and Allocation problems) can be described by mathematical models where it is required to identify points belonging to two or more sets of points. Different approaches and theories exist to treat the above problems: Machine Learning, Support Vector Machines, Cluster Analysis, Neural Systems (see, e.g., Bennet and Mangasarian, 1992; Mangasarian, 1994; Advances in Kernel Methods, 1999; Cristianini and Shawe-Taylor, 2000; Vapnik, 2000; Lee and Mangasarian, 2001; Cucker and Smale, 2001; Bagirov et al., 2003 and references therein).

Most of the existing tools provide some identification rule (a classifier) by means of which a given point is assigned (attributed) to one of the given sets. These sets are assumed to be *finite* representative samplings of some unknown sets.

*The work of the second author was supported by the Russian Foundation for Fundamental Studies (RFFI) under Grant No 03-01-00668.

There exist two main types of classification problems: supervised classification problems and unsupervised ones. In the present paper only the first class of problems is discussed (the case where one is able to evaluate the quality of a chosen classifier). The problem is to find a (possibly) simple rule to identify points. The quality of a classifier is usually measured by some functional (for example, by the amount of misclassified points). The problem of finding such a classifier is often reduced to some optimization problem in a multidimensional space (the problem of natural and surrogate functionals and related optimization problems is discussed in Demyanov, 2005). This multi-dimensional optimization problem is sometimes replaced by an optimization problem in a lower-dimensional space by choosing a small number of most informative coordinates (parameters or features) or their linear (or nonlinear) combinations. The problem of defining the most informative coordinates is solved, e.g., by ranking coordinates. Low-dimensional (especially one- and two-dimensional) identification problems can successfully be solved (an algorithm in the one-dimensional case is described in Section 6). Solving identification problems for different subsets of coordinates (features) one may get several classifiers for the same database. Each classifier can be viewed as a virtual expert. The problem of evaluation of experts' conclusions is important in the decision making theory. This problem is discussed in the present paper for the case of supervised classification. One may expect that the usage of conclusions of several "experts" can produce a new identification rule which is in one way or another better than the individual ones. In the case of unsupervised classification different approaches (based on probabilistic considerations) are often employed. They are not discussed here, as well as very important problems related to training and testing sets, dependency of parameters (features), multiple cross-validation procedures etc.

The paper is organized as follows. In Section 2 the problem of identification rules is stated. A new method (the method of virtual experts – the VE-method) is discussed in Section 3. A generalized method of virtual experts is outlined in Section 5. It is assumed that each classifier from the given collection of classifiers can be chosen from the corresponding family of classifiers. Therefore the problem of finding the best collection of classifiers arises leading to a new optimization problem (which is discontinuous and multiextremal). The notion of multi-dimensional classifier is introduced. Two illustrative examples are presented in Section 4. An algorithm for the one-dimensional identification problem (the 1D-identification problem) is described in Section 6.

It is also worth noting that *the classification problems* (*with the existing databases*) *represent a perfect testing ground for different numerical methods as well as for different theories and approaches to solving identification problems.*

2. Identification Rules

The identification problem can be formulated as follows. Assume that two sets $\mathcal{A} \subset R^n$ and $\mathcal{B} \subset R^n$ are given. These sets are assumed to be *finite* representative samplings of some unknown sets A and B.

Let $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$. It is required to find a rule (called an *identification rule*(*IR*)) to identify points of $\mathcal{C}$. Of course, one is interested to get as simple an IR as possible. Usually the identification is performed by means of some functional (called a *classifier*) in the following way.

If $f: R^n \to R$ is a classifier and $c \in \mathcal{C}$ is given then the point c is attributed to the set $\mathcal{A}$ if $f(c) > 0$, and to the set $\mathcal{B}$ if $f(c) < 0$. If $f(c) = 0$ then, by definition, the point c is considered as unidentifiable by the classifier f.

However, an identification rule may take a more complicated form (one of them will be described in Section 3). If every $c \in \mathcal{C}$ can be correctly identified (*well-classified*) by a functional f, we say that the sets $\mathcal{A}$ and $\mathcal{B}$ are *perfectly classified* by the functional f (by the classifier f). It may happen that some points of $\mathcal{C}$ are wrongly identified (*misclassified*) by the classifier f. Denote by $\mathcal{A}^+ \subset \mathcal{A}$ the set of points of $\mathcal{A}$ which are correctly identified by the classifier f, i.e.

$$\mathcal{A}^+ = \{c \in \mathcal{A} \mid f(c) > 0\},$$

and by $\mathcal{A}^- \subset \mathcal{A}$ – the set of points of $\mathcal{A}$ which are wrongly identified by the classifier f, i.e.

$$\mathcal{A}^- = \{c \in \mathcal{A} \mid f(c) \leqslant 0\}.$$

Analogously, let us denote by $\mathcal{B}^+ \subset \mathcal{B}$ the set of points of $\mathcal{B}$ which are correctly identified by the classifier f, i.e.

$$\mathcal{B}^+ = \{c \in \mathcal{B} \mid f(c) < 0\},$$

and by $\mathcal{B}^- \subset \mathcal{B}$ – the set of points of $\mathcal{B}$ which are wrongly identified by the classifier f, i.e.

$$\mathcal{B}^- = \{c \in \mathcal{B} \mid f(c) \geqslant 0\}.$$

Note that $\mathcal{A}^+ \cup \mathcal{A}^- = \mathcal{A}$, $\mathcal{B}^+ \cup \mathcal{B}^- = \mathcal{B}$. The sets $\mathcal{A}^+$, $\mathcal{A}^-$, $\mathcal{B}^+$, $\mathcal{B}^-$ depend on f.

The quality of an identification rule (in our case the quality of a classifier) is measured by the amount of misclassified points. For example, if $\mathcal{A}$ and $\mathcal{B}$ are finite sets (i.e. each consisting of a finite number of points), then to evaluate the quality of a classifier one can use one of the following criteria:

$$\varphi_1(f) = |\mathcal{A}^-(f)| + |\mathcal{B}^-(f)|, \tag{1}$$

$$\varphi_2(f) = \frac{|\mathcal{A}^-(f)|}{|\mathcal{A}|} + \frac{|\mathcal{B}^-(f)|}{|\mathcal{B}|}, \tag{2}$$

$$\varphi_3(f) = \max\left\{|\mathcal{A}^-(f)|, |\mathcal{B}^-(f)|\right\}, \tag{3}$$

$$\varphi_4(f) = \max\left\{\frac{|\mathcal{A}^-(f)|}{|\mathcal{A}|}, \frac{|\mathcal{B}^-(f)|}{|\mathcal{B}|}\right\}. \tag{4}$$

Here $|A|$ is the car dinality of the set A (the number of points of A).

If $\mathcal{F}$ is a family of classifiers (or, in general, identification rules) and $\varphi(f)$ is a chosen criterion then the identification problem can be formulated as follows.

Find an $f^* \in \mathcal{F}$ such that

$$\varphi(f^*) = \min_{f \in \mathcal{F}} \varphi(f). \tag{5}$$

If $\mathcal{A} \cap \mathcal{B} = \emptyset$ then, in principle, it is possible to construct a functional f which performs a *perfect identification* (each $c \in \mathcal{A} \cup \mathcal{B}$ can be correctly identified). For example, if one takes

$$f(c) = \begin{cases} 1, & c \in \mathcal{A}, \\ -1, & c \in \mathcal{B}, \\ 0, & c \notin \mathcal{A} \cup \mathcal{B}, \end{cases}$$

then the sets $\mathcal{A}$ and $\mathcal{B}$ are perfectly identified by f. However, this trivial solution is not acceptable from practical considerations.

