Available online at http://ijdea.srbiau.ac.ir

Int. J. Data Envelopment Analysis (ISSN 2345-458X)

Vol. 11, No. 3, Year 2023 Article ID IJDEA-00422, Pages 12-23
Research Article

International Journal of Data Envelopment Analysis

Science and Research Branch (IAU)

Data Envelopment Analysis-Discriminant Analysis by imprecise data for more than two groups: apply to the pharmaceutical stock companies

Sarah Navidi*, Mohsen Rostamy-Malkhalifeh
Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Received 24 March 2023, Accepted 18 June 2023

Abstract
One of the interesting subjects that amuse the mind of researchers is surmising the correct classification of a new sample by using available data. Data Envelopment Analysis (DEA) and Discriminant Analysis (DA) can classify data by each one alone. DEA classifies as efficient and inefficient groups and DA classify by historical data. Merge these two methods is a powerful tool for classifying the data. Since, in the real world, in many cases we do not have the exact data, so we use imprecise data (e.g. fuzzy and interval data) in these cases. So, in this paper, we represent our new DEA-DA method by using Mixed-Integer Nonlinear Programming (MINLP) to classify with imprecise data to more than two groups. Then we represent an empirical example of our purpose method on the Iranian pharmaceutical stock companies' data. In our research, we divided pharmaceutical stock companies into four groups with imprecise data (fuzzy and interval data). Since, most of the classical DA models used for two groups, the advantage of the proposed model is beheld. The result shows that the model can predict and classify more than two groups (as many as we want) with imprecise data so correct.

Keywords: Data Envelopment Analysis, Discriminant Analysis, Mixed-Integer Nonlinear Programming, Imprecise data, Classification.

* Corresponding author: Email: s_navidi@yahoo.com

1. Introduction

To predict the group membership of a new observation frequently is used Discriminant Analysis (DA). In DA, the observations classify to the specified group by some defined factors. By using the memberships of these observations, we can predict the membership of new observations.

Data Envelopment Analysis (DEA) is a powerful tool in optimization, management, and decision science for classifying the units to two groups such as efficient and inefficient. DEA has lots of applied models such as BCC (Banker et al. [1]), CCR (Charnes et al. [2]), Additive, and so on. In (1999) the additive model of DEA compared with represented GP approach for DA. The comparison showed their likeness and discrepancy. Sueyoshi believed that incorporate DA and DEA in the framework of GP is so useful and helps us to specify the group membership of new observation, more precise. So, he presented his DEA-DA method by using GP in 1999 then he completed his model in 2001 (Sueyoshi [3, 4]).

In real cases, often we have not accessed exact data, so we use imprecise data in these situations. Moreover, some data should be presented as an interval or fuzzy because of their nature. On discriminant analysis, there are numerable papers for interval data (Jahanshahloo et al. [5], Duarte Silva et al. [6, 7], Angulo et al. [8]) and fuzzy data (Hosseinzadeh Lotfi et al. [9, 10], Khalili-Damghani et al. [11], Omrani et al. [12], Ji et al. [13]).

There are lots of fuzzy methods have been presented to solve the fuzzy DEA model [14], the fuzzy DEA methods can be categorized into these groups: the tolerance approach [15], the α-cuts approach [16, 17], the fuzzy ranking approach [18, 19], the possibility approach [20, 21], other developments [22, 23, 24] and cross-efficiency fuzzy DEA [25].

Sueyoshi [26] represented another procedure to classify observations more correctly. He extended the GP approach by Mixed Integer Programming (MIP) models. A binary variable is counting the number of misclassified observations. Sueyoshi in 2004 represented the two-stage MIP version of DEA-DA and methodologically compared it with the extended GP model. Sueyoshi [27] in 2005 reformulated his two-stage MIP model into one stage. Most of the presented methods are just useful for classifying observation into two groups, Sueyoshi [28] in 2006 completing his model and presented his one stage MIP version of DEA-DA for more than two groups.

DEA is one of the best tools for portfolio selection and portfolio optimization. There are lots of researchers that want to find an easier and accurate method in this field (Navidi et al. [29, 30], and Banihashemi et al. [31, 32]). In portfolio cases, some data are imprecise (Peykani et al. [33]).

The DEA-DA method is used in lots of management cases. Most of the presented DEA-DA methods classify the observation into two groups. But, in the real world, we have more than two groups. e.g. customer clubs usually separate their customers with the platinum card (loyal customers), gold card (good customers), silver card (average customers), and blue card (new customers). This classification helps the manager to present the best services appropriate to each customer. Another e.g. assume that the financier who wants to inaugurate a company use discriminant analysis to decide what is best. By using discriminant analysis can distribute available stock companies by their historical data to defined groups such as great, good, average, and weak stock companies. So, in this paper, we distribute available pharmaceutical stock companies by their attainable data to defined groups (that are more than two groups) with

imprecise data; then by using this information we predict the group membership of new pharmaceutical stock companies.

The remainder of this paper is organized as follows: In section 2, we reviewed some previous works in the DEA-DA method. Our proposed method is described in section 3. The empirical example of our purpose method on the Iranian pharmaceutical stock companies is represented in section 4. The conclusion is represented in section 5.

