Gazi University Journal of Science
GU J Sci
24(1):77-84 (2011)

www.gujs.org

A Combining Mathematical Programming Method For Multi-Group Data Classification

H. Hasan ÖRKCÜ[1♠], Hasan BAL[1]

[1] *Gazi University, Arts and Science Faculty, Department of Statistics, 06500, Ankara, TURKEY*

Received: 23.03.2010 Accepted: 30.09.2010

ABSTRACT

In spite of the abundance of articles on mathematical programming models to the two-group classification problem, very few have addressed the multi-group classification problem using mathematical programming. This study presents a new multi-group data classification method based on mathematical programming. A new multi-group data classification model is proposed in this study that includes the strong properties of the mathematical programming models previously suggested for multi-group classification problems in the literature. The efficiency of proposed approach is tested on the well-known IRIS data set. The results on the IRIS data set show that our proposed method is usability and efficient on multi-group classification problems.

Key Words: *Multi-group data classification problem, mathematical programming, multi-layer neural network.*

1. INTRODUCTION

Discriminant analysis (DA) is a decision support tool with a wide range of applications, such as health applications, bankruptcy prediction, education planning, taxonomy problems, including engineering applications. DA is a multivariate statistical classification technique for separating distinct sets of objectives and allocating a new objective to a previously defined group. This technique uses the values of a set of variables associated with each unit to classify unit of unknown group membership. Discriminant methods then determine a function that classifies group membership based on the observed attributes. The resulting function is used to predict group membership of new units.

Over the last three decades, much interest has been generated in mathematical programming approaches to the statistical classification problem. The papers by Freed and Glover [1, 2] have triggered a series of papers examining both promising formulations and theoretical shortcomings of proposed methods. Since both the standard linear discriminant procedure [3] and the quadratic discriminant function procedure [4] are based on the assumption of multivariate normality for optimal performance, a number of researchers have examined mathematical programming formulations in situations in which the standard assumptions are violated. Most of these researches have focused on the two-group problem with papers proposing new mathematical programming models or evaluating the classificatory performance of

♠Corresponding author, e-mail: hhorkcu@gazi.edu.tr

proposed models against that of the standard parametric classification procedures.

Artificial neural networks (ANN), like mathematical programming methods, have also a wide ranging usage area in the classification problems. There are many studies related to the mathematical programming and artificial neural network approaches for the solutions of the two-group classification problems [5–16].

This paper presents a usability and efficient new multi-group mathematical programming approach for multi-group classification problems. In section 2, multi-group mathematical programming methods proposed in the literature are examined and proposed multi-group mathematical programming model is presented in section 3. The concepts of artificial neural networks, multi-layer network structure and back-propagation algorithm are handled in section 4. The results for IRIS data set is given in section 5. Lastly, the paper is concluded by presenting the conclusions and discussion of results.

2.MATHEMATICAL PROGRAMMING APPROACHES FOR MULTI-GROUP CLASSIFICATION

The multi-group problem has received relatively limited research interests in terms of mathematical programming approaches. Only a few papers have considered extensions into the three or more group case. A simple extension of two-group formulation to a multiple-group formulation is to use a pairwise analysis on all two-group combinations. Freed and Glover [2] proposed the decomposition of the h group problem into $\binom{h}{2}=\frac{h(h-1)}{2}$ two-group problems and the classification of the observations by pairwise application of any mathematical programming model for the two-group problem. It should be noted that this method may not yield the minimum number of misclassifications in the training sample. This is because the method may classify an observation into h different groups by the h pairwise discriminate functions, as pointed out by Pavur and Loucopoulos [17], Loucopoulos and Pavur [18]. For the multi-group classification problems, a more direct way is to use general multiple function classification model presented by Gehrlein [19].

Gehrlein [19] proposed a general multiple function classification (GMFC) model. The notations and the formulation for GMFC are given below.

$$y_j=\begin{cases}1, & \text{if unit } j \text{ is misclassified}\\ 0, & \text{if unit } j \text{ is properly classified}\end{cases},$$

Z_{ij} : observation value for variable i of unit j,

h : the number of groups,

M : a large positive constant,

k : the number of variables,

n : the total number of units.

