Journal of Global Optimization **12:** 111–126, 1998.

Pattern Classification by Linear Goal Programming and its Extensions

HIROTAKA NAKAYAMA and NAOKO KAGAKU
Department of Applied Mathematics, Konan University, 8-9-1 Okamoto, Higashinada, Kobe 658, Japan (E-mail: nakayama@edu2.math.konan-u.ac.jp)

(Received and accepted: 2 May 1997)

Abstract. Pattern classification is one of the main themes in pattern recognition, and has been tackled by several methods such as the statistic one, artificial neural networks, mathematical programming and so on. Among them, the multi-surface method proposed by Mangasarian is very attractive, because it can provide an exact discrimination function even for highly nonlinear problems without any assumption on the data distribution. However, the method often causes many slits on the discrimination curve. In other words, the piecewise linear discrimination curve is sometimes too complex resulting in a poor generalization ability. In this paper, several trials in order to overcome the difficulties of the multi-surface method are suggested. One of them is the utilization of goal programming in which the auxiliary linear programming problem is formulated as a goal programming in order to get as simple discrimination curves as possible. Another one is to apply fuzzy programming by which we can get fuzzy discrimination curves with gray zones. In addition, it will be shown that using the suggested methods, the additional learning can be easily made. These features of the methods make the discrimination more realistic. The effectiveness of the methods is shown on the basis of some applications.

Key words: Multiobjective programming, pattern classification, goal programming, portfolio selection.

1. Introduction

Machine learning for pattern classification is one of the most important subjects in artificial intelligence. Several approaches to this topic have been developed so far: in particular, statistic discrimination analysis, artificial neural networks (ANN), and the mathematical programming approach are popular in many real fields.

In ANN, the back propagation method is widely applied for learning. However, unless several parameters are decided upon moderately in advance, it tends to be trapped in local minima. Moreover, it is a problem how the structure of ANN should be decided upon [2, 3].

Although statistic discrimination analysis has been developed remarkably, it requires some assumption on data distribution. This assumption restricts the application.

On the other hand, the mathematical programming approach has been attracting many researcher's attention due to its applicability to a wide range of real problems and the simplicity of its algorithm. In particular, the Multisurface Method (MSM) is very attractive because it is very simple and can provide an exact discrimination function even for highly nonlinear problems without any assumptions on the data distribution.

The method finds a piecewise linear discrimination function by solving linear programming iteratively. One of the most prominent features of MSM is that (1) there are no parameters to be decided in advance, (2) relatively short computing time is required, and (3) the resulting solution is guaranteed to be the global optimal. Wolbeg–Mangasarian [10] reported its good performance in application to medical diagnosis problems. In addition, MSM can be interpreted to be a method for constructing cascaded neural networks [6,9].

However, we observed that the method tends to produce too complicated discrimination surfaces, which cause a poor generalization ability because it is affected by noise. In order to overcome this difficulty, we proposed a multi-objective optimization approach [8]. On the other hand, Benett–Mangasarian proposed a method called Robust Linear Programming Discrimination (RPLD) whose idea is based on the goal programming. Although these new methods provide simpler discrimination functions than by the original MSM, they still lead to too complicated ones in many cases. This is because they try to find an exact discrimination surface which classifies all data correctly. It has been observed that such perfect learning gives a poor generalization ability in many cases. This phenomenon is known as over-learning.

In order to avoid over-learning, in this paper, we suggest methods which do not require the perfect classification for given data, but allow unclassified data. This idea can be performed by applying fuzzy mathematical programming techniques which produce “gray zones”. The effectiveness of the modified methods can be proved by several applications to artificial classification problems. Next, these methods are extended to trinary classification problems. Finally, we apply them to a stock portfolio problem.

2. Multisurface Method (MSM)

Suppose that given data in a set X of an n-dimensional Euclidean space belong to one of two categories $\mathcal{A}$ and $\mathcal{B}$. Let A be a matrix whose row vectors denote points of the category $\mathcal{A}$. Similarly, let B be a matrix whose row vectors denote points of the category $\mathcal{B}$. For simplicity of notation, we denote the set of points of $\mathcal{A}$ by A. Similarly the set of points of $\mathcal{B}$ is denoted by B. MSM suggested by Mangasarian [5] finds a piecewise linear discrimination surface separating two sets A and B by solving linear programming problems iteratively. The main idea is to find two hyperplanes parallel to each other which classify as many given data as possible:

$$g(\mathbf{u}) = \mathbf{x}^T\mathbf{u} = \alpha$$

$$g(\mathbf{u}) = \mathbf{x}^T\mathbf{u} = \beta$$

This is performed by the following algorithm:

Step 1. Solve the following linear programming problem at the k-th iteration (set $k = 1$ at the beginning):

$$
\begin{aligned}
\text{(I)}\quad & \text{Maximize} && \phi(A,B) = \alpha - \beta \\
& \text{subject to} && A\mathbf{u} \geqq \alpha\mathbf{1} \\
& && B\mathbf{u} \leqq \beta\mathbf{1} \\
& && -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1} \\
& \text{where} && \mathbf{1}^T = (1,\ldots,1).
\end{aligned}
$$

In order to avoid a trivial solution $\mathbf{u} = 0, \alpha = 0, \beta = 0$, we add a constraint which is a linear approximation of $\mathbf{u}^T\mathbf{u} \geqq \frac{1}{2}$. Namely,

$$\mathbf{u}^T\mathbf{u} \cong \mathbf{p}^T\mathbf{p} + 2\mathbf{p}^T(\mathbf{u} - \mathbf{p}) \geqq \frac{1}{2}$$

which finally yields

$$\mathbf{p}^T\mathbf{u} \geqq \frac{1}{2}\left(\frac{1}{2} + \mathbf{p}^T\mathbf{p}\right).$$

Now our problem to be solved is given by

$$
\begin{aligned}
\text{(I}'\text{)}\quad & \text{Maximize} && \phi_i(A,B) = \alpha - \beta \\
& \text{subject to} && A\mathbf{u} \geqq \alpha\mathbf{1} \\
& && B\mathbf{u} \leqq \beta\mathbf{1} \\
& && -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1} \\
& && \mathbf{p}_i^T\mathbf{u} \geqq \tfrac{1}{2}\left(\tfrac{1}{2} + \mathbf{p}_i^T\mathbf{p}_i\right) \qquad (1)
\end{aligned}
$$

where $\mathbf{p}_i$ is given by one of

$$
\begin{aligned}
\mathbf{p}_i^T &= \left(\frac{1}{\sqrt{2}}, 0, \ldots, 0\right) \\
\mathbf{p}_2^T &= \left(-\frac{1}{\sqrt{2}}, 0, \ldots, 0\right) \\
&\ \ \vdots \\
\mathbf{p}_{2n}^T &= \left(0, \ldots, 0, -\frac{1}{\sqrt{2}}\right).
\end{aligned}
$$

After solving the LP problem (I′) for each i such that $1 \leqq i \leqq 2n$, we take a hyperplane which classifies correctly as many given data as possible. Let the solution be $\mathbf{u}^*, \alpha^*, \beta^*$, and let the corresponding value of the objective function be $\phi^*(A,B)$.

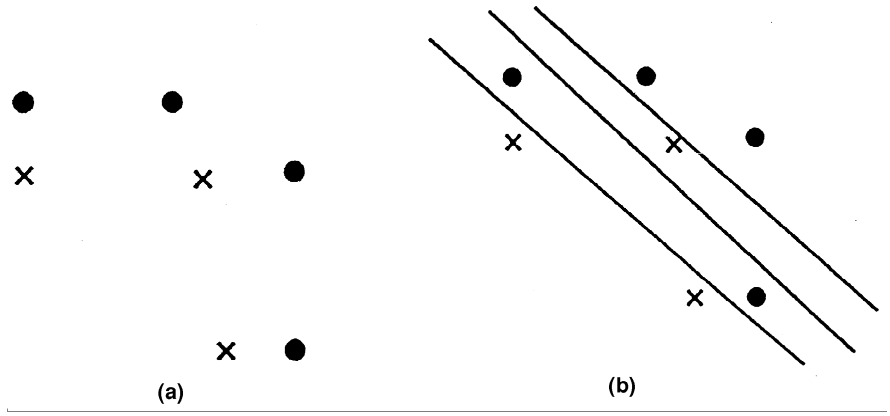

Figure 1. (a) Data. (b) Classification by MSM.

If $\phi^*(A, B) > 0$, then we have a complete separating hyperplane $g(\mathbf{u}^*) = (\alpha^* + \beta^*)/2$. Set $\tilde{A}^k = \{\mathbf{x} \in X | g(\mathbf{u}^*) \geqq (\alpha^* + \beta^*)/2\}$ and $\tilde{B}^k = \{\mathbf{x} \in X | g(\mathbf{u}^*) < (\alpha^* + \beta^*)/2\}$. $\tilde{A}^k$ and $\tilde{B}^k$ include the sets A and B in X, respectively, which is decided at this stage. Go to Step 3.

Otherwise, go to Step 2.