If $\mathcal{A} \cap \mathcal{B} \neq \emptyset$ then a perfect identification is impossible, and the best possible classifier (or IR) is the one which allows to identify all points of the set $(\mathcal{A} \cup \mathcal{B}) \setminus (\mathcal{A} \cap \mathcal{B})$.

The set $\mathcal{A} \cap \mathcal{B}$ is the set of *essentially unidentifiable* (by the classifier f) *points*.

The simplest case where each of the sets $\mathcal{A}$ and $\mathcal{B}$ contains a finite number of points in the one-dimensional space is discussed in Section 6. An algorithm for finding the best linear classifier is described in subsection 6.2 (see also Demyanova, 2004). The idea of isolation (see Astorino and Gaudioso, 2002) in the 1D-case is also easily implementable (see Petrova, 2004; Varis, 2004).

In subsection 6.3 it is shown how to use the one-dimensional identification procedure for ranking the parameters. Having obtained the mentioned ranking list, one is able to choose several most essential parameters (features) for constructing an identification rule for the entire sets $\mathcal{A}$ and $\mathcal{B}$ using the chosen features. Different approaches are used to find such an

IR (see, e.g., Bennet and Mangasarian, 1992; Advances in Kernel Methods, 1999; Vapnik, 2000; Lee and Mangasarian, 2001; Kokorina, 2002a; Kokorina, 2002b; Bagirov et al., 2003; Bagirov, 2005; Demyanov, 2005).

Another quite simple case is the two-dimensional one (where each of the sets $\mathcal{A}$ and $\mathcal{B}$ contains a finite number of points in the two-dimensional space). This case can easily be visualized, and a proper IR can be constructed.

In the subsequent section a new method is described which makes use of several classifiers (usually obtained by solving some low-dimensional identification problems) for constructing a better identification rule.

Remark 1. In the discussion above it was assumed that identification rules are of the deterministic nature. However, there are IRs where the classifier gives the answer in terms of probability: the probability that a point $c \in \mathcal{A} \cup \mathcal{B}$ belongs to the set $\mathcal{A}$ is $p_{\mathcal{A}}(c)$, and the probability that a point c belongs to $\mathcal{B}$ is $p_{\mathcal{B}}(c)$. In the sequel only deterministic identification rules are used though the approach described can be extended to probabilistic IRs as well.

Remark 2. The problem stated is *a supervised classification problem* since in the process of constructing a classifier, one is able to check the quality of the classifier. An unsupervised classification problem is to divide one given set (which is the union of some unknown sets) into two or more sets with similar features (and the problem is to define such a similarity) in the hope that the sets thus constructed are close to the initial ones.

3. The Method of Virtual Experts

Consider in detail the case where each of the sets $\mathcal{A}$ and $\mathcal{B}$ contains a finite number of points:

$$\mathcal{A} = \{a_i \in R^n \mid i \in I\}, \quad \mathcal{B} = \{b_j \in R^n \mid j \in J\},$$

where $I = 1:N_1,\ J = 1:N_2$. Put

$$\mathcal{C} = \mathcal{A} \cup \mathcal{B} = \{c_k \in R^n \mid i \in K\},$$

where $K = 1:N,\ N = N_1 + N_2$,

$$c_k = \begin{cases} a_k, & k \in 1:N_1, \\ b_{k-N_1}, & k \in (N_1+1):N. \end{cases}$$

Assume that $\mathcal{A} \cap \mathcal{B} = \emptyset$ (the sets have no common points).

Let $f_1, \dots, f_m$ be given classifiers. Every classifier $f_s, s \in 1:m$, produces the sets $\mathcal{C}_{s\mathcal{A}}$ and $\mathcal{C}_{s\mathcal{B}}$ such that $\mathcal{C}_{s\mathcal{A}} \subset \mathcal{C}$, $\mathcal{C}_{s\mathcal{B}} \subset \mathcal{C}$,

$$\mathcal{C}_{s\mathcal{A}} = \{c_k \in \mathcal{C} \mid i \in K, \ c_k \ \text{ is assigned to } \mathcal{A}\},$$
$$\mathcal{C}_{s\mathcal{B}} = \{c_k \in \mathcal{C} \mid i \in K, \ c_k \ \text{ is assigned to } \mathcal{B}\}.$$

Let us suppose that

$$\mathcal{C}_{s\mathcal{A}} \cap \mathcal{C}_{s\mathcal{B}} = \emptyset, \quad \mathcal{C}_{s\mathcal{A}} \cup \mathcal{C}_{s\mathcal{B}} = \mathcal{C}$$

(i.e. it is assumed that every classifier assignes each point $c \in \mathcal{C}$ to one of the sets $\mathcal{A}$ or $\mathcal{B}$, uncertainty is excluded).

Denote

$$\mathcal{A}_s^+ = \mathcal{A} \cap \mathcal{C}_{s\mathcal{A}}, \ \mathcal{B}_s^+ = \mathcal{B} \cap \mathcal{C}_{s\mathcal{B}}, \ \mathcal{A}_s^- = \mathcal{A} \cap \mathcal{C}_{s\mathcal{B}}, \ \mathcal{B}_s^- = \mathcal{B} \cap \mathcal{C}_{s\mathcal{A}}.$$

The set $\mathcal{A}_s^+$ is the subset of points of the set $\mathcal{A}$ which are correctly identified by the classifier f_s (well-classified points of the set $\mathcal{A}$), the set $\mathcal{B}_s^+$ is the subset of points of the set $\mathcal{B}$ which are correctly identified by the classifier f_s (well-classified points of the set $\mathcal{B}$), the set $\mathcal{A}_s^-$ is the subset of points of the set $\mathcal{A}$ which are incorrectly identified by the classifier f_s (misclassified points of $\mathcal{A}$), the set $\mathcal{B}_s^-$ is the subset of points of the set $\mathcal{B}$ which are incorrectly identified by the classifier f_s (misclassified points of $\mathcal{B}$).

The classifier f_s will be referred to as *a virtual expert*. We shall use the information provided by the classifiers $f_s, s \in 1:m$, to construct a new classifier in the following way. For every point $c \in \mathcal{C}$ put

$$e_s(c) = \begin{cases} 1, & c \in \mathcal{C}_{s\mathcal{A}}, \\ 2, & c \in \mathcal{C}_{s\mathcal{B}}, \end{cases}$$

$e(c) = (e_1(c), e_2(c), \dots, e_m(c))$.

The vector $e(c)$ can take 2^m values (m-dimensional vectors whose coordinates are equal to 1 or 2). By $\mathcal{E}$ denote the set of all possible values of $e(c)$. The set $\mathcal{C}$ will be divided into 2^m subsets $\mathcal{C}_E, \ E \in \mathcal{E}$:

$$\mathcal{C}_E = \{c \in \mathcal{C} \mid e(c) = E\}. \tag{6}$$

Some of the subsets $\mathcal{C}_E$ may be empty. Note that

$$\mathcal{C}_{E_1} \cap \mathcal{C}_{E_2} = \emptyset \ \forall E_1 \neq E_2; \quad \bigcup \{\mathcal{C}_E \mid E \in \mathcal{E}\} = \mathcal{C}.$$

For every $E \in \mathcal{E}$, let us construct the sets

$$\mathcal{A}_E = \mathcal{A} \cap \mathcal{C}_E, \quad \mathcal{B}_E = \mathcal{B} \cap \mathcal{C}_E. \tag{7}$$

We shall use the following identification rule for a point $c \in \mathcal{C}_E$:

$$\text{if } |\mathcal{A}_E| \geqslant |\mathcal{B}_E|, \text{ then } c \text{ is assigned to the set } \mathcal{A}; \tag{8}$$

$$\text{if } |\mathcal{A}_E| < |\mathcal{B}_E|, \text{ then } c \text{ is assigned to the set } \mathcal{B}. \tag{9}$$

Remind that $|A|$ is the cardinality of A (the number of points of A).