α_{ri} : the weight assigned to variable Z_{ij} for unit j in group G_r

α_{r0} : the shifting constant for group G_r (a threshold value for group G_r)

The variables y_j ($j=1,\ldots,n$) are binary and the criterion weight α_{ri} is sign-unrestricted ($r=1, 2,\ldots,h$, $i=1, 2,\ldots,k$).

$$Min \sum_{j=1}^{n} y_j$$

Subject to: (1)

$$\alpha_{r0}+\sum_{i=1}^{k}\alpha_{ri}Z_{ij}-\alpha_{t0}-\sum_{i=1}^{k}\alpha_{ti}Z_{ij}+My_j\geq e,$$
$$\forall j\in G_r,\ j=1, 2,\ldots,n$$
$$r=1, 2,\ldots,h,\ r\neq t$$

The GMFC model classifies a unit into the group with the largest discriminant score. Note that in constraints of this model a separate discriminant function is constructed for each group.

Gochet et al. [20] derive a linear programming formulation allowing for nonparallel separating hyperplanes between the groups, an outcome not possible by any other formulation presented previously. The formulation for Gochet et al. [20] model (GCH- LP^q) is given below:

$$Min \sum_{r\in S}\sum_{t\in S_{\sim r}}\sum_{j\in P_r}\beta_{rt}^{j}$$

Subject to:

$$\beta_{rt}^{j}+\left(\alpha^{r}-\alpha^{t}\right)Z_{jr}-\gamma_{rt}^{j}=\varepsilon,\ \ \forall j\in P_r,$$
$$t\in S_{\sim r},\ r\in S \qquad (2)$$

$$\sum_{r\in S}\sum_{t\in S_{\sim r}}\sum_{j\in P_r}\left(\gamma_{rt}^{j}-\beta_{rt}^{j}\right)=q$$

$$\beta_{rt}^{j},\ \gamma_{rt}^{j}\geq 0,\quad \forall j\in P_r,\ t\in S_{\sim r},\ r\in S$$

α^r,α^t are unrestricted in sing variables. In this model, variable β_{rt}^{j} denotes the badness of fit and γ_{rt}^{j} the goodness of fit of units $j\in P_r$ with respect to group $j\in G_{\sim r}$. The small number of ε is included in order to resolve the ambiguity arising from border line cases. The decision rule associated with the model is to classify a unit j to group h, provided that $\alpha^h z_r = \max_{r \in S} \left\{ \alpha^r z_j \,/\, j \in G_r,\ r = 1, \ldots, h \right\}$.

The measure of total goodness and badness are conceptually similar to internal and external deviations previously introduced by several researchers for the two-group case [21].

Sueyoshi [9] developed a linear programming model to solve the multi-group classification problem. The multiple-group model is based on mixed-integer programming model which is proposed previously by Sueyoshi [8]. Sueyoshi's multi-group classification model can be formulated as follows:

$$Min \sum_{r=1}^{h} \sum_{j \in G_r} y_j$$

Subject to:

$$\sum_{i=1}^{k} \left(\lambda_i^+ - \lambda_i^- \right) Z_{ij} - c_r + My_j \geq 0 \quad , \qquad j \in G_r, \quad r = 1, \ldots, h-1 \qquad (3)$$

$$\sum_{i=1}^{k} \left(\lambda_i^+ - \lambda_i^- \right) Z_{ij} - c_r - My_j \leq -\varepsilon , \qquad j \in G_{r+1}, \quad r = 1, \ldots, h-1$$

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

$$\xi_i^+ \geq \lambda_i^+ \geq \varepsilon \xi_i^+ \qquad i = 1, 2, \ldots, k$$

$$\xi_i^- \geq \lambda_i^- \geq \varepsilon \xi_i^- \qquad i = 1, 2, \ldots, k$$

$$\xi_i^+ + \xi_i^- \leq 1 \qquad i = 1, 2, \ldots, k$$

$$\sum_{i=1}^{k} \left(\xi_i^+ + \xi_i^- \right) = k$$

$$\lambda_i^+ \geq 0,\ \lambda_i^- \geq 0,\ (i = 1, \ldots, k)$$

In this model, c_r ($r = 1, \ldots, h-1$) unrestricted in sign, $y_j = 0/1$, $\xi_i^+ = 0/1$, $\xi_i^- = 0/1$ and other variables are non-negative ($j = 1, \ldots, n; i = 1, \ldots, k$),