Step 2. First, remove the points such that $\mathbf{x}^T\mathbf{u}^* > \beta^*$ from the set A. Let A^k denote the set of removed points. Take the separating hyperplane as $g(\mathbf{u}^*) = (\beta^* + \tilde{\beta})/2$ where $\tilde{\beta} = \text{Min}\,\{\mathbf{x}^T\mathbf{u}^* | \mathbf{x} \in A^k\}$. Let $\tilde{A}^k = \{\mathbf{x} \in X | g(\mathbf{u}^*) > (\beta^* + \tilde{\beta})/2\}$. The set $\tilde{A}^k$ denotes a subregion of the category $\mathcal{A}$ in X which is decided at this stage. Rewrite $X \backslash \tilde{A}^k$ by X and $A \backslash A^k$ by A.

Next, remove the points such that $\mathbf{x}^T\mathbf{u}^* < \alpha^*$ from the set B. Let B^k denote the set of removed points. Take the separating hyperplane as $g(\mathbf{u}^*) = (\alpha^* + \bar{\alpha})/2$ where $\tilde{\alpha} = \text{Max}\,\{\mathbf{x}^T\mathbf{u}^* | \mathbf{x} \in B^k\}$. Let $\tilde{B}^k = \{\mathbf{x} \in X | g(\mathbf{u}^*) < (\alpha^* + \tilde{\alpha})/2\}$. The set $\tilde{B}^k$ denotes a subregion of the category $\mathcal{B}$ in X which is decided at this stage. Rewrite $X \backslash \tilde{B}^k$ by X and $B \backslash B^k$ by B.

Set $k = k + 1$ and go to Step 1.

Step 3. Construct a piecewise linear separating hypersurface for A and B by adopting the relevant parts of the hyperplanes obtained above.

REMARK. At the final p-th stage, we have the region of $\mathcal{A}$ in X as $\tilde{A}^1 \cup \tilde{A}^2 \cup \cdots \cup \tilde{A}^p$ and that of $\mathcal{B}$ in X as $\tilde{B}^1 \cup \tilde{B}^2 \cup \cdots \cup \tilde{B}^p$. Given a new point, its classification is easily made. Namely, since the new point is either one of these subregions in X, we can classify it by checking which subregion it belongs to in the order of $1, 2, \ldots, p$.

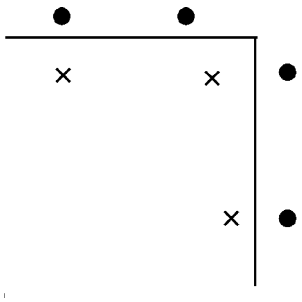

Figure 2. Classification by (II′).

3. Revision by Multi-objective Programming and Goal Programming

3.1. MULTI-OBJECTIVE PROGRAMMING APPROACH

It has been observed that MSM yields sometimes too complex discrimination hypersurfaces. For example, consider a problem given by Figure 1a. By applying MSM, we have a too complex discrimination surface as shown in Figure 1b, by which the classification is very sensitive to the noise of data. One of the reasons for this phenomenon may be that the discrimination surface is constructed by two parallel hyperplanes solving a LP problem at each iteration.

In order to overcome the difficulty of MSM, we can consider a hyperplane which has the set B on its one side and classifies correctly as many points of A as possible. This is performed by solving the following LP problem:

$$
\begin{aligned}
\text{(II)}\quad \text{Maximize}\quad & A\mathbf{u} - \beta\mathbf{1} \\
\text{subject to}\quad & B\mathbf{u} - \beta\mathbf{1} \leqq \mathbf{0} \\
& -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1}
\end{aligned}
$$

Problem (II) is a multi-objective optimization problem. We can take a scalarization of the linear sum of the objective functions. For this scalarization, we have

$$
\begin{aligned}
\text{(II}'\text{)}\quad \text{Maximize}\quad & \sum_{i=1}^{m} (y_i^+ - y_i^-) \\
\text{subject to}\quad & A\mathbf{u} - \beta\mathbf{1} = \mathbf{y}^+ - \mathbf{y}^- \\
& B\mathbf{u} - \beta\mathbf{1} \leqq \mathbf{0} \\
& -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1} \\
& \mathbf{y}^+, \mathbf{y}^- \geqq \mathbf{0}
\end{aligned}
$$

More generally, considering the symmetric role of A and B, we solve another LP problem in addition to (II′):

$$\begin{aligned}(\mathrm{III})\quad &\text{Minimize} && B\mathbf{u} - \alpha\mathbf{1}\\ &\text{subject to} && A\mathbf{u} - \alpha\mathbf{1} \geqq \mathbf{0}\\ &&& -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1}\end{aligned}$$

from which

$$\begin{aligned}(\mathrm{III}')\quad &\text{Minimize} && \sum_{i=1}^{m} (z_i^+ - z_i^-)\\ &\text{subject to} && A\mathbf{u} - \alpha\mathbf{1} \geqq \mathbf{0}\\ &&& B\mathbf{u} - \alpha\mathbf{1} = \mathbf{z}^+ - \mathbf{z}^-\\ &&& -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1}\\ &&& \mathbf{z}^+, \mathbf{z}^- \geqq \mathbf{0}\end{aligned}$$

In order to avoid a trivial solution $\mathbf{u} = 0$, $\mathbf{y} = 0$, $\mathbf{z} = 0$, $\alpha = 0$, $\beta = 0$, we add the constraint (1) to each LP problem above.