The described identification rule can be written in the traditional form (see the beginning of Section 2) by means of the following classifier f:

$$f(c) = \begin{cases} 1, & c \in \mathcal{C}_E, \ |\mathcal{A}_E| \geqslant |\mathcal{B}_E|, \\ -1, & c \in \mathcal{C}_E, \ |\mathcal{A}_E| < |\mathcal{B}_E|. \end{cases}$$

Thus, the new classifier f divides the set $\mathcal{C}$ into several (not more than 2^m) subsets, and in every subset a new identification rule is valid. It is natural to expect that the classifier thus constructed will be (in general) more informative and effective than the initial individual IRs.

The described procedure will be called *the method of virtual experts* (the VE-method). This method has been tested on several specific databases and proved to be quite effective. A detailed report on the results of its application to widely used databases will be published elsewhere.

Let us illustrate the proposed method by several examples.

4. Illustrative Examples

4.1. EXAMPLE 1.

Let $x = (x_1, x_2) \in R^2$, $\quad \mathcal{A} = \{a_i \in R^2 \ | \ i \in 1 : N_1\}, \mathcal{B} = \{b_j \in R^2 \ | \ j \in 1 : N_2\}$, where

$$\begin{aligned} &N_1 = 7, \ a_1 = (1,2), \ a_2 = (2,2), \ a_3 = (3,2), \\ &a_4 = (1,1), \ a_5 = (2,1), \ a_6 = (3,1), \ a_7 = (-2,-1); \\ &N_2 = 7, \ b_1 = (-1,3), \ b_2 = (-1,2), \ b_3 = (-1,1), \\ &b_4 = (2,-1), \ b_5 = (3,-1), \ b_6 = (4,-1), \ b_7 = (2,-2). \end{aligned}$$

Put $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$. Let the classifiers $f_1(x) = x_1$ and $f_2(x) = x_2$ be given with the following identification rules: a point $c = (c_1, c_2) \in \mathcal{C}$ is assigned by the classifier $f_s, s \in 1:2$, to the set $\mathcal{A}$ if $f_s(c) > 0$, and to $\mathcal{B}$ – if $f_s(c) < 0$.

It is clear (see Figure 1 and Section 3) that

$$\begin{aligned} &\mathcal{C}_{1\mathcal{A}} = \{a_1 - a_6, b_4 - b_6\}, \quad \mathcal{C}_{1\mathcal{B}} = \{b_1 - b_3, a_7\}, \\ &\mathcal{A}_1^+ = \mathcal{A} \cap \mathcal{C}_{1\mathcal{A}} = \{a_1 - a_6\}, \ \mathcal{B}_1^+ = \mathcal{B} \cap \mathcal{C}_{1\mathcal{B}} = \{b_1 - b_3\}, \\ &\mathcal{A}_1^- = \mathcal{A} \cap \mathcal{C}_{1\mathcal{B}} = \{a_7\}, \ \mathcal{B}_1^- = \mathcal{B} \cap \mathcal{C}_{1\mathcal{A}} = \{b_4 - b_7\}. \end{aligned}$$

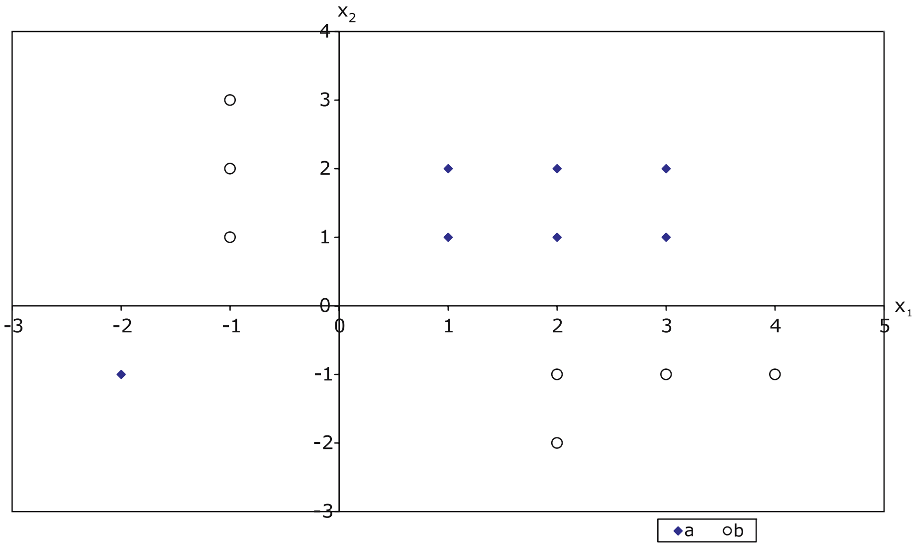

Figure 1.

We conclude that the points $a_7, b_4 - b_7$ are misclassified by the classifier f_1 (altogether 5 misclassified points).

In this section we use the following notation: $\{a_1 - a_6, b_4 - b_6\}$ means $\{a_i \mid i \in 1:6\} \cup \{b_j \mid j \in 4:6\}$.

Analogously, we have

$$\mathcal{C}_{2\mathcal{A}} = \{a_1 - a_6, b_1 - b_3\}, \quad \mathcal{C}_{2\mathcal{B}} = \{b_4 - b_7, a_7\},$$
$$\mathcal{A}_2^+ = \mathcal{A} \cap \mathcal{C}_{2\mathcal{A}} = \{a_1 - a_6\}, \ \mathcal{B}_2^+ = \mathcal{B} \cap \mathcal{C}_{2\mathcal{B}} = \{b_4 - b_7\},$$
$$\mathcal{A}_2^- = \mathcal{A} \cap \mathcal{C}_{2\mathcal{B}} = \{a_7\}, \ \mathcal{B}_2^- = \mathcal{B} \cap \mathcal{C}_{2\mathcal{A}} = \{b_1 - b_3\}.$$

Again, we conclude that the points $a_7, b_1 - b_3$ are misclassified by the classifier f_2 (altogether 4 misclassified points).

The set $\mathcal{E}$ contains 4 points: $\mathcal{E} = \{E_1, E_2, E_3, E_4\}$ where $E_1 = (1,1)$, $E_2 = (1,2)$, $E_3 = (2,1)$, $E_4 = (2,2)$. We have (see (6))

$$\mathcal{C}_{E_1} = \{c \in \mathcal{C} \mid e(c) = E_1\} = \{a_1 - a_6\}, \ \mathcal{C}_{E_2} = \{c \in \mathcal{C} \mid e(c) = E_2\} = \{b_4 - b_7\},$$
$$\mathcal{C}_{E_3} = \{c \in \mathcal{C} \mid e(c) = E_3\} = \{b_1 - b_3\}, \ \mathcal{C}_{E_4} = \{c \in \mathcal{C} \mid e(c) = E_4\} = \{a_7\}.$$

Now let us construct the sets (see (7)) $\quad \mathcal{A}_{E_s} = \mathcal{A} \cap \mathcal{C}_{E_s}, \quad \mathcal{B}_{E_s} = \mathcal{B} \cap \mathcal{C}_{E_s}:$

$$\mathcal{A}_{E_1} = \mathcal{C}_{E_1} = \{a_1 - a_6\}, \quad \mathcal{B}_{E_1} = \emptyset, \quad \mathcal{A}_{E_2} = \emptyset, \quad \mathcal{B}_{E_2} = \{b_4 - b_7\},$$
$$\mathcal{A}_{E_3} = \emptyset, \quad \mathcal{B}_{E_3} = \{b_1 - b_3\}, \quad \mathcal{A}_{E_4} = \{a_7\}, \quad \mathcal{B}_{E_4} = \emptyset.$$