The units in groups are classified as follows:

If $\sum_{i=1}^{k} \lambda_i Z_{im} \geq c_1$, then the observation belongs to G_1.

If $c_{r-1} - \varepsilon \geq \sum_{i=1}^{k} \lambda_i Z_{im} \leq c_r$, then the observation belongs to G_r ($r = 1, \ldots, h-1$).

If $c_{h-1} - \varepsilon \geq \sum_{i=1}^{k} \lambda_i Z_{im}$, then the observation belongs to G_h.

This model produces $h-1$ different discriminant scores (c_1 to c_{h-1}) in order to classify h groups. It maintains same weight scores λ_i ($i = 1, \ldots, k$). When solving this model, $c_1 > c_2 > \ldots > c_{h-1}$ is required optimality. If such a requirement is not satisfied on optimality, then additional side constraints: $c_1 > c_2 + \varepsilon$, $c_2 > c_3 + \varepsilon, \ldots$, and $c_{h-2} > c_{h-1} + \varepsilon$ are needed to be incorporated into the model.

As noted by Sueyoshi [9], his model can solve a specific type of multiple classifications, where the "specific" implies that a whole data set can be arranged in a particular ordering. In other words, the particular ordering data implies the one that is classified into multi groups by several separation functions whose slopes (i.e., weights) are same but having different intercepts. If a data set does not have such a special ordering structure, Sueyoshi model may produce an infeasible solution or a low classification rate.

3. A NEW MATHEMATICAL PROGRAMMING MODEL FOR MULTI-GROUPS

The mathematical programming approaches proposed for multi-group classification problems in the literature have both advantages and disadvantages. Multi-group DEA-DA model proposed by Sueyoshi [9] depends on the assumption that data set has a particular ordering according to the groups; and this model are also mixed integer classification models. Multiple function classification model of Gehrlein does not require such a particular ordering in the data structure; the absence of appropriate normalization constraint in the model causes irrational solutions such as weight coefficients to be zero. Moreover, the fact that the model is a mixed integer model brings forth additional difficulty in the solutions. There is no assumption that data set has a particular ordering in the LP^q approach proposed by Gochet et al. [20] as well as the fact that all the variables are not negative in the model which shows that the model is not an integer model. This model's disadvantage, however, is the non-appropriation of normalization constraint and the problem of the decision for q constant. The model proposed by Gochet et al. [20] also produces infeasible solutions for some situations [21].

DEA-DA models, proposed by Sueyoshi for two-group and multi-group classification problems, are methodologically stronger than other models since it allows negative variables in the data and attempts the classification of units by including them in a convex sets [9]. The strong features of DEA-DA model by Sueyoshi can be combined with Gochet et al.'s classification model in our study. A new multi-group mathematical programming classification model, which is a combination of Sueyoshi's DEA-DA model, and Gochet et al.'s multi-group classification model, is given below:

$$Min \sum_{r=1}^{h}\sum_{t\neq r}^{h}\sum_{j=1}^{n} n_{rt}^{j}$$

Subject to:

$$\alpha_{r0} + \sum_{i=1}^{k}\alpha_{ri}Z_{ij} - \alpha_{t0} - \sum_{i=1}^{k}\alpha_{ti}Z_{ij} + n_{rt}^{j} - p_{rt}^{j} = \varepsilon\,,$$

$$\forall j \in G_r\,,\; j = 1,\, 2, \ldots, n \qquad (4)$$

$$\sum_{r=1}^{h}\sum_{i=0}^{k}\alpha_{ri} = 1 \qquad , r = 1,\, 2, \ldots, h\,,\; r \neq t$$

$$\alpha_{ri} \geq 0 \quad , r = 1,\, 2, \ldots, h\,,\; i = 0,\, 1, \ldots, k$$

ε is a small number.

