Applied Mathematical Sciences, Vol. 1, 2007, no. 15, 723 - 737

Discriminant Analysis of Imprecise Data

F. Hosseinzadeh Lotfi

Department of Mathematics, Science and Research Branch
Islamic Azad University
P.O. Box 14155/775 and 14155/4933, Post code: 1477893855, Tehran, Iran
hosseinzadeh_lotfi@yahoo.com

G. R. Jahanshahloo

Department of Mathematics
Teacher Training University, Tehran, Iran

F. Rezai Balf

Department of Mathematics
Islamic Azad University, Qaemshahr, Iran

H. Zhiani Rezai

Department of Mathematics
Islamic Azad University, Mashhad, Iran

Abstract

Data envelopment analysis technique which is developed based on the mathematical programming, evaluates the relative efficiency of a set of homogeneous decision making units. This paper shows the method of Discriminant Analysis (DA), on Imprecise Data by Data Envelopment

Analysis (DEA). DEA-Discriminant Analysis (DEA-DA) is designed to identify the existence or non-existence of an overlap between two groups, by separating hyperplane. In addition it predicts a new observation belong to which group. However there are similarities between DEA and DA. DA is a method for separating two sets with previous knowledge meanwhile DEA is a technique for separating two sets efficient and inefficient without previous knowledge. Also goal programming method can be used for both of these methods.

Mathematics Subject Classification: Operations Research, No. 90

Keywords: Data Envelopment Analysis, Imprecise Data, Discriminate Analysis

1. Introduction

Data envelopment analysis (DEA), occasionally called frontier analysis, was first put forward by Charnes, Cooper and Rhodes in 1978 [2]. It is a performance measurement technique which can be used for evaluating the relative efficiency of decision-making units (DMU) in organizations. Unlike the original DEA models, we assume further that the levels of data (inputs and outputs) are not known exactly. In fact, they are as imprecise, such as interval data, ordinal data and fuzzy data. Cooper et al. (1999) addressed the problem of imprecise data in DEA, in its general form. Discriminant Analysis (DA) is a method to predict group membership of a new sampled observation. In DA, it is assumed that all observations (G) are classified into two groups, Group1 (G_1) and Group 2 (G_2), which have n_1 and n_2 members, respectively. Based on the above description, in the absence of an overlap, we can have a separating hyperplane between two groups, to form $p^t z = d$ where, p is the normal vector and d is a constant value. Otherwise, in the existence of an overlap between the two groups, discriminant process is achieved by the following two-steps: (a) The classification and overlap identification and (b) The overlap handling. Sueyoshi (1999) has been proposed a non-parametric approach for discriminant analysis. This approach is referred to as "DEA-Discriminant Analysis (DEA-DA)" . In this paper, we extend it for Imprecise data . The remaining structure made in this article, is as follow: In Section 2 we give a backgroung, in section 3 a model is introduced under discriminant analysis for imprecise data. In section 4 we give an illustrative example and finally in section 5 conclusions are presented.

2. Background

Suppose that there are n observations $z_j = (x_j, y_j)$, $(j = 1, ..., n)$ where $x_j = (x_{1j}, ..., x_{mj})$ is the $j-th$ input vector and $y_j = (y_{1j}, ..., y_{sj})$ is the $j-th$ output vector $(j = 1, ..., n)$. Suppose that each observation has k independent factors, denoted by $(x_{1j}, ..., x_{mj}, y_{1j}, ..., y_{sj})$ such that $m + s = k$. These observations can be classified into two groups G_1 and G_2 , with n_1 and n_2 observations, respectively. Also it is assumed that $n_1 + n_2 = n$ and $G_1 \bigcup G_2 = G$. All we need to do, is finding a hyperplane with the form of $(\alpha, \beta)^t(x, y) = d$, such that $(\alpha, \beta)^t(x, y) \geq d$ for $(x, y) \in G_1$ and $(\alpha, \beta)^t(x, y) \leq d - \epsilon$ for $(x, y) \in G_2$. Note that $(\alpha, \beta)^t(x, y)$ is a linear discriminant function, d is a discriminant score of the first group (G_1), and is a discrimination score of the second group . The small positive number ϵ is used to avoid an obvious solution (i.e., all the weights are zero). Therefore we have:

$$\sum_{i=1}^{m} \alpha_i x_{ij} + \sum_{r=1}^{s} \beta_r y_{rj} \geq d,\ j \in J_1\ ,\ \sum_{i=1}^{m} \alpha_i x_{ij} + \sum_{r=1}^{s} \beta_r y_{rj} \leq d - \epsilon,\ j \in J_2 \quad (1)$$

where $J_1 = \{j \mid z_j = (x_j, y_j) \in G_1\}, \quad J_2 = \{j \mid z_j = (x_j, y_j) \in G_2\}$

In system (1) all $(x_{ij}, y_{ij})(i = 1, ..., m,\ r = 1, ..., s,\ j = 1, ..., n)$ have certain values.