2. Background

In this section, we define previously presented models.

Sueyoshi [27] in 2005 represented his MIP model of the DEA-DA method for two groups.

Assume that there are n observations $j=(1,...,n)$ that belong to two groups (G_1 *and* G_2), Each observation defines by k independent factors $i=(1,...,k)$ indicated by Z_{ij}.

The MIP model of DEA-DA is formulated as follows:

$$min \sum_{j\in G_1} y_j + \sum_{j\in G_2} y_j \qquad (1)$$

$$s.t.\ \sum_{i=1}^{k}(\lambda_i^+ - \lambda_i^-)Z_{ij} - c + My_j \geq 0,\ j\in G_1$$

$$\sum_{i=1}^{k}(\lambda_i^+ - \lambda_i^-)Z_{ij} - c - My_j \leq -\varepsilon,\ j\in G_2$$

$$\sum_{i=1}^{k}(\lambda_i^+ + \lambda_i^-) = 1$$

$$\zeta_i^+ \geq \lambda_i^+ \geq \varepsilon\zeta_i^+, \qquad i=1,...,k$$

$$\zeta_i^- \geq \lambda_i^- \geq \varepsilon\zeta_i^-, \qquad i=1,...,k$$

$$\zeta_i^+ + \zeta_i^- \leq 1, \qquad i=1,...,k$$

$$\sum_{i=1}^{k}(\zeta_i^+ + \zeta_i^-) = k$$

$$c: unrestricted,\ \zeta_i^+ = 0/1\ ,\ \zeta_i^- = 0/1,$$

$$y_j = 0/1\ \ ,\ \ all\ other\ variable \geq 0.$$

All the observed factors Z_{ij} are connected by $\sum_{i=1}^{k}\lambda_i Z_{ij}$ where λ_i is a weight for the ith factor. These weights are limited in the way that the sum of total values of $\lambda_i = (\lambda_i^+ - \lambda_i^-)$ for all $i=1,...,k$ is unity. The binary variable y_j in the objective function minimizes the total number of incorrect classifications of the observations. The discriminant score declares by $c\,(j\in G_1)$ and $c-\varepsilon\,(j\in G_2)$, respectively. M is a given large number and ε is a given small number. Various choosing of M and ε cusses producing different weight estimates (λ_i^*, c^* (the optimal solutions of (1))). That is the defect of the MIP versions of DEA-DA. These numbers attain by a try and error approach (e.g. Sueyoshi [28] used $M = 1000$, $\varepsilon = 0.0001$ and $M = 1000000$, $\varepsilon = 0.001$).

The constraints number 4, 5, and 6 in the model (1) are nonlinear conditions (NLC). $\zeta_i^+ \geq \lambda_i^+ \geq \varepsilon\zeta_i^+$ specify the upper and lower bounds of λ_i^+ and $\zeta_i^- \geq \lambda_i^- \geq \varepsilon\zeta_i^-$ specify the upper and lower bounds of λ_i^-. If $\lambda_i^+ \geq \varepsilon > 0$ and $\lambda_i^- \geq \varepsilon > 0$ at the same time, then $\zeta_i^+ + \zeta_i^- = 2$ and the result becomes infeasible. So, these 3 conditions together, make the results feasible.

The constraint number 7 in the model (1) is the nonzero condition (NZC). If we use model (1) without $\sum_{i=1}^{k}(\zeta_i^+ + \zeta_i^-) = k$ also we have $\lambda_i^+ = 0$ and $\lambda_i^- = 0$ at the same time, the results will have many zeros in

the weight estimation therefore the classification will not be correct.

The new sample that is rth observation, whose value is defined by Z_{ir}, can be classified by the following principle:

I. $If \ \sum_{i=1}^{k} \lambda_i^* Z_{ir} \geq c^* \ \ then \ \ r \in G_1$

II. $If \ \sum_{i=1}^{k} \lambda_i^* Z_{ir} \leq c^* - \varepsilon \ \ then \ \ r \in G_2$

Sueyoshi [28] in 2006 expanded his MIP model of the DEA-DA method for more than two groups.

The MIP model of DEA-DA for more than two groups is formulated as follows:

$$min \sum_{g=1}^{h} \sum_{j \in G_g} y_j \qquad (2)$$

$$s.t. \begin{cases} \sum_{i=1}^{k} (\lambda_i^+ - \lambda_i^-) Z_{ij} - c_g + M y_j \geq 0, \\ j \in G_g \quad , \quad g = 1,...,h-1 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{k} (\lambda_i^+ - \lambda_i^-) Z_{ij} - c_g - M y_j \leq -\varepsilon, \\ j \in G_{g+1} \quad , \quad g = 1,...,h-1 \end{cases}$$

$$\sum_{i=1}^{k} (\lambda_i^+ + \lambda_i^-) = 1$$

$$\zeta_i^+ \geq \lambda_i^+ \geq \varepsilon \zeta_i^+, \qquad i = 1,...,k$$

$$\zeta_i^- \geq \lambda_i^- \geq \varepsilon \zeta_i^-, \qquad i = 1,...,k$$

$$\zeta_i^+ + \zeta_i^- \leq 1, \qquad i = 1,...,k$$

$$\sum_{i=1}^{k} (\zeta_i^+ + \zeta_i^-) = k$$

$c_g \ (g = 1,...,h-1)$: *unrestricted* , $\zeta_i^+ = 0/1$,

$\zeta_i^- = 0/1$, $y_j = 0/1$, *all other variable* ≥ 0

The different h groups separate with discriminant scores $c_g \ (g = 1,...,h-1)$.