A unit is assigned into the group that has its highest classification score with the aid of this model. The first constraint in the model (4), while the units which belong in the r^{th} group is to be assigned to its group by the aid of $\alpha_{r0} + \sum_{i=1}^{k}\alpha_{ri}Z_{ij}$ discriminant function, the units which belong in the t^{th} group is to be assigned to its group by the aid of $\alpha_{t0} + \sum_{i=1}^{k}\alpha_{ti}Z_{ij}$ discriminant function. In this model, sum of the α_{ri} classification weights are restricted to unity. Classifying the units to appropriate groups is made by adding n_{rt}^{j} and p_{rt}^{j} deviation variables to this constraint and minimizing the undesired n_{rt}^{j} deviation variable.

4.MULTI-LAYER ARTIFICIAL NEURAL NETWORKS AND USING CLASSIFICATION PROBLEMS

A neural network is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionist approach [22]. Artificial neural networks are parallel computational models which are able to map any nonlinear functional relationship between an input and an output hyperspace to desired accuracy. They are constituted by individual processing units called neurons or nodes and differ among each other in the way these units are connected to process the information and, consequently in the kind of learning protocol adopted [23].

In particular, the neurons of a feed-forward neural network are organized in three layers: the input units receive information from the outside world, usually in the form of a data file; the intermediate neurons, contained in one or more hidden layers, allow nonlinearity in the data processing, the output layer is used to provide an answer for a given set of input values. In a fully connected artificial neural network, each neuron in a given layer is connected to each neuron in the following layer by an associated numerical weight (w_{ij}). The weight connection two neurons regulate the magnitude of signal that passes between them. In addition, each neuron possesses a numerical bias term corresponding to an input of -1 whose associated weight has the meaning of a threshold value.

Rumelhart et al. [24] popularized the use of backpropagation for learning internal representation in neural networks. Backpropagation (BP) algorithm is the most widely used search technique for training neural networks. Information in an ANN is stored in the connection weights which can be thought of as the memory of the system. The purpose of BP training is to change iteratively the weights between the neurons in a direction that minimizes the error E, defined as the squared difference between the desired and the actual outcomes of the output nodes, summed over training patterns (training set data) and the output neurons. The algorithm uses a sample-by-sample updating rule for adjusting connection weights in the network. In one algorithm iteration, a training sample is presented to the network. The signal is then fed in a forward manner through the network until the network output is obtained. The error between the actual and desired network outputs is calculated and used to adjust the connection weights. Basically, the adjustment procedure, derived from a gradient descent method, is used to reduce the error magnitude. The procedure is firstly applied to the connection weights in the output layer, followed by the connection weights in the hidden layer next to output layer. This adjustment is continued backward through to network until connection weights in the first hidden layer are reached. The iteration is completed after all connection weights in the network have been adjusted.

In this study, training of the multi-layer neural networks is implemented with back-propagation algorithm and network structure that has been trained with back-propagation algorithm has been used in the solutions of the multi-group classification models.

5. COMPARISON OF CLASSIFICATION MODELS USING IRIS DATA SET

In this section, the efficiency of the proposed model is tested on the well-known IRIS data set. This data set was provided by Fisher [3] for the sampling of statistical discriminant analysis. There existed 3 different species of ornamental flowers, meaning 3 different groups; *setosa*, *versicolor*, and *virginica*. Data set comprised of 150 flowers; having 50 in each of the three groups. 4 different species were observed from each flower; *sepal* length, *sepal* width, *petal* length, and *petal* width. This data set is studied in the multi-group classification problems and pattern recognition.

Data is separated into two different sets in the discriminant analysis, like the artificial neural networks, for the purpose of comparing their classification performances. The first set is used for obtaining the discriminant function and is termed as training sample or development sample. This set, in which the discriminant function is obtained, is equivalent to the training set in artificial neural networks. The latter set is termed as holdout sample and these samples are not used for obtaining the discriminant function. The accurate performance of the method is tested by this holdout sample. The holdout sample in the discriminant analysis is equivalent to the test set in artificial neural networks.