Sueyoshi (1999) proposed a non-parametric approach for DEA-DA. Also he introduced a type of non-parametric approach, referred to as "Extended DEA-Discriminant Analysis" in the two stage classification possesses. Discriminant process is achieved by the following two-steps: (a) The classification and overlap identification and (b) The overlap handling.

Overlap Identification and Classification: The first step of DEA-DA is mathematically formulated as follows:

$$
\begin{array}{ll}
Min & \varphi = \sum_{j\in J_1} s^{+}_{1j} + \sum_{j\in J_2} s^{-}_{2j} \\
S.t. & \sum_{i=1}^{k} \alpha_i z_{ij} + s^{+}_{1j} - s^{-}_{1j} = d, \quad j \in J_1 \\
 & \sum_{i=1}^{k} \beta_i z_{ij} + s^{+}_{2j} - s^{-}_{2j} = d - \epsilon, \quad j \in J_2 \\
 & \sum_{i=1}^{k} \alpha_i = 1, \\
 & \sum_{i=1}^{k} \beta_i = 1, \\
 & all\ \ slacks \geq 0, \alpha_i \geq 0, \beta_i \geq 0,\ i = 1, ..., k\ \ and\, d:\ unrestricted
\end{array}
\tag{2}
$$

It is assumed that the vectors α and β are normalized.

2.1. Fuzzy and Interval Data

A crisp set is normally defined as a collection of elements $x \in X$. Every element can be either belong to or not belong to a set A, such as, $A \subseteq X$. But, for a fuzzy set, a element of X can be belong to the set A, by a degree of membership. Therefore, if X is a collection of objects denoted generally by x, then a fuzzy set $\tilde{A}$ in X is a of ordered pairs as follows:
$\tilde{A} = \{(x, \mu_{\tilde{A}}(x))|\, x \in X\}$.
Where, $\mu_{\tilde{A}}(x)$ is called the membership function or grade of membership x in $\tilde{A}$. The range of the membership function is a subset of the nonnegative real numbers belong to $(0, 1]$. If $Sup_x \mu_{\tilde{A}}(x) = 1$, then fuzzy set $\tilde{A}$ is called normal. Otherwise, A nonempty fuzzy set $\tilde{A}$ can always be normalized by dividing $\mu_{\tilde{A}}(x)$ by $Sup_x \mu_{\tilde{A}}(x)$. Hence, we can for convenience, assume that fuzzy sets are normalized.
The set of elements that belong to the fuzzy set $\tilde{A}$ at least to the degree $\delta\ (\delta > 0)$ is called the $\delta-$level set and we show with A_δ:
$A_\delta = \{x \in X|\, \mu_{\tilde{A}}(x) \geq \delta\}$
Note that, if $0 < \delta_1 < \delta_2$ then $A_{\delta_2} \subseteq A_{\delta_1}$. Finally, a fuzzy number $\tilde{A}$ is $LR-$type if there exist reference function L (for left) , R (for right) and scalars $\delta_1 > 0,\ \ \delta_2 > 0$ with

$$\mu_{\tilde{A}}(x) = \begin{cases} L(\frac{m-x}{\delta_1}) & x \le m \\ R(\frac{x-m}{\delta_2}) & x \ge m \end{cases}$$