The new sample that is rth observation, whose value is defined by Z_{ir}, can be classified by the following principle:

I. $If \ \sum_{i=1}^{k} \lambda_i^* Z_{ir} \geq c_1^* \ \ then \ \ r \in G_1$

II. $If \ \ c_{g-1}^* - \varepsilon \geq \sum_{i=1}^{k} \lambda_i^* Z_{ir} \geq c_g^*$

$then \quad r \in G_g \ (g = 2,...,h-1)$

III. $If \ \sum_{i=1}^{k} \lambda_i^* Z_{ir} \leq c_{h-1}^* - \varepsilon \ \ then \ \ r \in G_h$

($c_g^* \ (g = 1,...,h-1)$ and $\lambda_i^* = (\lambda_i^{+^*} - \lambda_i^{-^*}, \ i = 1,...,k)$ are the optimal solutions of (2))

3. Proposed method

As in the real world, we do not have certain data and the data might be an imprecise. Also, some data express as an interval or fuzzy because of their features. Based on the Sueyoshi [28] method, we propose our method that classifies observations into more than two groups for imprecise data.

Assume that X is a set. The fuzzy subset A of X is specified by a membership function $\mu_A(x)$ which associates with each $x \in X$ a number in [0,1] indicating to what degree x is a member of A: $A = \{(x, \mu_A(x)) \mid x \in X\}$.

The α-cuts of a fuzzy set A is defined as $A_\alpha = \{x \in X \mid \mu_A(x) \geq \alpha\}$.

Assume that there are n observations $j = (1,...,n)$ that are belong to h groups $(g = 1,...,h)$, Each observation defines by k independent factors $i = (1,...,k)$ indicated by Z_{ij} . Z_{ij} is presented as fuzzy sets with membership function $\mu_{Z_{ij}}$. The α-cut set of Z_{ij} is the interval $[z_{ij}^L, z_{ij}^U]$ where z_{ij}^L and z_{ij}^U are the lower and upper bounds of any α-cut sets, respectively. So, for the specified $\alpha \in [0,1]$ we can suppose that Z_{ij} are interval data as

$Z_{ij} \in [z_{ij}^L, z_{ij}^U]$. Therefore, we can assume that all Z_{ij} are interval data (fuzzy data are interval data with α-cut), $Z_{ij} \in [z_{ij}^L, z_{ij}^U]$ with permanent lower and upper bounds of the interval (for more explanation of imprecise data in DEA, see Cooper et al. [34]). Then we will have:

$$\theta = min \sum_{g=1}^{h} \sum_{j \in G_g} y_j \qquad (3)$$

$$s.t. \quad \begin{cases} \sum_{i=1}^{k} \lambda_i [z_{ij}^L, z_{ij}^U] - c_g + My_j \geq 0, \\ j \in G_g \ , \ g = 1, ..., h-1 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{k} \lambda_i [z_{ij}^L, z_{ij}^U] - c_g - My_j \leq -\varepsilon, \\ j \in G_{g+1} \ , \ g = 1, ..., h-1 \end{cases}$$

$$\sum_{i=1}^{k} |\lambda_i| = 1$$

$$c_g \ (g = 1, ..., h-1) : unrestricted,$$

$$\lambda_i : unrestricted \ , \ y_j = 0/1$$

All the observed factors $Z_{ij} \in [z_{ij}^L, z_{ij}^U]$ are connected by $\sum_{i=1}^{k} \lambda_i [z_{ij}^L, z_{ij}^U]$ where λ_i is a weight for the ith factor. These weights are limited in the way that the sum of total values of $|\lambda_i|$ for all $i = 1, ..., k$ is unity. The different h groups separate with discriminant scores $c_g (g = 1, ..., h-1)$. The binary variable y_j in the objective function minimizes the total number of incorrect classifications of the observations.

M is a given large number and ε is a given small number. Various choosing of M and ε cusses producing different weight estimates (λ_i^*, c_g^*). That is the defect of the MIP versions of DEA-DA. These numbers attain by a try and error approach (See, Sueyoshi [28]).