Therefore

$$|\mathcal{A}_{E_1}| = 6,\ |\mathcal{B}_{E_1}| = 0; \quad |\mathcal{A}_{E_2}| = 0,\ |\mathcal{B}_{E_2}| = 4,$$
$$|\mathcal{A}_{E_3}| = 0,\ |\mathcal{B}_{E_3}| = 3, \quad |\mathcal{A}_{E_4}| = 1,\ |\mathcal{B}_{E_4}| = 0.$$

According to (8)–(9), we have the following identification rule:

since $|\mathcal{A}_{E_1}| - |\mathcal{B}_{E_1}| = 6 > 0$, then $c \in \mathcal{C}_{E_1}$ is assigned to the set $\mathcal{A}$;
since $|\mathcal{A}_{E_2}| - |\mathcal{B}_{E_2}| = -4 < 0$, then $c \in \mathcal{C}_{E_2}$ is assigned to $\mathcal{B}$;
since $|\mathcal{A}_{E_3}| - |\mathcal{B}_{E_3}| = -3 < 0$, then $c \in \mathcal{C}_{E_3}$ is assigned to $\mathcal{B}$;
since $|\mathcal{A}_{E_4}| - |\mathcal{B}_{E_4}| = 1 > 0$, then $c \in \mathcal{C}_{E_4}$ is assigned to $\mathcal{A}$.

As a result of the identification, all points of the set $\mathcal{C}$ are correctly classified. Note that both "experts" f_1 and f_2 misclassified the point a_7 assigning it to the set $\mathcal{B}$, nevertheless, the new identification rule classified this point correctly. Of course, in general one can not expect such a perfect classification, we presented this example just to demonstrate the potential effectiveness of the VE-method.

Remark 3. Consider the following families of classifiers

$$\mathcal{F}_1 = \{f(x) = x_1 + \alpha \mid \alpha \in R = (-\infty, +\infty)\},$$
$$\mathcal{F}_2 = \{f(x) = x_2 + \beta \mid \beta \in R = (-\infty, +\infty)\}.$$

It is easy to check that for any functional $\varphi(f)$ described in (1)–(4), the classifier $f_1(x) = x_1$ is a minimizer of the problem (see (5)):

$$\varphi(f_1) = \min_{f \in \mathcal{F}_1} \varphi(f)$$

(the minimum is attained at $\alpha = 0$), and the classifier $f_2(x) = x_2$ is a minimizer of the problem:

$$\varphi(f_2) = \min_{f \in \mathcal{F}_2} \varphi(f)$$

(the minimum is attained at $\beta = 0$). Note also, that the original classifiers $f_1, \ldots, f_m$ are not supposed to be minimizers for some families of classifiers (of course, if they are, it is better).

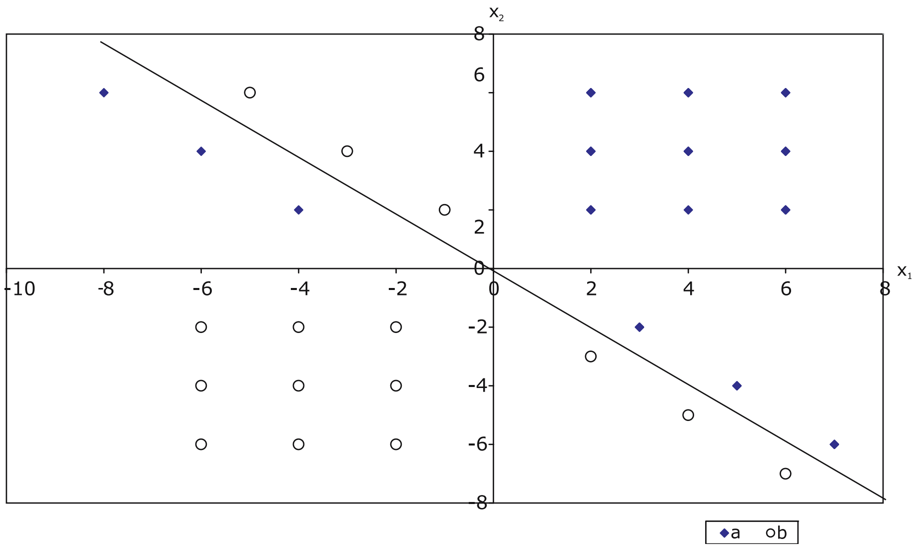

Figure 2.

4.2. EXAMPLE 2.

Let $x=(x_1,x_2)\in R^2$, $\mathcal{A}=\{a_i\in R^2 \mid i\in 1:N_1\}, \mathcal{B}=\{b_j\in R^2 \mid j\in 1:N_2\}$, where $N_1=15$, $a_1=(2,6)$, $a_2=(4,6)$, $a_3=(6,6)$, $a_4=(2,4)$, $a_5=(4,4), a_6=(6,4)$, $a_7=(2,2)$, $a_8=(4,2)$, $a_9=(6,2)$, $a_{10}=(-8,6), a_{11}=(-6,4)$, $a_{12}=(-4,2)$, $a_{13}=(3,-2)$, $a_{14}=(5,-4)$, $a_{15}=(7,-6)$; $N_2=15$, $b_1=(-6,-2)$, $b_2=(-4,-2)$, $b_3=(-2,-2)$, $b_4=(-6,-4), b_5=(-4,-4)$, $b_6=(-2,-4)$, $b_7=(-6,-6)$, $b_8=(-4,-6), b_9=(-2,-6)$, $b_{10}=(-5,6)$, $b_{11}=(-3,4)$, $b_{12}=(-1,2), b_{13}=(2,-3)$, $b_{14}=(4,-5)$, $b_{15}=(6,-7)$.
Put $\mathcal{C}=\mathcal{A}\cup\mathcal{B}$. Let the classifiers $f_1(x)=x_1$, $f_2=x_2$ and $f_3(x)=x_1+x_2$ be given with the following identification rules: a point $c=(c_1,c_2)\in\mathcal{C}$ is assigned by the classifier $f_s, s\in 1:3$, to the set $\mathcal{A}$ if $f_s(c)>0$, and to $\mathcal{B}$ – if $f_s(c)<0$.

It is clear (see Figure 2 and Section 3) that

$$\mathcal{C}_{1\mathcal{A}}=\{a_1-a_9, a_{13}-a_{15}, b_{13}-b_{15}\},\quad \mathcal{C}_{1\mathcal{B}}=\{b_1-b_9, b_{10}-b_{12}, a_{10}-a_{12}\},$$

$$\mathcal{A}_1^+=\mathcal{A}\cap\mathcal{C}_{1\mathcal{A}}=\{a_1-a_9, a_{13}-a_{15}\},\ \mathcal{B}_1^+=\mathcal{B}\cap\mathcal{C}_{1\mathcal{B}}=\{b_1-b_{12}\},$$

$$\mathcal{A}_1^-=\mathcal{A}\cap\mathcal{C}_{1\mathcal{B}}=\{a_{10}-a_{12}\},\ \mathcal{B}_1^-=\mathcal{B}\cap\mathcal{C}_{1\mathcal{A}}=\{b_{13}-b_{15}\}.$$

We conclude that the points $a_{10}-a_{12}$ and $b_{13}-b_{15}$ are misclassified by the classifier f_1 (altogether 6 misclassified points).