where the real number m is called mean value of $\tilde{A}$ and δ_1, δ_2 are called the left and right spreads, respectively. Symbolically, $\tilde{A}$ is denoted to $(m, \delta_1, \delta_2)_{LR}$. Now, suppose that $\tilde{x_{ij}}$ and $\tilde{y_{rj}}$ are the $i-th$ input and the $r-th$ output for DMU_j with continues membership function, then the $\delta-$level set of $\tilde{x_{ij}}$ and $\tilde{y_{rj}}$ is the interval $[x^L_{ij,\delta}, x^U_{ij,\delta}]$ and $[y^L_{rj,\delta}, y^U_{rj,\delta}]$, where $x^L_{ij,\delta}$ and $y^L_{rj,\delta}$ are lower bounds and $x^U_{ij,\delta}$ and $y^U_{rj,\delta}$ are upper bounds of any $\delta-$level sets, respectively. Then for a fixed $\delta \in (0, 1]$, we can suppose that the $i-th$ input and the $r-th$ output of DMU_j are as interval data as follow:
$x_{ij} \in [x^L_{ij,\delta}, x^U_{ij,\delta}], \quad y_{rj}[y^L_{rj,\delta}, y^U_{rj,\delta}]$

Now,let us suppose that all observations $(x_j, y_j), (j = 1, ..., n)$ are interval data (recall that fuzzy data are interval data with $\delta-$level), i.e., $x_{ij} \in [x^L_{ij}, x^U_{ij}],\ (i = 1, ..., m,\ j = 1, ..., n)$ and $y_{rj} \in [y^L_{rj}, y^U_{rj}],\ \ (r = 1, ..., s,\ j = 1, ..., n)$ with positive constant lower and upper bounds of the interval . Then, according to (1) we have:

$$\sum_{i=1}^m \alpha_i[x^L_{ij}, x^U_{ij}] + \sum_{r=1}^s \beta_r[y^L_{rj}, y^U_{rj}] \ge d,\ \ j \in J_1$$
$$\sum_{i=1}^m \alpha_i[x^L_{ij}, x^U_{ij}] + \sum_{r=1}^s \beta_r[y^L_{rj}, y^U_{rj}] \le d - \epsilon,\ \ j \in J_2 \qquad (3-a)$$

It is obvious that the preceding system is satisfied, incase the foregoing system is satisfied.

Note that, system (3-a) in order to fuzzy data with $\alpha-$level convert as follow:

$$\sum_{i=1}^m \alpha_i[x^L_{ij,\delta}, x^U_{ij,\delta}] + \sum_{r=1}^s \beta_r[y^L_{rj,\delta}, y^U_{rj,\delta}] \ge d,\ \ j \in J_1$$
$$\sum_{i=1}^m \alpha_i[x^L_{ij,\delta}, x^U_{ij,\delta}] + \sum_{r=1}^s \beta_r[y^L_{rj,\delta}, y^U_{rj,\delta}] \le d - \epsilon,\ \ j \in J_2 \qquad (3-b)$$

With respect to (3-a) and (3-b), we obtain system (4-a) and system (4-b).

$$\sum_{i=1}^m \alpha_i x^{t_{ij}}_{ij} + \sum_{r=1}^s \beta_r y^{t_{rj}}_{rj} \ge d,$$
$$j \in J_1, t_{ij} \in \{L_{ij}, U_{ij}\}, (i = 1, ..., m),\ \ t_{rj} \in \{L_{rj}, U_{rj}\}, (r = 1, ..., s)$$

$$\sum_{i=1}^{m} \alpha_i x_{ij}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj}^{t_{rj}} \leq d - \epsilon,$$
$$j \in J_1, t_{ij} \in \{L_{ij}, U_{ij}\}, (i = 1, ..., m),\ t_{rj} \in \{L_{rj}, U_{rj}\}, (r = 1, ..., s) \qquad (4-a)$$

$$\sum_{i=1}^{m} \alpha_i x_{ij,\delta}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj,\delta}^{t_{rj}} \geq d,$$
$$j \in J_1, t_{ij} \in \{L_{ij,\delta}, U_{ij,\delta}\}, (i = 1, ..., m),\ t_{rj} \in \{L_{rj,\delta}, U_{rj,\delta}\}, (r = 1, ..., s)$$
$$\sum_{i=1}^{m} \alpha_i x_{ij,\delta}^{t_{ij}} + \sum_{r=1}^{s} \beta_r y_{rj,\delta}^{t_{rj}} \leq d - \epsilon,$$
$$j \in J_1, t_{ij} \in \{L_{ij,\delta}, U_{ij,\delta}\}, (i = 1, ..., m),\ t_{rj} \in \{L_{rj,\delta}, U_{rj,\delta}\}, (r = 1, ..., s) \quad (4-b)$$

Note that there are 2^{m+s} constraints for each DMU in systems (4). In other words, any interval observation is corresponded to 2^{m+s} vertex points. Figure 1 depicts how a hyperplane exactly separates two groups G_1 and G_2 . Such a hyperplane is called a "strongly separating hyperplane". Meanwhile, Figure 2 depicts another case, in which all or parts of some interval observations in G_1 , may not belong to G_1 . Similarly, all or parts of some interval observations in G_2 may not belong to G_2 . The hyperplane separating the two groups in case 2, is only called a "separating hyperplane". In this case, an overlap has occurred.