At the beginning, set $k = 1$. Let A^k be the set of points of A such that $\mathbf{x}^T\mathbf{u}^* > \beta^*$ for the solution to (II′), and let n_A denote the number of elements of A^k. Similarly, let B^k be the set of points of B such that $\mathbf{x}^T\mathbf{u}^* < \alpha^*$ for the solution to (III′), and let n_B denote the number of elements of B^k.

If $n_A \geqq n_B$, then we take a separating hyperplane as $g(\mathbf{u}^*) = (\beta^* + \tilde{\beta})/2$ where β^* is the solution to (II′) and $\tilde{\beta} = \mathrm{Min}\{\mathbf{x}^T\mathbf{u}^* | \mathbf{x} \in A^k\}$. Let $\tilde{A}^k = \{\mathbf{x} \in X | g(\mathbf{u}^*) > (\beta^* + \tilde{\beta})/2\}$. The set $\tilde{A}^k$ denotes a subregion of the category $\mathcal{A}$ in X which is decided upon at this stage. Rewrite $X\backslash\tilde{A}^k$ by X and $A\backslash A^k$ by A.

If $n_A < n_B$, then we take a separating hyperplane as $g(\mathbf{u}^*) = (\alpha^* + \tilde{\alpha})/2$ where α^* is the solution to (III′) and $\tilde{\alpha} = \mathrm{Max}\{\mathbf{x}^T\mathbf{u}^* | \mathbf{x} \in B^k\}$. Let $\tilde{B}^k = \{\mathbf{x} \in X | g(\mathbf{u}^*) < (\alpha^* + \tilde{\alpha})/2\}$. The set $\tilde{B}^k$ denotes a subregion of the category $\mathcal{B}$ in X which is decided upon at this stage. Rewrite $X\backslash\tilde{B}^k$ by X and $B\backslash B^k$ by B.

Setting $k = k + 1$, solve the problem (II′) and (III′) until we have $A\backslash\tilde{A}^k = \emptyset$ and $B\backslash\tilde{B}^k = \emptyset$.

REMARK. At the final p-th stage, we have the region of $\mathcal{A}$ in X as $\tilde{A}^1 \cup \tilde{A}^2 \cup \cdots \cup \tilde{A}^p$ and that of $\mathcal{B}$ in X as $\tilde{B}^1 \cup \tilde{B}^2 \cup \cdots \cup \tilde{B}^p$. Given a new point, its classification is easily made. Namely, since the new point is either one of these subregion in X, we can classify it by checking which subregion it belongs to in the order of $1, 2, \ldots, p$.

3.2. Goal Programming Approach

Unlike the method by multi-objective optimization given above, we can also consider another method obtaining a hyperplane which has one of the sets on its one side and minimizes the sum of distances between misclassified data of another set and the hyperplane at each iteration. This is formulated as goal programming, and was originally given by Benett–Mangasarian [1], who named it the robust linear programming discrimination (RLPD). The algorithm is summarized as follows:

In order to minimize the sum of distances between misclassified data in the class $\mathcal{A}$ and the hyperplane by the separating hyperplane, we can solve the linear programming problem

$$\begin{aligned} \text{(IV)}\quad \text{Minimize}\quad & \sum_{j=1}^{k} y_j \\ \text{subject to}\quad & -A\mathbf{u} + \beta\mathbf{1} \leqq \mathbf{y} \\ & B\mathbf{u} - \beta\mathbf{1} \leqq \mathbf{0} \\ & -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1} \\ & \mathbf{y} \geqq \mathbf{0} \end{aligned}$$

Similarly, for class $\mathcal{B}$, we have

$$\begin{aligned} \text{(V)}\quad \text{Minimize}\quad & \sum_{i=1}^{m} z_i \\ \text{subject to}\quad & A\mathbf{u} - \alpha\mathbf{1} \geqq \mathbf{0} \\ & B\mathbf{u} - \alpha\mathbf{1} \leqq \mathbf{z} \\ & -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1} \\ & \mathbf{z} \geqq \mathbf{0} \end{aligned}$$

In order to avoid a trivial solution $\mathbf{u} = 0$, $\mathbf{y} = 0$, $\mathbf{z} = 0$, $\alpha = 0$, $\beta = 0$, we add the constraint (1) to each LP problem above. We can make the same procedure as the multi-objective optimization stated above taking (IV) in place of (II′) and (V) in place of (III′).