For the classifier f_2 we have

$$\mathcal{C}_{2\mathcal{A}}=\{a_1-a_{12},b_{10}-b_{12}\},\quad \mathcal{C}_{2\mathcal{B}}=\{b_1-b_9,b_{13}-b_{15},a_{13}-a_{15}\},$$
$$\mathcal{A}_2^+=\mathcal{A}\cap\mathcal{C}_{2\mathcal{A}}=\{a_1-a_{12},\},\quad \mathcal{B}_2^+=\mathcal{B}\cap\mathcal{C}_{2\mathcal{B}}=\{b_1-b_9,b_{13}-b_{15}\},$$
$$\mathcal{A}_2^-=\mathcal{A}\cap\mathcal{C}_{2\mathcal{B}}=\{a_{13}-a_{15}\},\quad \mathcal{B}_2^-=\mathcal{B}\cap\mathcal{C}_{2\mathcal{A}}=\{b_{10}-b_{12}\}.$$

We conclude that the points $a_{13}-a_{15}$ and $b_{10}-b_{12}$ are misclassified by the classifier f_2 (altogether 6 misclassified points).

Analogously, for the classifier f_3 we have

$$\mathcal{C}_{3\mathcal{A}}=\{a_1-a_9,a_{13}-a_{15},b_{10}-b_{12}\},\quad \mathcal{C}_{3\mathcal{B}}=\{b_1-b_9,b_{13}-b_{15},a_{10}-a_{12}\},$$
$$\mathcal{A}_3^+=\mathcal{A}\cap\mathcal{C}_{3\mathcal{A}}=\{a_1-a_9,a_{13}-a_{15},\},\ \mathcal{B}_3^+=\mathcal{B}\cap\mathcal{C}_{3\mathcal{B}}=\{b_1-b_9,b_{13}-b_{15}\},$$
$$\mathcal{A}_3^-=\mathcal{A}\cap\mathcal{C}_{3\mathcal{B}}=\{a_{10}-a_{12}\},\ \mathcal{B}_3^-=\mathcal{B}\cap\mathcal{C}_{3\mathcal{A}}=\{b_{10}-b_{12}\}.$$

We conclude that the points $a_{10}-a_{12}$ and $b_{10}-b_{12}$ are misclassified by the classifier f_3 (altogether 6 misclassified points). Thus, every of the three classifiers produces 6 misclassified points.

The set $\mathcal{E}$ contains 8 points: $\mathcal{E}=\{E_i \mid i\in 1:8\}$ where $E_1=(1,1,1)$, $E_2=(1,1,2)$, $E_3=(1,2,1)$, $E_4=(1,2,2)$, $E_5=(2,1,1)$, $E_6=(2,1,2)$, $E_7=(2,2,1)$, $E_8=(2,2,2)$. We have (see (6))

$$\mathcal{C}_{E_1}=\{c\in\mathcal{C}\mid e(c)=E_1\}=\{a_1-a_9\},$$
$$\mathcal{C}_{E_2}=\{c\in\mathcal{C}\mid e(c)=E_2\}=\emptyset,$$
$$\mathcal{C}_{E_3}=\{c\in\mathcal{C}\mid e(c)=E_3\}=\{a_{13}-a_{15}\},$$
$$\mathcal{C}_{E_4}=\{c\in\mathcal{C}\mid e(c)=E_4\}=\{b_{13}-b_{15}\},$$
$$\mathcal{C}_{E_5}=\{c\in\mathcal{C}\mid e(c)=E_5\}=\{a_1-a_9\},$$
$$\mathcal{C}_{E_6}=\{c\in\mathcal{C}\mid e(c)=E_6\}=\{a_{10}-a_{12}\},$$
$$\mathcal{C}_{E_7}=\{c\in\mathcal{C}\mid e(c)=E_7\}=\emptyset,$$
$$\mathcal{C}_{E_8}=\{c\in\mathcal{C}\mid e(c)=E_8\}=\{b_1-b_9\}.$$

Note that the sets $\mathcal{C}_{E_2}$ and $\mathcal{C}_{E_7}$ are empty (therefore they are not considered in the sequel). Now let us construct the sets (see (7)) $\mathcal{A}_{E_s}=\mathcal{A}\cap\mathcal{C}_{E_s}$, $\mathcal{B}_{E_s}=\mathcal{B}\cap\mathcal{C}_{E_s}$:

$$\mathcal{A}_{E_1}=\mathcal{C}_{E_1}=\{a_1-a_9\},\quad \mathcal{B}_{E_1}=\emptyset,\quad \mathcal{A}_{E_3}=\mathcal{C}_{E_3}=\{a_{13}-a_{15}\},$$
$$\mathcal{B}_{E_3}=\emptyset,\mathcal{A}_{E_4}=\emptyset,\quad \mathcal{B}_{E_4}=\mathcal{C}_{E_4}=\{b_{13}-b_{15}\},\quad \mathcal{A}_{E_5}=\emptyset,$$
$$\mathcal{B}_{E_5}=\mathcal{C}_{E_5}=\{b_{10}-b_{12}\},\quad \mathcal{A}_{E_6}=\mathcal{C}_{E_6}=\{a_{10}-a_{12}\},$$
$$\mathcal{B}_{E_6}=\emptyset,\quad \mathcal{A}_{E_8}=\emptyset,\quad \mathcal{B}_{E_8}=\mathcal{C}_{E_8}=\{b_1-b_9\}.$$

Therefore

$$|\mathcal{A}_{E_1}|=9,\ |\mathcal{B}_{E_1}|=0;\quad |\mathcal{A}_{E_3}|=3,\ |\mathcal{B}_{E_3}|=0;\quad |\mathcal{A}_{E_4}|=0,\ |\mathcal{B}_{E_4}|=3;$$
$$|\mathcal{A}_{E_5}|=0,\ |\mathcal{B}_{E_5}|=3;\quad |\mathcal{A}_{E_6}|=3,\ |\mathcal{B}_{E_6}|=0;\quad |\mathcal{A}_{E_8}|=0,\ |\mathcal{B}_{E_8}|=9.$$

According to (8)–(9), we have the following identification rule:

$$
\begin{aligned}
&\text{since } |\mathcal{A}_{E_1}|-|\mathcal{B}_{E_1}|=9>0, \text{ then } c\in\mathcal{C}_{E_1} \text{ is assigned to the set } \mathcal{A};\\
&\text{since } |\mathcal{A}_{E_3}|-|\mathcal{B}_{E_3}|=3>0, \text{ then } c\in\mathcal{C}_{E_3} \text{ is assigned to } \mathcal{A};\\
&\text{since } |\mathcal{A}_{E_4}|-|\mathcal{B}_{E_4}|=-3<0, \text{ then } c\in\mathcal{C}_{E_4} \text{ is assigned to } \mathcal{B},\\
&\text{since } |\mathcal{A}_{E_5}|-|\mathcal{B}_{E_5}|=-3<0, \text{ then } c\in\mathcal{C}_{E_5} \text{ is assigned to the set } \mathcal{B};\\
&\text{since } |\mathcal{A}_{E_6}|-|\mathcal{B}_{E_6}|=3>0, \text{ then } c\in\mathcal{C}_{E_6} \text{ is assigned to } \mathcal{A};\\
&\text{since } |\mathcal{A}_{E_8}|-|\mathcal{B}_{E_8}|=-9<0, \text{ then } c\in\mathcal{C}_{E_8} \text{ is assigned to } \mathcal{B}.
\end{aligned}
$$

As a result of the identification, all points of the set $\mathcal{C}$ are correctly classified. Of course, in general one can not expect such a perfect classification, we presented this example again just to demonstrate the potential effectiveness of the VE-method.