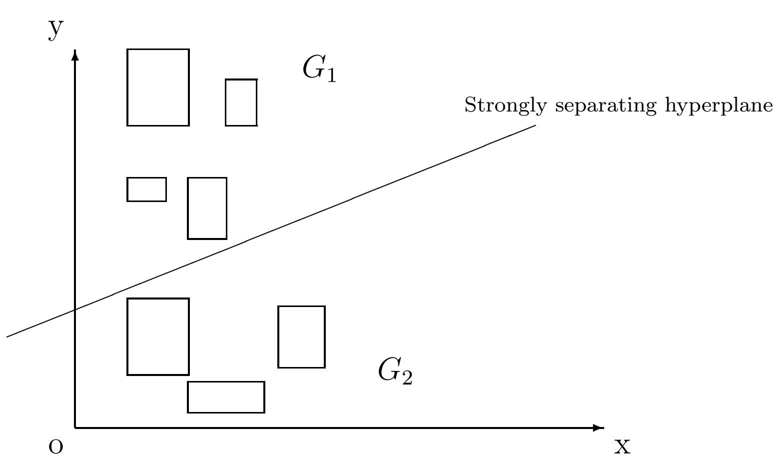

Fig. 1. A strongly separating hyperplane

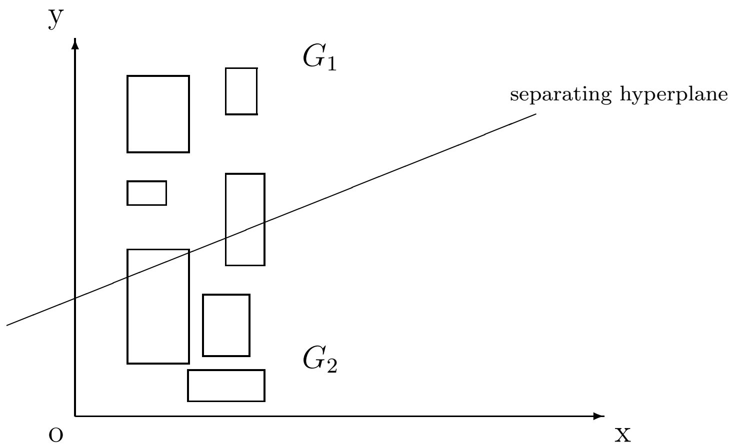

Fig. 2. A separating hyperplane

Now, we show that Sueyoshi model always can not specify an overlap between two sets as G_1 and G_2. In order to this purpose, first we show our claim with an example. Then we introduce Sueyoshi modified model.

2.2. A Numerical Example

Consider G_1 and G_2 to be the two groups of data with one input and one output factor as follows (Table 1):

Table 1. The value of inputs and output

Groups	DMU_j	x	y
Group 1 (G_1)	P_1	1	2
	P_2	2	3
	P_3	1	4
	P_4	2	5
Group 2 (G_2)	Q_1	5	6
	Q_2	6	6
	Q_3	6	8
	Q_4	7	9
	Q_5	8	7

Using model (2) for the classification of such observations, the optimal solution is obtained as follows:

$\alpha_1^* = 0, \alpha_2^* = 1, \beta_1^* = 1, \beta_2^* = 0, d^* = 6.1, \varphi* = 13.4$

In order words, we will have:

$0x + 6y \geq 6.1, j \in J_1$

$1x + 0y < 6.1, j \in J_2$

Since the total deviations's value is 13.4, there is overlap between two groups G_1 and G_2 , (Fig 3). Clearly, classification between two groups G_1 and G_2 cannot achieve correctly (With regards to Figure 3). Now let us substitute G_1 with G_2 in Table 1, and use model (2) for classification of them. Then the optimal solution is obtained as follows:

$\alpha_1^* = 1, \alpha_2^* = 0, \beta_1^* = 0.0333, \beta_2^* = 0.9667, d^* = 5, \varphi* = 0$

In other words, we will have:

$1x + 0y \geq 5, j \in J_1$

$0.0333x + 0.9667y < 5, j \in J_2$

This means, there is not overlap between two groups (Fig 4).