3.3. Classification of New Data

A point $\mathbf{x} \in R^n$ is classified into $\mathcal{A}$ or $\mathcal{B}$ as follows: The judgment is made by hyperplanes obtained as solutions to corresponding LP problems. Suppose that the given new point is not classified definitely for the first $k-1$ hyperplanes, and that it is classified definitely by the k-th hyperplane (i.e., it is in the subset removed by the k-th hyperplane).

- If the k-th hyperplane is obtained by solving (II′) or (IV), then the point is judged to be in $\mathcal{A}$.
- If the k-th hyperplane is obtained by solving (III′) or (V), then the point is judged to be in $\mathcal{B}$.

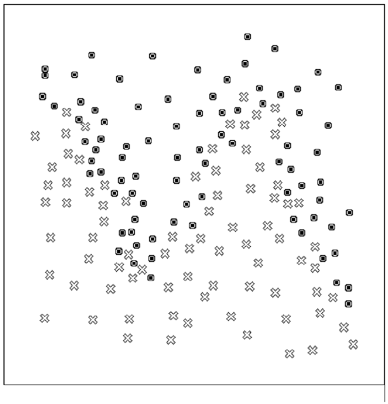

Figure 3. Data of case 1.

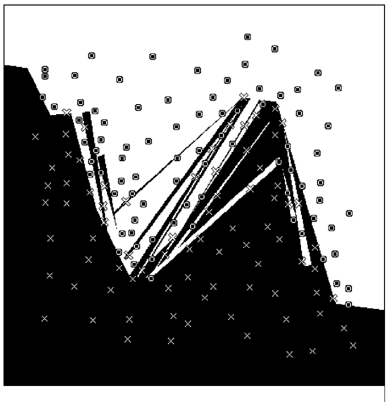

Figure 4. MSM (case 1).

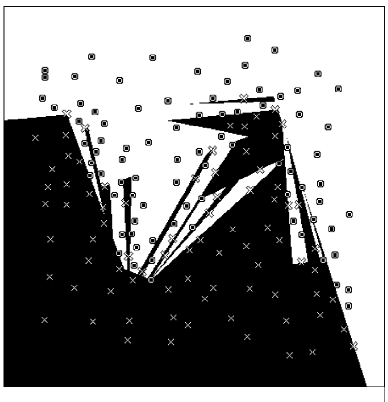

Figure 5. RLPD (case 1).

The result by MSM and RLPD for the data of Figure 3 is given in Figure 4 and Figure 5, respectively.

4. Application of Fuzzy Mathematical Programming

The multi-objective programming approach and the goal programming approach make a perfect classification for given teacher's data as well as MSM. As a result, we have too complex discrimination surfaces which cause a poor generalization ability due to the high sensitivity to the noise of data.

In order to overcome this difficulty, we can introduce fuzzy mathematical programming into the methods so that some critical points may be unclassified. In the following, a case for the goal programming approach will be stated.

In problem (IV), the constraints $B\mathbf{u} - \beta\mathbf{1} \leqq \mathbf{0}$ may be fuzzified into

$$B\mathbf{u} - \beta\mathbf{1} \lesssim \mathbf{0}$$

where $\lesssim$ implies "almost $\leqq$". The degree of satisfaction of the inequality $B\mathbf{u}-\beta\mathbf{1} \leqq \mathbf{0}$ can be represented by the membership function $M_B(\mathbf{x})$ denoting the degree of membership of $\mathbf{x}$ in B:

$$M_B(\mathbf{x}) = \begin{cases} 1, & -\dfrac{\mathbf{x}^T\mathbf{u} - \beta}{e} \geqq 1 \\ -\dfrac{\mathbf{x}^T\mathbf{u} - \beta}{e}, & 0 < -\dfrac{\mathbf{x}^T\mathbf{u} - \beta}{e} < 1 \\ 0, & -\dfrac{\mathbf{x}^T\mathbf{u} - \beta}{e} \leqq 0 \end{cases}$$

Similarly, the membership function $M_A(\mathbf{x})$ of the point $\mathbf{x}$ in A is given by

$$M_A(\mathbf{x}) = \begin{cases} 0, & -\dfrac{\mathbf{x}^T\mathbf{u} - \alpha}{e} - 1 \geqq 0 \\ \dfrac{\mathbf{x}^T\mathbf{u} - \alpha}{e} - 1, & -1 < \dfrac{\mathbf{x}^T\mathbf{u} - \alpha}{e} - 1 < 0 \\ -1, & \dfrac{\mathbf{x}^T\mathbf{u} - \alpha}{e} - 1 \leqq -1 \end{cases}$$