Remark 4. Consider the following families of classifiers

$$
\begin{aligned}
&\mathcal{F}_1=\{f(x)=x_1+\alpha \mid \alpha\in R=(-\infty,+\infty)\},\\
&\mathcal{F}_2=\{f(x)=x_2+\beta \mid \beta\in R=(-\infty,+\infty)\},\\
&\mathcal{F}_3=\{f(x)=x_1+x_2+\gamma \mid \gamma\in R=(-\infty,+\infty)\}.
\end{aligned}
$$

It is not difficult to check that for any functional $\varphi(f)$ described in (1)–(4), the classifier $f_1(x)=x_1$ is a minimizer of the problem (see (5)):

$$\varphi(f_1)=\min_{f\in\mathcal{F}_1}\varphi(f)$$

(the minimum is attained at $\alpha=0$), the classifier $f_2(x)=x_2$ is a minimizer of the problem:

$$\varphi(f_2)=\min_{f\in\mathcal{F}_2}\varphi(f)$$

(the minimum is attained at $\beta=0$), and the classifier $f_3(x)=x_1+x_2$ is a minimizer of the problem:

$$\varphi(f_3)=\min_{f\in\mathcal{F}_3}\varphi(f)$$

(the minimum is attained at $\gamma=0$).

Remark 5. If one takes only classifiers f_1 and f_2 for the sets $\mathcal{A}$ and $\mathcal{B}$ described in Example 2, and performs all calculations as in Example 1, the following identification rule will be obtained for a point $c=(c_1,c_2)\in\mathcal{C}$:

if $c_1 > 0, c_2 > 0$ then c is assigned to the set $\mathcal{A}$;
if $c_1 > 0, c_2 < 0$ then c is assigned to the set $\mathcal{A}$;
if $c_1 < 0, c_2 > 0$ then c is assigned to the set $\mathcal{A}$;
if $c_1 < 0, c_2 < 0$ then c is assigned to the set $\mathcal{B}$.

This new classifier produces 6 misclassified points – the same amount as the one produced by every of the classifiers f_1 and f_2. However, if a point $c = (c_1, c_2)$ belogs to the first or the third quadrant of the plane (i.e., if $c_1 c_2 > 0$) then the point c is correctly identified. In this respect the new classifier is more informative.

Remark 6. The VE-method is a supervised classification method. In the case of unsupervised classification (when one is unable to check the quality of an identification rule), if several IRs are available with some characterization of each classifier (e.g., it is known that the "expert" f_s says that the probability that a point $c \in \mathcal{A} \cup \mathcal{B}$ belongs to the set $\mathcal{A}$ is $p_{s\mathcal{A}}(c)$, and the probability that c belongs to $\mathcal{B}$ is $p_{s\mathcal{B}}(c)$), then it is possible to derive a new identification rule based on the probabilistic arguments.

5. A Generalized Method of Virtual Experts

Assume that we have several families of classifiers

$$\mathcal{F}_1 = \{f_{1\alpha_1}(x) = f_1(x, \alpha_1) \mid \alpha_1 \in \Omega_1 \subset R^{n_1}\}, \dots ,$$
$$\mathcal{F}_m = \{f_{m\alpha_m}(x) = f_m(x, \alpha_m) \mid \alpha_m \in \Omega_m \subset R^{n_m}\}.$$

Let

$$\mathcal{F} = \{(f_{1\alpha_1}, \dots , f_{m\alpha_m}) \mid f_{i\alpha_i} \in \mathcal{F}_i \ \forall i \in 1:m\}.$$

Any collection of m classifiers of the type $F(\alpha) = F(\alpha_1, \dots , \alpha_m) = (f_{1\alpha_1}, \dots , f_{m\alpha_m}) \in \mathcal{F}$ (where $(\alpha) = (\alpha_1, \dots , \alpha_m) \in \Omega = (\Omega_1 \times \cdots \times \Omega_m)$) will be called an m-dimensional classifier. Using the VE-method (applied to the collection of classifiers $F(\alpha)$), one gets a new identification rule described in Section 3 (see (8)–(9)).

Denote by $\mathcal{A}^-(F(\alpha)) \subset \mathcal{A}$ the set of points of $\mathcal{A}$ which are wrongly identified by the classifier $F(\alpha)$ and by $\mathcal{B}^-(F(\alpha)) \subset \mathcal{B}$ – the set of points of $\mathcal{B}$ which are wrongly identified by the classifier $F(\alpha)$. Let us choose one of the functionals $\varphi_i(F(\alpha)), i \in 1:4$, described in (1)–(4). Denote it by $\varphi(\alpha)$.

Now it is possible to state the problem of finding a minimizer of the α^* of the functional φ:

$$\varphi(\alpha^*) = \min_{\alpha \in \Omega} \varphi(\alpha). \tag{10}$$

To illustrate the idea, consider again Example 1 discussed in subsection 4.1. Let two families of classifiers

$$\mathcal{F}_1 = \Big\{ f_{1\alpha_1}(x) = \alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13} | \alpha_{11} \in R,\ \alpha_{12} \in R,\\ \alpha_{13} \in R,\ \alpha_{11}^2 + \alpha_{12}^2 = 1 \Big\}$$

and

$$\mathcal{F}_2 = \Big\{ f_{2\alpha_2}(x) = \alpha_{21}x_1 + \alpha_{22}x_2 + \alpha_{23} | \alpha_{21} \in R,\ \alpha_{22} \in R,\ \alpha_{23} \in R,\\ \alpha_{21}^2 + \alpha_{22}^2 = 1 \Big\}$$

be given with the following identification rules:

if $f_1(x, \alpha_1) \in \mathcal{F}_1$ and $f_2(x, \alpha_2) \in \mathcal{F}_2$ then a point $c = (c_1, c_2) \in \mathcal{C}$ is assigned by the classifier $f_s, s \in 1:2$, to the set $\mathcal{A}$ if $f_s(c, \alpha_s) > 0$, and to $\mathcal{B}$ – if $f_s(c, \alpha_s) < 0$.

Note that the families $\mathcal{F}_1$ and $\mathcal{F}_2$ coincide, i.e. we can choose a pair of classifiers from the same family of classifiers. It follows from the results of Example 1 that the point $\alpha^* = (\alpha_1^*, \alpha_2^*)$ where $\alpha_1^* = (1, 0, 0),\ \alpha_2^* = (0, 1, 0)$ is a minimizer of the functional φ since $\varphi(\alpha^*) = 0$ (and the functional φ is nonnegative). Observe also that the point $\alpha^{**} = (\alpha_1^{**}, \alpha_2^{**})$ where $\alpha_1^{**} = (0, 1, 0),\ \alpha_2^{**} = (1, 0, 0)$ is also a minimizer of the functional φ since again $\varphi(\alpha^{**}) = 0$.

Remark 7. The functional $\varphi(\alpha)$ defined by (10) is discontinuous (it is also multiextremal) therefore the stated problem is a difficult and challenging one from both theoretical and practical points of view. Like in [8] (see Section 3 there), this discrete optimization problem can be approximated (to within any given accuracy) by a continuous (though still multiextremal) one.

6. One-dimensional Identification by a Separation Technique

We use the following separation algorithm to identify points of two one-dimensional finite sets of disjoint points. Let sets $A \subset \mathbf{R}$ and $B \subset \mathbf{R}$ be given:

$$A = \{a_i \in \mathbf{R} \mid i \in I\}, \quad B = \{b_j \in \mathbf{R} \mid j \in J\},$$

where $I = 1 : N_1,\ J = 1 : N_2$. Assume that a_i and b_j are ordered:

$$a_1 < a_2 < \cdots < a_{N_1}, b_1 < b_2 < \cdots < b_{N_2}$$

and that $a_i \neq b_j \quad \forall i \in I, j \in J$ (i.e. $A \cap B = \emptyset$).