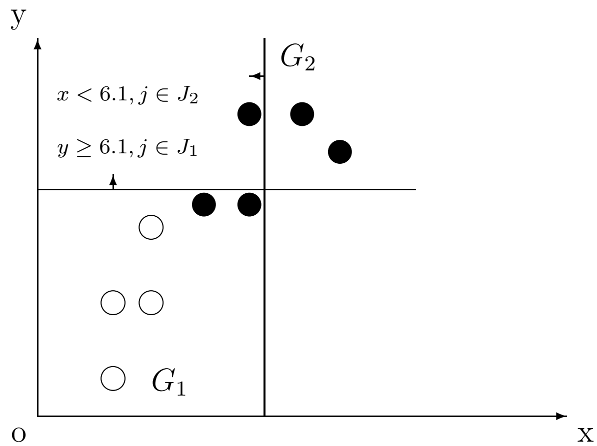

Fig. 3. Existence of overlap

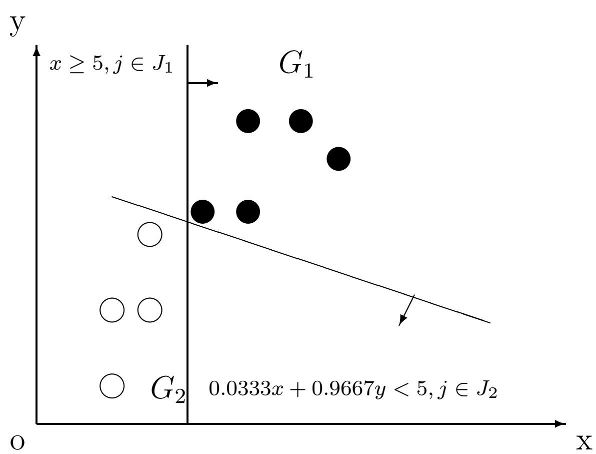

Fig. 4. Not existence of overlap

The above example shows that the Sueyoshi model should be modified as follow:

$$
\begin{array}{ll}
Min & \varphi = \sum_{j\in J_1} s_{1j}^{+} + \sum_{j\in J_2} s_{2j}^{-} \\
S.t. & \sum_{i=1}^{k} \alpha_i z_{ij} + s_{1j}^{+} - s_{1j}^{-} = d, \quad j \in J_1 \\
 & \sum_{i=1}^{k} \beta_i z_{ij} + s_{2j}^{+} - s_{2j}^{-} = d - \epsilon, \quad j \in J_2 \\
 & \sum_{i=1}^{k} \alpha_i = 1 - 2u, \\
 & \sum_{i=1}^{k} \beta_i = 1 - 2v,
\end{array}
\tag{5}
$$

$$u, v \in \{0, 1\}$$
$$all\ \ slacks \geq 0, \alpha_i \geq 0, \beta_i \geq 0,\ i = 1, ..., k\ \ and\ d:\ unrestricted$$

where $z_j = (x_j, y_j)$. Since $u, v \in \{0, 1\}$ we have: $\sum_{i=1}^{k} \alpha_i = 1$, or $\sum_{i=1}^{k} \alpha_i = -1$ and $\sum_{i=1}^{k} \beta_i = 1$, or $\sum_{i=1}^{k} \beta_i = -1$.

By using model (5) for above example, when $G_1 = \{P_1, P_2, P_3, P_4\}$ and $G_2 = \{Q_1, Q_2, Q_3, Q_4, Q_5\}$ we obtain: $\alpha_1^* = -1, \alpha_2^* = 0, \beta_1^* = -2.95, \beta_2^* = 1.95, d^* = -2, \varphi * = 0$. In other words we have:

$-x \geq -2$, for $(x, y) \in G_1$ and $-2.95x + 1.95y \leq -2$, for $(x, y) \in G_2$. This means that there is not overlap between two groups (Fig 5).