For solving model (3), we can solve models (4) and (5) that are it's upper and lower bound:

$$\theta^L = min \sum_{g=1}^{h} \sum_{j \in G_g} y_j \qquad (4)$$

$$s.t. \quad \begin{cases} \sum_{i=1}^{k} \lambda_i z_{ij}^U - c_g + My_j \geq 0, \\ j \in G_g \ , \ g = 1, ..., h-1 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{k} \lambda_i z_{ij}^L - c_g - My_j \leq -\varepsilon, \\ j \in G_{g+1} \ , \ g = 1, ..., h-1 \end{cases}$$

$$\sum_{i=1}^{k} |\lambda_i| = 1$$

$$c_g \ (g = 1, ..., h-1) : unrestricted,$$

$$\lambda_i : unrestricted \ , \ y_j = 0/1$$

$$\theta^U = min \sum_{g=1}^{h} \sum_{j \in G_g} y_j \qquad (5)$$

$$s.t. \quad \begin{cases} \sum_{i=1}^{k} \lambda_i z_{ij}^L - c_g + My_j \geq 0, \\ j \in G_g \ , \ g = 1, ..., h-1 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{k} \lambda_i z_{ij}^U - c_g - My_j \leq -\varepsilon, \\ j \in G_{g+1} \ , \ g = 1, ..., h-1 \end{cases}$$

$$\sum_{i=1}^{k} |\lambda_i| = 1$$

$$c_g \ (g = 1, ..., h-1) : unrestricted,$$

$$\lambda_i : unrestricted \ , \ y_j = 0/1$$

The objective function $\theta \in [\theta^L, \theta^U]$ minimizes the total number of incorrect classifications of the observations.

The new sample that is mth observation, whose value is defined by $Z_{im} \in [z_{im}^L, z_{im}^U]$, can be classified by the following principle:

I. $If \ \sum_{i=1}^{k} \lambda_i^{L^*} z_{im}^{L} \geq c_1^{L^*} \ \ then \ \ m \in G_1$

II. $If \ c_{g-1}^{L^*} \geq \sum_{i=1}^{k} \lambda_i^{L^*} z_{im}^{L} \geq c_g^{L^*} \ \& \ c_{g-1}^{U^*} \geq \sum_{i=1}^{k} \lambda_i^{U^*} z_{im}^{U} \geq c_g^{U^*}$
$then \quad m \in G_g \ (g = 2,...,h-1)$

III. $If \ c_{g-1}^{L^*} \geq \sum_{i=1}^{k} \lambda_i^{L^*} z_{im}^{L} \geq c_g^{L^*} \ \& \ c_g^{U^*} \geq \sum_{i=1}^{k} \lambda_i^{U^*} z_{im}^{U} \geq c_{g+1}^{U^*}$
$then \quad m \in G_g \ (g = 2,...,h-1)$

IV. $If \ \sum_{i=1}^{k} \lambda_i^{U^*} z_{im}^{U} \leq c_{h-1}^{U^*} \ \ then \ \ m \in G_h$

($\lambda_i^{L^*}$ and $c_g^{L^*}$ are the optimal solutions of model (4), $\lambda_i^{U^*}$ and $c_g^{U^*}$ are the optimal solutions of model (5))

Theorem 1. The optimal value of the objective function of the model (3) is finite.

Proof. Let

$\lambda = e_1$

$c_g = 0$, $(g = 1,...,h-1)$

$y_j = 0$, $(j = 1,...,n)$ `

Then, with this select model (3) has a feasible solution. On the other hand, we always have:

$$0 \leq \theta = min \sum_{g=1}^{h} \sum_{j \in G_g} y_j$$

Therefore, model (3) has a bounded optimal solution, and the proof is completed.

Theorem 2. Let $\theta^*, \theta^{L^*}, \theta^{U^*}$ be the optimal solution for model (3), (4), (5) respectively. Then $\theta^{L^*} \leq \theta^* \leq \theta^{U^*}$.

Proof. Let θ^*, c_g^* be the optimal solution of model (3). Because of $z_{ij} \leq z_{ij}^U$ and

$$\sum_{i=1}^{k} \lambda_i Z_{ij} \leq c_g + My_j - \varepsilon$$

We have

$$\sum_{i=1}^{k} \lambda_i Z_{ij} \leq c_g^* + My_j - \varepsilon$$

$$\sum_{i=1}^{k} \lambda_i z_{ij}^U \leq c_g^U + My_j - \varepsilon$$

$$c_g^* + My_j - \varepsilon \leq c_g + My_j - \varepsilon \leq c_g^U + My_j - \varepsilon$$

$$\sum_{i=1}^{k} \lambda_i z_{ij} \leq \sum_{i=1}^{k} \lambda_i z_{ij}^U$$

So

$$\sum_{i=1}^{k} \lambda_i Z_{ij} \leq c_g^U + My_j - \varepsilon$$

Therefore $\theta^* \leq \theta^U$, we can prove $\theta^L \leq \theta^*$ likewise.

4. Empirical example

4.1. Data collection

The dataset was collected from the 23 Iranian pharmaceutical stock companies, from 2015 till 2019. The dataset was obtained from http://www.fipiran.com. All of the stock companies are shown by their company's symbol. Table 2 represents Z_{ij} from Iranian pharmaceutical stock companies. In this research, we used 10 financial indexes ($i1,\ldots,i10$) that express in Table 1, to distribute the 23 Iranian pharmaceutical stock companies to 4 groups. More details are in section 4.2.