We shall use the following identification procedure: Assume that a function $F(x,c):\mathbf{R}^2\to\mathbf{R}$ is given. The function $F(x,c)$ will be referred to as an identifier (or classifier). Take any $c\in A\cup B$. If $F(x,c)>0$, the point c is identified as a point of the set A. If $F(x,c)<0$, the point c is identified as a point of the set B. In the case $F(x,c)=0$ the point c is considered as unidentifiable by the identifier $F(x,c)$. In what follows we consider the case $F(x,c)=c-x$. (Sometimes we shall use the function $F(x,c)=x-c$.)

Let $m_1(x)=|A^-|$ where $A^-=\{a_i\in A \mid a_i\leqslant x\}, m_2(x)=|B^-|$ where $B^-=\{b_j\in B \mid b_j\geqslant x\}$. Here $|C|$ is the number of points in a set C. Thus, $m_1(x)$ ($m_2(x)$) represents the number of points of the set A (respectively, B) incorrectly identified (misclassified) by the identifier $F(x,c)$. Now, as a performance (criterion) function, let us take the function

$$m(x)=\max\{m_1(x),m_2(x)\}. \tag{11}$$

The problem is to find

$$\min_{x\in\mathbf{R}} m(x)=m^*.$$

The functions $m(x),m_1(x),m_2(x)$ take only integer values and are discontinuous and piecewise constant. The set A is the set of discontinuity points of the function m_1 and the set B is the set of discontinuity points of the function m_2. The function $m_1(x)$ is nondecreasing and, hence, quasiconvex, while the function $m_1(x)$ is nonincreasing and also quasiconvex (see Zabotin et al., 1972). The function $m(x)$, as the maximum of quasiconvex functions, is also quasiconvex (see Zabotin et al., 1972). Of course, one may choose another function as a criterion function. For example, the function $f_1(x)=m_1(x)+m_2(x)$ is also interesting in this respect. However, the set of (global) minimizers of the function m is convex and there exist no local minimizers while the function f_1 may have local minimizers which are not global ones and even the set of global minimizers may happen to be nonconvex.

6.1. NECESSARY AND SUFFICIENT OPTIMALITY CONDITIONS

Let M^* be the set of minimizers of the function m defined by (11). The function m is quasiconvex, therefore M^* is convex (see Zabotin et al., 1972). Since we consider the one-dimensional case, M^* has one of the forms: $M^*=[p,q]$, $M^*=(p,q]$, $M^*=[p,q)$, $M^*=(p,q)$. The function m is constant on $\mathbf{R}\setminus A\cup B$, hence the points p and q belong to $A\cup B$.

By $m_1(x^+)$ let us denote the value $m_1(z)$ for $z\in(x,x')$ where x' is such that $(x,x')\cap[A\cup B]=\emptyset$ (i.e., in the interval (x,x') there are no points of the sets A and B). By $m_1(x^-)$ let us denote the value $m_1(z)$ for $z\in(x',x)$

where x' is such that $(x', x) \cap [A \cup B] = \emptyset$ (i.e., in the interval (x', x) there are no points of the sets A and B).

The functions $m_2(x^+)$ and $m_2(x^-)$ are introduced in the same way.

Put

$$m(x^+) = \max\{m_1(x^+), m_2(x^+)\}, \;\; m(x^-) = \max\{m_1(x^-), m_2(x^-)\}.$$

Clearly,

$$m_1(x^+) \geqslant m_1(x), \quad m_2(x^+) \leqslant m_2(x),$$
$$m_1(x^-) \leqslant m_1(x), \quad m_2(x^-) \geqslant m_2(x).$$

If $a_i \in A, b_j \in B$ then

$$m_1(a_i^+) = m_1(a_i), \;\; m_1(a_i^-) = m_1(a_i) - 1,$$
$$m_1(b_j^+) = m_1(b_j^-) = m_1(b_j), \;\; m_2(a_i^+) = m_2(a_i^-) = m_2(a_i),$$
$$m_2(b_j^+) = m_2(b_j) - 1, \;\; m_2(b_j^-) = m_2(b_j).$$

It is easy to see that

$$m(a_i^-) \leqslant m(a_i) = m(a_i^+), \quad m(b_j^+) \leqslant m(b_j) = m(b_j^-).$$

Now let us show that M^* is of the form $M^* = (p, q)$. Assuming the contrary, one has:

In the case $p \in M^*$: if $p = b_j \in B$ then $m_1(b_j^-) = m_1(b_j)$, $m_2(b_j^-) = m_2(b_j)$, and the interval M^* can be enlarged (to the left); if $p = a_i \in A$ then

$$m_1(a_i^-) = m_1(a_i) - 1, \quad m_2(a_i^-) = m_2(a_i),$$

and the interval M^* can be enlarged (to the left).

In the case $q \in M^*$: if $p = b_j \in B$ then

$$m_1(b_j^+) = m_1(b_j), \quad m_2(b_j^+) = m_2(b_j) - 1,$$

and the interval M^* can be enlarged (to the right); if $p = a_i \in A$ then

$$m_1(a_i^+) = m_1(a_i), \quad m_2(a_i^+) = m_2(a_i),$$

and the interval M^* can be enlarged (to the right).

Thus, $M^* = (p, q)$. Next, let us prove that $p \in B, q \in A$. First we show that $p \in B$. Assume the contrary, let $p = a_i$. Then

$$m_1(a_i) = m_1(a_i^+), \;\; m_1(a_i^-) = m_1(a_i) - 1 = m_1(a_i^+) - 1,$$
$$m_2(a_i^-) = m_2(a_i^+) = m_2(a_i),$$

and therefore $m(a_i^-) \leqslant m(a_i^+)$ (i.e., points close to a_i from the left belong to M^*). This is a contradiction, hence, $p \in B$.

Now show that $q \in A$. Assume the contrary, let $q = b_j$. Then

$$m_2(b_j^+) = m_2(b_j) - 1 = m_2(b_j^-) - 1, \;\; m_1(b_j^+) = m_1(b_j^-) = m_1(b_j),$$

and therefore $m(b_j^+) \leqslant m(b_j^-)$ (i.e., points close to b_j from the right belong to M^*). This is a contradiction, hence, $q \in A$.

Thus, we have proved that $M^* = (b_j, a_i)$. Now we shall demonstrate that $m_2(b_j^+) = m_1(a_i^-) = m^*$. Indeed, assume, e.g., that $m_2(b_j^+) < m^*$. Since

$$\begin{aligned} &m_2(b_j^-) = m_2(b_j) = m_2(b_j^+) + 1 \leqslant m^*, \\ &m_1(b_j^-) = m_1(b_j^+) \leqslant m^*, \end{aligned}$$

i.e. $m(b_j^-) \leqslant m^*$, and since $m(z) \geqslant m^* \quad \forall z$, then $m(b_j^-) = m^*$. Thus, the set M^* can be enlarged (to the left) which is impossible.

Similarly, assuming that $m_1(a_i^-) < m^*$, one gets

$$m_1(a_i^+) = m_1(a_i^-) + 1 \leqslant m^*, \;\; m_2(a_i^+) = m_2(a_i^-) \leqslant m^*,$$

i.e. $m(a_i^+) \leqslant m^*$, and since $m(z) \geqslant m^* \quad \forall z$, then $m(a_i^+) = m^*$. Thus, the set M^* can be enlarged (to the right) which is again impossible.

Hence, we have just proved the following

THEOREM 1. *The set M^* is of the form $M^* = (b_j, a_i)$ and, moreover,*

$$m_2(b_j^+) = m_1(a_i^-) = m^*.$$

6.2. A NUMERICAL METHOD FOR MINIMIZING $m(x)$

Using the above Theorem, it is possible to describe the following numerical procedure for finding the set M^*.