Using the usual technique in fuzzy mathematical programming, we have the following LP problems taking into account the fuzziness stated above:

$$\begin{aligned} \text{(VI)} \quad \text{Minimize} \quad & \sum_{j=1}^{k} y_j - w\lambda \\ \text{subject to} \quad & -A\mathbf{u} + \beta\mathbf{1} \leqq \mathbf{y} \\ & B\mathbf{u} - \beta\mathbf{1} + e\lambda\mathbf{1} \leqq e\mathbf{1} \\ & -\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1} \\ & \mathbf{y} \geqq \mathbf{0} \end{aligned}$$

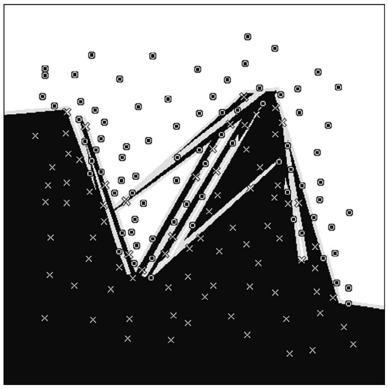

Figure 6. Fuzzy RLPD (case 1).

$$\text{(VII)} \quad \text{Minimize} \quad \sum_{i=1}^{m} z_i - w\lambda$$
$$\text{subject to} \quad A\mathbf{u} - \alpha\mathbf{1} - e\lambda\mathbf{1} \geqq -e\mathbf{1}$$
$$B\mathbf{u} - \alpha\mathbf{1} \leqq \mathbf{z}$$
$$-\mathbf{1} \leqq \mathbf{u} \leqq \mathbf{1}$$
$$\mathbf{z} \geqq \mathbf{0}$$

REMARK. The parameter e represents the width of the "gray zone" of the separating hyperplane, and w does the weight for considering the membership function. The fuzziness of the boundary between the set A and B varies according to the value of these parameters. They should be decided by users on the basis of their experiences.

The procedure is the same as in the previous section using (VI) in place of (IV) and (VII) in place of (V). The result is given in Figure 6.

5. Trinary Classification

For the trinary classification problem with three categories $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, solve a sequence of binary classification problems:

(1) $A' = A, B' = B \cup C$

(2) $A'' = A \cup C, B'' = B$

First, solve the binary classification problem with categories $\mathcal{A}'$ and $\mathcal{B}'$ associated with (1), Next, solve the binary classification problem with categories $\mathcal{A}''$ and $\mathcal{B}''$ associated with (2).

For a given point $\mathbf{x}$, we judge its category as follows:

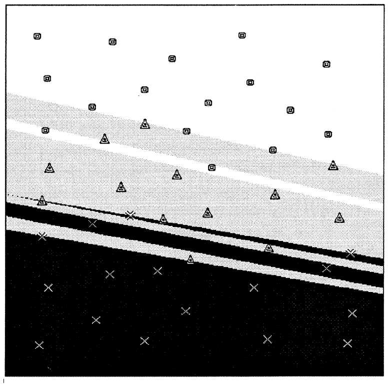

Figure 7. MSM (case 2).

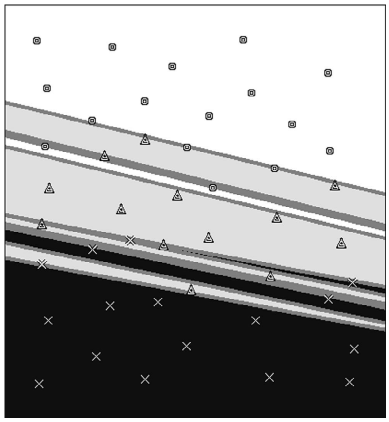

Figure 8. Fuzzy MSM (case 2).

i) if point **x** is judged to be in A' and A'', then it is judged to be in A,
ii) if point **x** is judged to be in b' and B'', then it is judged to be in B,
iii) if point **x** is judged to be neither in A nor in B, then it is judged to be in C.

In case with fuzzy mathematical programming, we decide as follows:
i) if point **x** is judged to be in A' and A'', then it is judged to be in A,
ii) if point **x** is judged to be in b' and B'', then it is judged to be in B,
iii) if point **x** is in a gray zone, then it is judged to be in a gray zone.
iv) if point **x** is judged to be neither in A nor in B nor in a gray zone, then it is judged to be in C.

An example by MSM, RLPD, the fuzzy MSM and the fuzzy RLPD is given in Figures 7–10.