Table 1. The financial indexes

$i1$	Total Current Assets	$i6$	Total Assets
$i2$	Total Current Liabilities	$i7$	Total Liabilities
$i3$	Total Stockholder Equity	$i8$	Capital
$i4$	Profit Margin	$i9$	Retained Earnings
$i5$	Gross Profit Ratio	$i10$	Cash

Table 2. Financial data of Iranian pharmaceutical stock companies

Obs	z_{1j}	z_{2j}	z_{3j}	z_{4j}
	(l,m,u)	(l,m,u)	(l,m,u)	(l,m,u)
DSOB1	(0.103,0.108,0.113)	(0.007,0.012,0.017)	(0.383,0.431,0.478)	(0.377,0.435,0.493)
FTIR1	(0.059,0.06,0.06)	(0.022,0.026,0.03)	(0.107,0.117,0.127)	(0.125,0.126,0.126)
PDRO1	(0.179,0.201,0.223)	(0.164,0.165,0.166)	(0.116,0.138,0.159)	(0.216,0.326,0.436)
JAMD1	(0.018,0.019,0.019)	(0.008,0.009,0.009)	(0.023,0.024,0.024)	(0.02,0.025,0.03)
DRZK1	(0.282,0.295,0.307)	(0.245,0.26,0.275)	(0.165,0.166,0.167)	(0.238,0.253,0.268)
IRDR1	(0.071,0.078,0.084)	(0.058,0.063,0.068)	(0.046,0.053,0.06)	(0.03,0.031,0.031)
AMIN1	(0.117,0.134,0.151)	(0.091,0.102,0.113)	(0.146,0.155,0.163)	(0.099,0.115,0.13)
ROZD1	(0.046,0.06,0.073)	(0.066,0.067,0.067)	(0.069,0.073,0.077)	(0.005,0.012,0.019)
DSIN1	(0.14,0.142,0.144)	(0.079,0.129,0.178)	(0.15,0.153,0.155)	(0.193,0.227,0.26)
ABDI1	(0.247,0.302,0.356)	(0.216,0.249,0.282)	(0.183,0.241,0.298)	(0.182,0.231,0.279)
DSNZ1	(0.138,0.14,0.142)	(0.14,0.15,0.159)	(0.085,0.091,0.096)	(0.025,0.026,0.027)
DALZ1	(0.242,0.329,0.416)	(0.129,0.225,0.32)	(0.257,0.314,0.37)	(0.345,0.372,0.398)
DJBR1	(0.288,0.306,0.324)	(0.193,0.214,0.235)	(0.265,0.27,0.275)	(0.218,0.296,0.373)
KIMI1	(0.135,0.137,0.139)	(0.109,0.114,0.118)	(0.106,0.124,0.141)	(0.147,0.152,0.156)
KSPZ1	(0.158,0.165,0.171)	(0.151,0.156,0.161)	(0.086,0.09,0.093)	(0.14,0.14,0.14)
DTDZ1	(0.099,0.104,0.108)	(0.087,0.092,0.096)	(0.056,0.065,0.074)	(0.05,0.056,0.062)
DDPK1	(0.083,0.089,0.095)	(0.08,0.084,0.088)	(0.035,0.04,0.045)	(0.033,0.036,0.038)
DLGM1	(0.078,0.088,0.098)	(0.08,0.093,0.105)	(0.064,0.08,0.095)	(0.026,0.028,0.03)
EXIR1	(0.294,0.311,0.328)	(0.288,0.298,0.308)	(0.086,0.09,0.094)	(0.083,0.087,0.091)
BRKT1	(0.097,0.115,0.132)	(0.237,0.262,0.286)	(0.457,0.582,0.706)	(0.239,0.252,0.265)
DPAK1	(0.333,0.363,0.393)	(0.397,0.436,0.475)	(0.147,0.164,0.181)	(0.162,0.212,0.262)
DFRB1	(0.266,0.295,0.323)	(0.312,0.313,0.314)	(0.173,0.187,0.2)	(0.186,0.205,0.223)
DZAH1	(0.317,0.318,0.318)	(0.453,0.463,0.473)	(0.099,0.103,0.107)	(0.014,0.07,0.125)