Choose $M_0 = (p_0, q_0)$ where

$$p_0 \leqslant \min\left\{\min_{i \in I} a_i, \;\; \min_{j \in J} b_j\right\}, \quad q_0 > \max\left\{\max_{i \in I} a_i, \max_{j \in J} b_j\right\}.$$

Then

$$m(p_0) = m_2(p_0) = N_1, \quad m(q_0) = m_1(q_0) = N_2, \quad m_1(p_0) = m_2(q_0) = 0.$$

Let $M_k = (p_k, q_k)$ be found such that $m_1(p_k) < m_2(p_k), \;\; m_2(q_k) < m_1(q_k)$. Take $c_k = \frac{p_k + q_k}{2}$.

If $m_1(c_k) = m_2(c_k)$ then $m(c_k) = m^*$ (i.e., c_k is a minimizer since m_2 is a decreasing function and m_1 is an increasing function). Now find a point $b_{j_k} \in B$ nearest to c_k from the left and a point $a_{i_k} \in A$ nearest to c_k from the right. Then $M^* = (b_{j_k}, a_{i_k})$.

If $m_1(c_k) < m_2(c_k)$ then put $M_{k+1} = (p_{k+1}, q_{k+1})$ where $p_{k+1} = c_k, \; q_{k+1} = q_k$. Note that

$$m_1(p_{k+1}) = m_1(c_k) < m_2(c_k) = m_2(p_{k+1}),$$
$$m_2(q_{k+1}) = m_2(q_k) < m_1(q_k) = m_1(q_{k+1}).$$

Finally, if $m_1(c_k) > m_2(c_k)$ then put $M_{k+1} = (p_{k+1}, q_{k+1})$ where $p_{k+1} = p_k, \; q_{k+1} = c_k$. Note again that

$$m_1(p_{k+1}) = m_1(p_k) < m_2(p_k) = m_2(p_{k+1}),$$
$$m_2(q_{k+1}) = m_2(c_k) < m_1(c_k) = m_1(q_{k+1}).$$

Continuing in the same way, we construct a sequence of intervals $\{M_k\}$. It is not difficult to show that the process terminates in a finite number of steps producing the set M^* we are looking for. The mentioned number of steps does not exceed the value K where K is such that

$$\frac{q_0 - p_0}{2^K} < d = \min_{i \in I, j \in J} |a_i - b_j|.$$

6.3. THE RANKING OF PARAMETERS VIA THE ID-IDENTIFICATION

Let sets $A \subset \mathbf{R}^n$ and $B \subset \mathbf{R}^n$ be given: $A = \{a_i \in \mathbf{R}^n \mid i \in I\}, \; B = \{b_j \in \mathbf{R}^n \mid j \in J\}$, where

$$I = 1 : N_1, \;\; J = 1 : N_2, \;\; a_i = (a_{i1}, \ldots, a_{in}), \;\; b_j = (b_{j1}, \ldots, b_{jn}).$$

For simplicity assume that $N_1 = N_2 = N$. For each coordinate k we have the one-dimensional sets

$$A_k = \{a_{ik} \mid i \in I\}, \;\; B_k = \{b_{jk} \mid j \in J\}.$$

Applying the above algorithm, we perform the identification procedure for the sets A_k and B_k and find the value m_k^*. The value $\mu_k = \frac{m_k^*}{N}$ may be viewed as the ratio of misclassified points of the sets A_k and B_k. Rearranging the values μ_k's in the increasing order, we can rank them according to their informative values. The described ranking is not based on statistical considerations (like the ranking procedure proposed in Kokorina, 2002a).

Remark 8. We used the described ranking procedure to perform n-dimensional identification by the separation technique. Many important conclusions were obtained even in a very simple case $n=2$. Note that 2 parameters with small informative values may have a high cumulative (joint) informative value (see examples in Section 4).

Remark 9. The case $N_1 \neq N_2$ can be treated in a similar way. The case of “weighted” a_i’s and b_j’s is of practical value and can be handled by a modified separation technique.

References

1. *Advances in Kernel Methods. Support Vector Learning* (1999). In: Schoelkopf, B., Burges, C.J.C. and Smola, A.J. (eds.). The MIT Press, Cambridge, Mass.; London, England.
2. Astorino, A. and Gaudioso, M. (2002), Polyhedral separability through Successive LP. *Journal of Optimization Theory and Applications* 112 (4), 265–293.
3. Bagirov, A.M., Rubinov, A.M., Soukhoroukova, N.V. and Yerwood, J. (2003), Unsupervised and supervised data classification via nonsmooth and global optimization. *Topology* 11(1), 1–93.
4. Bagirov, A.M. (2005), Max-min separability. *Optimization Methods and Software (OMS)* 20(2–3), 275–294.
5. Bennett, K.P. and Mangasarian, O.L. (1992), Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software* 1(1), 22–34.
6. Cristianini, N. and Shawe-Taylor, J. (2000), *An Introduction to Support Vector Machines and Other Kernel Based Methods*. Cambridge University Press.
7. Cucker, F. and Smale, S. (2001), On the mathematical foundations of learning. *Bulletin (New Series) of the American Mathematical Society* 39(1), 1–49.
8. Demyanov, V.F. (2005), Mathematical Diagnostics via Nonsmooth Analysis. *Optimization Methods and Software (OMS)* 20(2–3), 197–218.
9. Demyanova, V.V. (2004), Psychodiagnostics of reliability of competition activity in sports by means of mathematical modelling. *Longevity, Aging and Degradation Models. Vol. 2* (Proceedings of the International Conference LAD’2004). The St.-Petersburg Polytechnical University Press, pp. 96–100; St.-Petersburg, Russia.
10. Kokorina, A.V. (2002a), Ranking the parameters in the mathematical diagnostics problems. Comments to the paper (Bagirov et al., 2003), pp. 86–89.
11. Kokorina, A.V. (2002b), Ranking the parameters in the data mining problems (in Russian). *Control Processes and Stability. Proceedings of XXXIII Scientific Conference of students and Ph.D. students of Applied Mathematics Dept. of St.-Petersburg State University. 15–18 April, 2002*. The St.-Petersburg University Press, pp. 277–281; St. Petersburg, Russia.
12. Kokorina, A.V. (2004), Ranking the Parameters in Classification Databases. *Longevity, Aging and Degradation Models. Vol. 2* (Proceedings of the International Conference LAD’2004). The St.Petersburg Polytechnical University Press, pp. 191–193; St.Petersburg, Russia.
13. Lee, Yuh-Jye. and Mangasarian, O.L. (2001), SSVM: A smooth support vector machine for classification. *Computational Optimization and Applications* 20(1), 5–22.
14. Mangasarian, O.L. (1994), Misclassification minimization. *Journal of Global Optimization* 5, 309–323.

15. Petrova, N.V. (2004), Separation of two discrete one-dimensional sets by the isolation method (in Russian). *Control Processes and Stability. Proceedings of XXXV Scientific Conference of students and Ph.D. students of Applied Mathematics Dept. of St.Petersburg State University. 14–16 April, 2004*. The St.Petersburg University Press, pp. 328–330; St.Petersburg, Russia.
16. Vapnik, V. (2000), *The Nature of Statistical Learning Theory.* Springer-Verlag, New York.
17. Varis, Ya.V. (2004), One-dimensional identification of two discrete sets by means of two intervals (in Russian). *Control Processes and Stability. Proceedings of XXXV Scientific Conference of students and Ph.D. students of Applied Mathematics Dept. of St.Petersburg State University. 140-16 April, 2004*. The St.Petersburg University Press, pp. 291–293; St.Petersburg, Russia.
18. Zabotin, Ya.I., Korablev, A.I. and Khabibulin, R.F. (1972), On minimization of quasiconvex functionals. *Izvestiya VUZov. Matematika*, No. 10, 27–33.