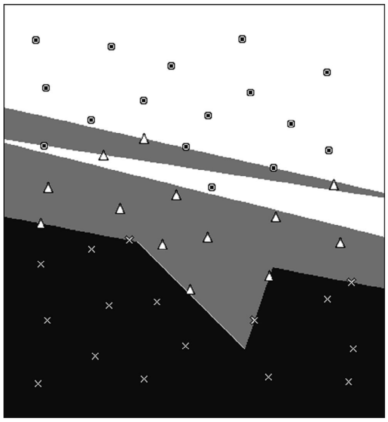

Figure 9. RLPD (case 2).

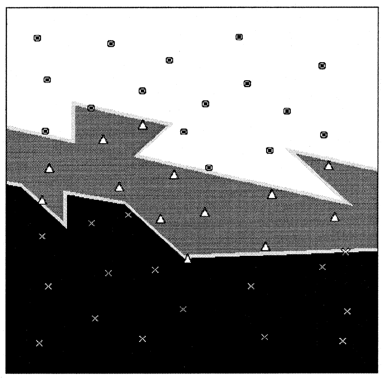

Figure 10. Fuzzy RLPD (case 2).

6. Additional Learning and Application to Portfolio Selection

The obtained discrimination surface follows just the teacher's data given in advance. In general, however, when the discrimination surface can not classify additional new data correctly, it needs to be modified. In other words, machine learning also needs additional learning like human beings so that it can follow the change of situation adaptively. In the stated methods, the additional learning can be easily made by restarting the procedure from the stage at which the new data is misjudged. The effect of additional learning can be shown along an example of application to portfolio selection as follows:

Our problem is to judge whether a stock is to be sold or to be bought (in addition, to be held in the case of trinary classification). Seven economic indices are taken into account. We have the data in 119 periods in the past for which it is already known to be sold or to be bought. We obtained a discrimination function by the

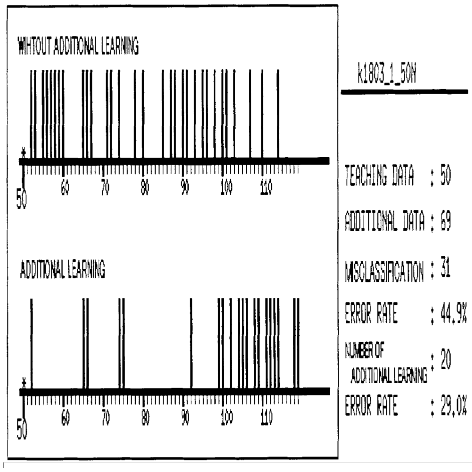

Figure 11. MSM (binary classification).

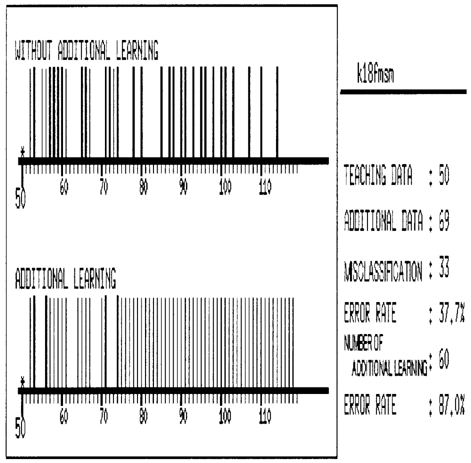

Figure 12. Fuzzy MSM (binary classification).

stated methods taking the first 50 points as the teacher's ones, and examined the ability of classification for the remaining 69 points. The following figures compare the result without additional learning and with additional learning. Flags with thick lines represent misclassified data, and those with thin lines represent data in a gray zone in the fuzzy approach.

It can be seen from Figure 11, for example, that the additional learning of the mis-classified 52-nd data enables correct judgment for the data until 64-th period while the method without additional learning misclassified many of them. On the other hand, many of data after the 99-th period were misclassified by the method with additional learning, but not by the one without additional learning. One of the

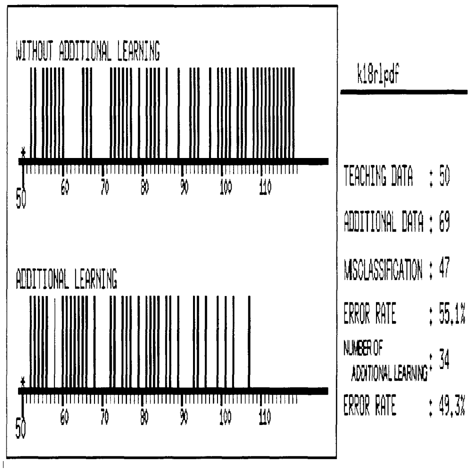

Figure 13. Fuzzy RLPD (binary classification).