Table 2. Continued

Obs	z_{5j}	z_{6j}	z_{7j}
	(l,m,u)	(l,m,u)	(l,m,u)
DSOB1	(0.001,0.059,0.117)	(0.16,0.189,0.217)	(0.007,0.012,0.016)
FTIR1	(0.115,0.116,0.117)	(0.058,0.064,0.069)	(0.021,0.026,0.03)
PDRO1	(0.581,0.586,0.59)	(0.145,0.158,0.171)	(0.15,0.159,0.167)
JAMD1	(0.016,0.018,0.019)	(0.014,0.015,0.015)	(0.007,0.008,0.008)
DRZK1	(0.218,0.226,0.233)	(0.211,0.217,0.222)	(0.222,0.231,0.24)
IRDR1	(0.05,0.053,0.055)	(0.052,0.059,0.065)	(0.052,0.059,0.065)
AMIN1	(0.101,0.112,0.122)	(0.113,0.129,0.145)	(0.083,0.097,0.111)
ROZD1	(0.028,0.03,0.032)	(0.077,0.08,0.083)	(0.075,0.077,0.079)
DSIN1	(0.185,0.198,0.211)	(0.121,0.148,0.174)	(0.091,0.128,0.164)
ABDI1	(0.366,0.409,0.451)	(0.202,0.255,0.307)	(0.196,0.231,0.266)
DSNZ1	(0.075,0.083,0.09)	(0.122,0.126,0.13)	(0.133,0.135,0.137)
DALZ1	(0.272,0.312,0.352)	(0.176,0.264,0.352)	(0.112,0.199,0.286)
DJBR1	(0.158,0.195,0.232)	(0.238,0.239,0.239)	(0.176,0.191,0.205)
KIMI1	(0.078,0.087,0.096)	(0.11,0.11,0.11)	(0.103,0.104,0.104)
KSPZ1	(0.127,0.141,0.155)	(0.123,0.123,0.123)	(0.131,0.138,0.145)

DTDZ1	(0.066,0.078,0.089)	(0.085,0.089,0.092)	(0.083,0.096,0.108)
DDPK1	(0.045,0.046,0.047)	(0.061,0.067,0.073)	(0.071,0.076,0.081)
DLGM1	(0.033,0.038,0.042)	(0.088,0.1,0.112)	(0.094,0.105,0.116)
EXIR1	(0.212,0.213,0.214)	(0.216,0.223,0.23)	(0.25,0.257,0.264)
BRKT1	(0.001,0.001,0.001)	(0.49,0.559,0.627)	(0.444,0.487,0.53)
DPAK1	(0.275,0.305,0.334)	(0.294,0.311,0.328)	(0.373,0.396,0.418)
DFRB1	(0.191,0.195,0.199)	(0.219,0.237,0.255)	(0.274,0.275,0.276)
DZAH1	(0.16,0.172,0.184)	(0.273,0.286,0.299)	(0.404,0.421,0.437)

Table 2. Continued

Obs	z_{8j}	z_{9j}	z_{10j}
	(l,m,u)	(l,m,u)	(l,m,u)
DSOB1	(0.265,0.28,0.294)	(0.428,0.436,0.444)	(0.032,0.04,0.048)
FTIR1	(0.071,0.084,0.097)	(0.114,0.116,0.118)	(0.078,0.087,0.096)
PDRO1	(0.038,0.047,0.056)	(0.21,0.216,0.221)	(0.144,0.226,0.307)
JAMD1	(0.012,0.017,0.022)	(0.023,0.024,0.025)	(0.073,0.079,0.084)
DRZK1	(0.105,0.107,0.109)	(0.213,0.227,0.24)	(0.443,0.475,0.506)
IRDR1	(0.043,0.069,0.094)	(0.022,0.024,0.026)	(0.031,0.075,0.119)
AMIN1	(0.147,0.151,0.154)	(0.075,0.117,0.158)	(0.172,0.241,0.309)
ROZD1	(0.124,0.147,0.169)	(0.002,0.021,0.04)	(0.063,0.097,0.13)
DSIN1	(0.054,0.063,0.071)	(0.224,0.235,0.246)	(0.341,0.439,0.536)
ABDI1	(0.089,0.193,0.296)	(0.182,0.229,0.276)	(0.142,0.213,0.283)
DSNZ1	(0.101,0.127,0.152)	(0.028,0.029,0.03)	(0.076,0.157,0.237)
DALZ1	(0.12,0.224,0.327)	(0.368,0.377,0.385)	(0.102,0.198,0.293)
DJBR1	(0.101,0.118,0.134)	(0.346,0.377,0.407)	(0.064,0.073,0.081)
KIMI1	(0.043,0.054,0.064)	(0.133,0.154,0.174)	(0.016,0.024,0.031)
KSPZ1	(0.027,0.034,0.041)	(0.134,0.14,0.145)	(0.156,0.262,0.368)
DTDZ1	(0.062,0.064,0.066)	(0.032,0.042,0.052)	(0.024,0.026,0.028)
DDPK1	(0.023,0.035,0.046)	(0.037,0.038,0.038)	(0.025,0.036,0.047)
DLGM1	(0.06,0.07,0.08)	(0.05,0.054,0.057)	(0.036,0.045,0.054)
EXIR1	(0.101,0.105,0.109)	(0.064,0.074,0.083)	(0.102,0.105,0.107)
BRKT1	(0.598,0.747,0.896)	(0.232,0.239,0.245)	(0.03,0.077,0.124)
DPAK1	(0.111,0.122,0.133)	(0.129,0.184,0.239)	(0.139,0.203,0.267)
DFRB1	(0.08,0.085,0.09)	(0.26,0.265,0.269)	(0.332,0.357,0.381)
DZAH1	(0.058,0.249,0.44)	(0.038,0.089,0.14)	(0.081,0.082,0.083)

By using $\alpha - \text{cut} = 0.001$, we assume that all Z_{ij} are interval data as $Z_{ij} \in [z_{ij}^L, z_{ij}^U]$.