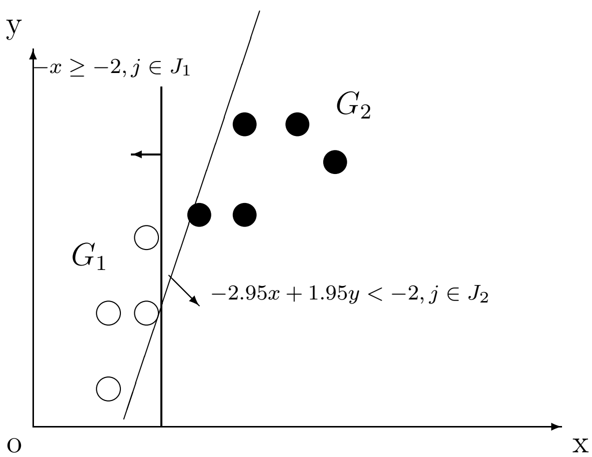

Fig. 5. No existence of overlap with model (2)

3. The Proposed Method

The general method in the overlap identification of fuzzy data are as follows: First, we set:

$\Gamma^- = \{t_{ij} \mid t_{ij} \in \{L_{ij}, U_{ij}\}, i = 1, ..., m\}$

$\Gamma^+ = \{t_{rj} \mid t_{rj} \in \{L_{rj}, U_{rj}\}, r = 1, ..., s\}$

Classification and Overlap Identification

The proposed model for overlap identification with respect to system (3-b) is given as follows:

$$Min\ \ \varphi = \sum_{j \in J_1} \sum_{(t_{1j}...t_{mj}) \in \Gamma^-} s_{1j}^{+t_{1j}...t_{mj}} + \sum_{j \in J_2} \sum_{(t_{1j}...t_{sj}) \in \Gamma^+} s_{2j}^{-t_{1j}...t_{sj}}$$

S.t.

$\sum_{i=1}^{m}\alpha_i x_{ij,\delta}^{t_{ij}}+\sum_{r=1}^{s}\beta_r y_{rj,\delta}^{t_{rj}}+s_{1j}^{+t_{1j}\dots t_{mj}}-s_{1j}^{-t_{1j}\dots t_{sj}}=d, \quad j\in J_1, t_{ij}\in\{L_{ij,\delta},U_{ij,\delta}\}, (i=1,...,m),\ t_{rj}\in\{L_{rj,\delta},U_{rj,\delta}\},(r=1,...,s)$

$\sum_{i=1}^{m}\alpha_i x_{ij,\delta}^{t_{ij}}+\sum_{r=1}^{s}\beta_r y_{rj,\delta}^{t_{rj}}+s_{2j}^{+t_{1j}\dots t_{mj}}-s_{2j}^{-t_{1j}\dots t_{sj}}=d-\epsilon, \quad j\in J_1, t_{ij}\in\{L_{ij,\delta},U_{ij,\delta}\},(i=1,...,m),\ t_{rj}\in\{L_{rj,\delta},U_{rj,\delta}\},(r=1,...,s)$ (6)

$\sum_{i=1}^{m}\alpha_i+\sum_{r=1}^{s}\beta_r=1-2u$

$s_{1j}^{+t_{1j}\dots t_{mj}}, s_{1j}^{-t_{1j}\dots t_{sj}}, s_{2j}^{+t_{1j}\dots t_{mj}}, s_{2j}^{-t_{1j}\dots t_{sj}} \geq 0, j=1,...,n, t_{ij}\in\{L_{ij,\delta},U_{ij,\delta}\},(i=1,...,m),\ t_{rj}\in\{L_{rj,\delta},U_{rj,\delta}\},(r=1,...,s),\ u,v\in\{0,1\}\ ,\alpha,\beta\ \ d:$ *unrestricted*