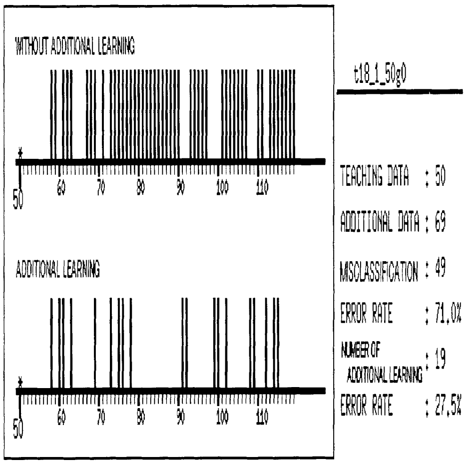

Figure 14. MSM (trinary classification).

reasons for this phenomenon may be that the rule around here is not affected by the near past, but by the far past. Another reason may be that the discrimination surface became more and more complex with additional learning. This suggests the necessity of not only additional learning but also moderate forgetting. This will be a very attractive subject of machine learning in the future.

7. Concluding Remarks

Throughout our experiments, it has been observed that the effectiveness of the stated methods depends on problems, although multi-objective programming approaches

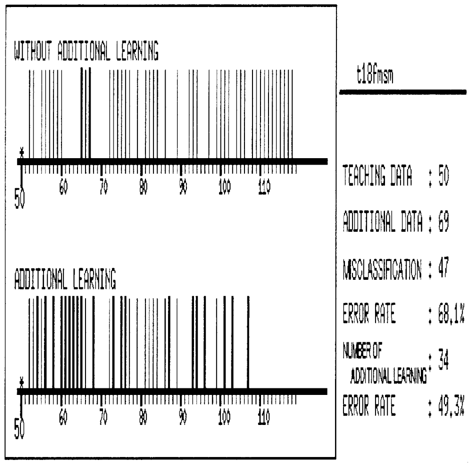

Figure 15. Fuzzy MSM (trinary classification).

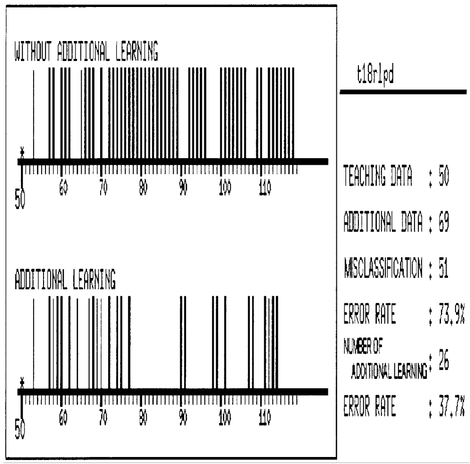

Figure 16. Fuzzy RLPD (trinary classification).

including RLDP show in general a tendency providing simpler discrimination surfaces than the original MSM. In particular, the degree of fuzzification, i.e., the width of gray zone is sensitive to the parameters e and w, and too many unclassified data appear depending on their values as was shown in Figure 12. In order to improve the effectiveness of the fuzzy mathematical programming approach, some appropriate way for deciding upon these parameters is inevitable and remains a future subject.

References

1. Bennett, K.P. and Mangasarian, O.L. (1992), Robust Linear Programming Discrimination of Two Linearly Inseparable Sets, *Optimization Methods and Software* **1**, 23–34.
2. Bishop, C.M. (1995), *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.
3. Hassoun, M.H. (1995), *Fundamentals of Artificial Neural Networks*, The MIT Press, Cambridge, Massachusetts.
4. Kagaku, N. and Nakayama, H. (1995), Pattern Classification Using Fuzzy Linear Programming, *Proceedings of 34th SICE Annual Conference*, 443–444 (in Japanese).
5. Mangasarian, O.L. (1968), Multisurface Method of Pattern Separation, *IEEE Transactions on Information Theory* **IT-14**, 801–807.
6. Mangasarian, O.L. (1993), Mathematical Programming in Neural Networks, *ORSA Journal on Computing* **5**, 349–360.
7. Nakayama, H. and Kagaku, N. (1996), Pattern Classification by Linear Goal Programming and its Applications, *International Conference on Multi-Objective Programming and Goal Programming*, Cordoba & Malaga, Spain.
8. Nakayama, H. and Nasu, K. (1994), Pattern Classification by Linear Programming, *XI-th International Conference on Multiple Criteria Decision Making* Coimbra, Portugal.
9. Takiyama, R. and Fukudome, K. (1992), A Comparison of Serial and Parallel Processings by Neurons, *Proceedings of International Joint Conference on Neural Networks*, Beijin, China, II, 811–814.
10. Wolbeg, W.H. and Mangasarian, O.L. (1990), Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology, *Proceedings of National Academy of Sciences USA* **87**, 9193–9196.