4.2. Apply the proposed method

In this section, we apply our represented method that contains models (4), (5) to our imprecise data. Table 3 present the $\lambda_i^*, (i = 1,...,10)$ and $c_g^*, (g = 1,2,3)$.

Table 3. Weight estimates and discriminant scores

c_1^{*L}	0.01	c_1^{*U}	0.02
c_2^{*L}	0	c_2^{*U}	0
c_3^{*L}	-0.02	c_3^{*U}	-0.01
λ_1^{*L}	-0.02	λ_1^{*U}	0.01
λ_2^{*L}	-0.03	λ_2^{*U}	0.05
λ_3^{*L}	0	λ_3^{*U}	0.06
λ_4^{*L}	-0.09	λ_4^{*U}	0.03
λ_5^{*L}	0.03	λ_5^{*U}	0.06
λ_6^{*L}	-0.33	λ_6^{*U}	0.36
λ_7^{*L}	-0.4	λ_7^{*U}	0.07
λ_8^{*L}	-0.01	λ_8^{*U}	0.14
λ_9^{*L}	-0.05	λ_9^{*U}	0.13
λ_{10}^{*L}	-0.01	λ_{10}^{*U}	0.06

In this research we used 10 financial indexes to distributed the 23 Iranian pharmaceutical stock companies to 4 following groups:

- G_1= {Great pharmaceutical stock companies}
- G_2= {Good pharmaceutical stock companies}
- G_3= {Average pharmaceutical stock companies}
- G_4= {Weak pharmaceutical stock companies}

Table 4 present the group membership and prediction of the group membership of Iranian pharmaceutical stock companies.

As you see in Table 4, all of the 23 pharmaceutical stock companies are classified as 100% correct. Models (4), (5) are the simple and convenient models that are used instead of solving model (3) that can correctly predict group membership. By using these models, we can predict the group membership of new pharmaceutical stock companies easily.

Table 4. Classification

Obs	Group	Prediction
DSOB1	G_1	G_1
FTIR1	G_1	G_1
PDRO1	G_1	G_1
JAMD1	G_2	G_2
DRZK1	G_2	G_2
IRDR1	G_2	G_2
AMIN1	G_2	G_2
ROZD1	G_2	G_2
DSIN1	G_2	G_2
ABDI1	G_2	G_2
DSNZ1	G_2	G_2
DALZ1	G_2	G_2
DJBR1	G_2	G_2
KIMI1	G_3	G_3
KSPZ1	G_3	G_3
DTDZ1	G_3	G_3
DDPK1	G_3	G_3
DLGM1	G_3	G_3
EXIR1	G_3	G_3
BRKT1	G_3	G_3
DPAK1	G_4	G_4
DFRB1	G_4	G_4
DZAH1	G_4	G_4

5. Conclusion

As we know, one of the important and useful subjects is surmising the correct classification of a new sample by using available data. Most of the methods represented in this field can classify observation into two groups with certain data, and there are a few of them that are used for more than two groups with imprecise data. In this paper, we represented our method by using the DEA-DA method with MINLP that can be classified observation into more than two groups as many groups as we want for imprecise data (fuzzy and interval data). We applied our purpose method on the Iranian pharmaceutical stock companies with imprecise data. As shown in Table 4, our represented method predicted all of the pharmaceutical stock companies' group membership 100% correct.

Future work can expand our framework to other alterations of the

DEA methods. Furthermore, it can be expanded into integrated numerical optimization use of our framework.

References

[1]. Banker RD, Charnes A, Coopper WW, (1984), Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis, Journal of Management Science, 30:1078-1092.

[2]. Charnes A, Cooper W.W, Rhodes E, (1978), Measuring the efficiency of decision-making units, European Journal of Operational Research, 2:429–444.

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

[4]. Sueyoshi T, (2001), Extended DEA–discriminant analysis, European Journal of Operational Research, 131:324–351.

[5]. Jahanshahloo G.R, Hosseinzadeh Lotfi F, Rezai Balf F, Zhiani Rezai H, (2007), Discriminant analysis of interval data using monte carlo method in assessment of overlap, Applied Mathematics and Computation, 191:521–532.

[6]. Duarte Silva A.P, Brito P, (2006), Linear discriminant analysis for interval data, Computational Statistics, 21:289–308.

[7]. Duarte Silva A.P, Brito P, (2015), Discriminant analysis of interval data: an assessment of parametric and distance-based approaches, Journal of Classification, 32:516–541.

[8]. Angulo C, Anguita D, Gonzalez-Abril L, Ortega J.A, (2008), Support vector machines for interval discriminant analysis, Neurocomputing, 71:1220–1229.

[9]. Hosseinzadeh Lotfi F, Jahanshahloo G.R, Rezai Balf F, Zhiani Rezai H, (2007), Discriminant Analysis of Imprecise Data, Applied Mathematical Sciences, 1:723 – 737.