where $s_{1j}^{+t_{1j}\dots t_{mj}}$ and $s_{1j}^{-t_{1j}\dots t_{sj}}$ are respectively positive and negative deviations of a linear discriminant function $h(\delta)=\sum_{i=1}^{m}\alpha_i x_{ij,\delta}+\sum_{r=1}^{s}\beta_r y_{rj,\delta}$ from a discriminant score d in G_1. The positive deviation $(s_{1j}^{+t_{1j}\dots t_{mj}}>0, j\in J_1, t_{ij}\in\{L_{ij,\delta},U_{ij,\delta}\},(i=1,...,m))$ is used to minimize the incorrect classification of the $j-th$ observation in G_1 . Meanwhile, the negative deviation $(s_{1j}^{-t_{1j}\dots t_{sj}}>0, j\in J_1, t_{rj}\in\{L_{rj,\delta},U_{rj,\delta}\},(r=1,...,s))$ shows the correct classification of the $j-th$ observation in G_1 . Also, $s_{2j}^{+t_{1j}\dots t_{mj}}$ and $s_{2j}^{-t_{1j}\dots t_{sj}}$ are respectively positive and negative deviations of the same linear discriminant function $h(\delta)=\sum_{i=1}^{m}\alpha_i x_{ij,\delta}+\sum_{r=1}^{s}\beta_r y_{rj,\delta}$ from a discriminant score $d-\epsilon$ in G_2. The positive number, ϵ , is used to avoid obvious solution (all weights equal to zero). In this case, the negative deviation $(s_{2j}^{-t_{1j}\dots t_{sj}}>0, j\in J_2, t_{rj}\in\{L_{rj,\delta},U_{rj,\delta}\},(r=1,...,s))$ indicates an incorrect group classification, while the positive deviation $(s_{2j}^{+t_{1j}\dots t_{mj}}>0, j\in J_2, t_{ij}\in\{L_{ij,\delta},U_{ij,\delta}\},(i=1,...,m))$ represents a correct classification.

Now, let α^*,β^* vectors and d^* be the optimal solutions of (6). Then, a new observation with $\gamma-$level, $(x_{p,\gamma},y_{p,\gamma})\in([x_{p,\gamma}^L,x_{p,\gamma}^U],[y_{p,\gamma}^L,y_{p,\gamma}^U])$ is classified by the following rule:

(a) If $\sum_{i=1}^{m}\alpha_i^* x_{ip,\gamma}^{t_{ip}}+\sum_{r=1}^{s}\beta_r^* y_{rp,\gamma}^{t_{rp}}\geq d^*,\ t_{ip}\in\{L_{ip,\gamma},U_{ip,\gamma}\},(i=1,...,m),\ t_{rp}\in\{L_{rp,\gamma},U_{rp,\gamma}\},(r=1,...,s)$ then $p\in J_1$ with $\gamma-$level (7).

(b) If $\sum_{i=1}^{m}\alpha_i^* x_{ip,\gamma}^{t_{ip}}+\sum_{r=1}^{s}\beta_r^* y_{rp,\gamma}^{t_{rp}}\leq d^*-\epsilon, t_{ip}\in\{L_{ip,\gamma},U_{ip,\gamma}\},(i=1,...,m),\ t_{rp}\in\{L_{rp,\gamma},U_{rp,\gamma}\},(r=1,...,s)$ then $p\in J_2$. with $\gamma-$level

(c) If $\sum_{i=1}^{m}\alpha_i^* x_{ip,\gamma}^{t_{ip}}+\sum_{r=1}^{s}\beta_r^* y_{rp,\gamma}^{t_{rp}}<d^*, \exists t_{ip}\in\{L_{ip,\gamma},U_{ip,\gamma}\},(i=1,...,m),\ t_{rp}\in$

$\{L_{rp,\gamma}, U_{rp,\gamma}\}, (r = 1, ..., s)$ then then all or parts of the new fuzzy observation belong to G_2. And also, if $\sum_{i=1}^{m} \alpha_i^* x_{ip,\gamma}^{t_{ip}} + \sum_{r=1}^{s} \beta_r^* y_{rp,\gamma}^{t_{rp}} > d^*$, $\exists t_{ip} \in \{L_{ip,\gamma}, U_{ip,\gamma}\}, (i = 1, ..., m)$, $t_{rp} \in \{L_{rp,\gamma}, U_{rp,\gamma}\}, (r = 1, ..., s)$ then all or parts of the new fuzzy observation belong to G_1 . Hence, we say that, there is overlap in this case.

Theorem 1. Model (6) has always a bounded optimal solution. [5]

Theorem 2. Suppose that $F_1 = \{j | j \in J_1\}$ and $F_2 = \{j | j \in J_2\}$ also assume that $C(F_1)$ and $C(F_2)$ be convex hull of F_1 and F_2, respectively. Then $C(F_1) \bigcap C(F_2) = \emptyset$ if and only if $\varphi^* = 0$ in model (6).[5]

4. An Illustrative Example

Consider two groups of fuzzy data with one input and one output (Table 2) as follows:

Table 2. The inputs and outputs values

DMU_s	*Observation* j	*Input* (m, δ_1, δ_2)	output (m, δ_1, δ_2)
Group 1 (G_1)	A	(5,3,4)	(6,2,4)
	B	(7,1,3)	(10,2,6)
	C	(9,2,5)	(10,2,6)
Group 2 (G_2)	D	(3,2,5)	(4,4,2)
	E	(7,3,2)	(5,4,3)

Using model (6), and $\delta-$level ($\delta = 0.1$), we obtain: $\alpha_1^{*+} = 0.2289, \alpha_1^{*-} = 0, \alpha_2^{*+} = 0.7711, \alpha_2^{*-} = 0, d^* = 6$. Then, we will obtain: $0.2289x + 0.7711y = 6$

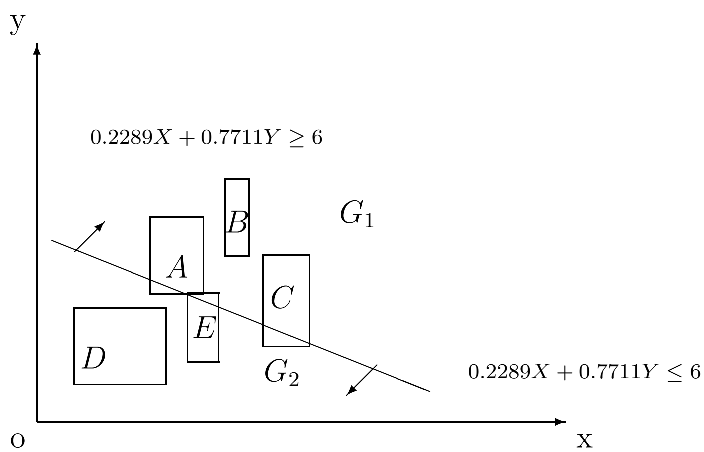

Fig.6. Illustration of example 4

Figure 6 shows a separating hyperplane between two groups G_1 and G_2 in dimension space (x, y). It is clear that we haven't got a strongly separating hyperplane. Therefore, an overlap occurs between two groups, G_1 and G_2 . This hyperplane is: $h(0.1) = 0.2289X + 0.7711Y = 6$.

5. Conclusion

The purpose of this study was an extension of DEA-DA (Sueyoshi, 1999), upon imprecise data, it identifies the measure of the overlapped regions of imprecise data, if there are any. When all or parts of an observation are overlapped between two groups, then we can clarify both the similarities that each of two groups has in common, and the differences the two groups owe. Furthermore, it is distinguished how much of an individual unit owes the characteristics of either groups. Also, in cause of data's nature, the overlap identification may be all or parts of an observation.

References

[1]Bazara, MS, HD Sherali and CM Shetty (1993) Nonlinear Programming: Theory and Algorithms (2 nd Edition), John Wiley and Sons, New York.

[2] Charnes, A, WW Cooper and E Rhodes (1978). Measuring the Efficiency of Decision Making Units. European Journal of Operating Research, 2(6), pp. 429-444.

[3] Cooper, WW, KS Park and G Yu (1999). IDEA and AR-IDEA: Models for dealing with imprecise data in DEA. Management Science 45, 597-607.

[4] Jahanshahloo, GR, F Hosseinzadeh Lotfi, H Zhiani Rezai and F Rezai Balf (2005). Using Monte Carlo method for ranking efficient DMUs. Applied Mathematics and Computation, 162, 371-379.

[5] Jahanshahloo, GR, F Hosseinzadeh Lotfi, F Rezai Balf and H Zhiani Rezai (2005). Discriminant Analysis of Interval Data Using Monte Carlo Method in Assessment of Overlap,(AMC inpress)

[6] Reuven, YR (1981). Simulation and the Monte Carlo Method, John Wiley and Sons, New York.

[7] Sueyoshi T., Kirihara, Y., (1998), Efficiency Measurement and strategic classification of Japanese banking institutions. International Journal of Systems Science, 29, 1249-1263.

[8] Sueyoshi, T(1999). DEA-discriminant analysis in the view of goal programming. European Journal of Operational Research 115, 564-582.

[9] Sueyoshi T., (2001), Extended DEA-Discriminant Analysis, European Journal of Operational Research 131, 324-351.

[10] Sueyoshi T., (2004), Mixed integer programming approach of extended DEA discriminant analysis, European Journal of Operational Research 152,

45-55.

Received: April 11, 2006