[10]. Hosseinzadeh Lotfi F, Mansouri B, (2008), The Extended Data Envelopment Analysis/Discriminant Analysis Approach of Fuzzy Models, Applied Mathematical Sciences, 2:1465 – 1477.

[11]. Khalili-Damghani K, Tavana M, J. Santos-Arteaga F, (2016), A comprehensive fuzzy DEA model for emerging market assessmentand selection decisions, Applied Soft Computing, 38:676–702.

[12]. Omrani H, Shafaat Kh, Emrouznejad A, (2018), An Integrated Fuzzy Clustering Cooperative Game Data Envelopment Analysis Model with application in Hospital efficiency, Expert Systems with Applications, 114:615-628.

[13]. Ji A, Qiao Y, Liu C, (2019), Fuzzy DEA-based classifier and its applications in healthcare management, Health Care Manag Sci 22:560–568.

[14]. Hatami-Marbini A, Emrouznejad A, Tavana M, (2011), A taxonomy and review of the fuzzy Data Envelopment Analysis literature: Two decades in the making, Eur J Oper Res 214:457–472.

[15]. Sengupta JK, (1992), A fuzzy systems approach in data envelopment Analysis, Computers and Mathematics with Applications 24(8–9):259–266.

[16]. Kao C, Liu S.T, (2003), A mathematical programming approach to fuzzy efficiency ranking, Int J Prod Econ 86(2):145–154.

[17]. Puri J, Yadav S.P, (2013), A concept of fuzzy input mix-efficiency in fuzzy DEA and its application in banking, Expert Syst Appl 40(5):1437–1450.

[18]. Guo P, Tanaka H, (2001), Fuzzy DEA: a perceptual evaluation method, Fuzzy Sets Syst 119(1):149–160.

[19]. León T, Liern V, Ruiz J.L, Sirvent I, (2003), A fuzzy mathematical programming approach to the assessment of efficiency with DEA models, Fuzzy Sets Syst 139(2):407–419.

[20]. Guo P, Tanaka H, Inuiguchi M, (2000), Self-organizing fuzzy aggregation models to

rank the objects with multiple attributes, IEEE Transactions on Systems, Man and Cybernetics, Part A – Systems and Humans 30(5):573–580.

[21]. Lertworasirikul S, Shu-Cherng F, Joines J.A, Nuttle H.L.W, (2003), Fuzzy data envelopment analysis (DEA): a possibility approach, Fuzzy Sets Syst 139(2):379–394.

[22]. Muren Z.M, Cui W, (2014), Generalized fuzzy data envelopment analysis methods, Appl Soft Comput 19:215–225.

[23]. Ghasemi M-R, Ignatius J, Lozano S, Emrouznejad A, Hatami-Marbini A, (2015), A fuzzy expected value approach under generalized data envelopment analysis, Knowl-Based Syst 89:148–159.

[24]. Tavana M, Shiraz R.K, Hatami-Marbini A, Agrell PJ, Paryab K, (2013), Chance-constrained DEA models with random fuzzy inputs and outputs, Knowl-Based Syst 52:32–52.

[25]. Dotoli M, Epicoco N, Falagario M, Sciancalepore F, (2015), Acrossefficiency fuzzy Data Envelopment Analysis technique for performance evaluation of Decision-Making Units under uncertainty, Comput Ind Eng 79:103–114.

[26]. Sueyoshi T, (2004), Mixed integer programming approach of extended DEA-discriminant analysis, European Journal of Operational Research, 152:45–55.

[27]. Sueyoshi T, (2005), Financial ratio analysis of electric power industry. Asia-Pacific Journal of Operational Research, 22:349-376.

[28]. Sueyoshi T, (2006), DEA–discriminant analysis: Methodological comparison among eight discriminant analysis approaches, European Journal of Operational Research, 169:47–272.

[29]. Navidi S, Banihashemi Sh, Sanei M, (2016), Three steps method for portfolio optimization by using Conditional Value at Risk measure, Journal of New Researches in Mathematics, 2:43–60.

[30]. Navidi S, Rostamy-Malkhalifeh M, Banihashemi Sh, (2020), Using MODEA and MODM with Different Risk Measures for Portfolio Optimization, Advances in Mathematical Finance and Applications, 5:29-51.

[31]. Banihashemi Sh, Navidi S, (2017), Portfolio performance evaluation in Mean-CVaR framework: A comparison with non-parametric methods value at risk in Mean-VaR analysis, Operations Research Perspectives, 4:21– 28.

[32]. Banihashemi Sh, Navidi S, (2018), Portfolio Optimization by Using MeanSharp-βVaR and Multi Objective MeanSharp-βVaR Models, Filomat, 32:815–823.

[33]. Peykani P, Mohammadi E, Rostamy-Malkhalifeh M, Hosseinzadeh Lotfi F, (2019), Fuzzy data envelopment analysis approach for ranking of stocks with an application to Tehran stock exchange, Advances in Mathematical Finance and Applications, 4:31-43.

[34] Cooper W.W, Park K.S, Yu G, (1999), IDEA and AR-IDEA: models for dealing with imprecise data in DEA, Management Science, 45:597–607.