We selected 24 data samples randomly for the training process where each group is represented by exactly the same number of samples. Training sample data set is shown in Table 1 and test set is shown in Table 2. Training sample data set is used for evaluating the performance of Fisher's linear discriminant function (FLDF), Gehrlein's multiple function classification model (GMFC), Sueyoshi's multi-group classification model (DEA-DA), Gochet et al.'s multi-group classification model (GCH- LP^q), and multi-layer structure networks trained with back-propagation algorithm.

Network structure containing seven hidden layers were used when multi-layer structure was trained with back-propagation algorithm. Maximum iteration number was chosen as 10000 while the learning rate was chosen as 0.05. All of the results are obtained by using the programs MATLAB 7.0 and WINQSB.

Table 1. IRIS data training (development) set.

	Setosa				*Versicolor*				*Virginica*			
Unit	Z_1	Z_2	Z_3	Z_4	Z_1	Z_2	Z_3	Z_4	Z_1	Z_2	Z_3	Z_4
1	5.7	4.4	1.5	0.4	5.6	3	4.1	1.3	6.5	3	5.8	2.2
2	4.8	3	1.4	0.1	5.5	2.3	4	1.3	6.3	2.9	5.6	1.8
3	5.2	4.1	1.5	0.1	5.7	3	4.2	1.2	6.3	2.5	5	1.9
4	5.5	4.2	1.4	0.2	6.7	3.1	4.7	1.5	6.7	3.1	5.6	2.4
5	4.9	3.1	1.5	0.1	5.8	2.7	4.1	1	6	3	4.8	1.8
6	4.8	3.4	1.6	0.2	5	2.3	3.3	1	5.8	2.7	5.1	1.9
7	5.4	3.7	1.5	0.2	6	2.7	5.1	1.6	6.5	3.2	5.1	2
8	5.2	3.5	1.5	0.2	5	2	3.5	1	6.8	3	5.5	2.1

Z_1 : *sepal* length, Z_2 : *sepal* width, Z_3 : *petal* length, Z_4 : *petal* width

Explicit solutions of the models and the correct classification performances obtained from training and test sets are summarized as follows:

●*Fisher's linear discriminant function (FLDF)*;

$d_1 = -138.501 + 60.310Z_1 - 4.256Z_2 - 5.400Z_3 - 76.697Z_4$ for group 1 (*Setosa*)

$d_2 = -119.289 + 50.050Z_1 - 15.651Z_2 + 13.738Z_3 - 50.435Z_4$ for group 2 (*Versicolor*)

$d_3 = -137.028 + 43.947Z_1 - 16.070Z_2 + 15.776Z_3 - 22.142Z_4$ for group 3 (*Virginica*)

%95.8 performance in training set and %89.6 performance in test set.

●*Gehrlein's multiple function classification model (GMFC)* has infeasible solution.

●*Gochet et al.'s multi-group classification model (GCH- LP^q)*;

$d_1 = 120.9900 + 0.0304Z_1 + 0.0098Z_2 - 209.4370Z_3 + 142.4284Z_4$ for group 1 (*Setosa*)

$d_2 = 120.9700 + 0.0233Z_1 - 0.1160Z_2 - 209.1665Z_3 + 141.9951Z_4$ for group 2 (*Versicolor*)

$d_3 = 120.4800 - 0.0503Z_1 + 0.1040Z_2 - 209.2884Z_3 + 142.5395Z_4$ for group 3 (*Virginica*)

%95.8 performance in training set and %84.1 performance in test set.

●*Sueyoshi's multi-group classification model (DEA-DA)*

$\lambda_1^+ = 0.2060$, $\lambda_2^+ = 0.0100$, $\lambda_3^+ = 0.0100$, $\lambda_4^+ = 0$, $\lambda_1^- = 0$, $\lambda_2^- = 0$, $\lambda_3^- = 0$, $\lambda_4^- = 0.7740$, $\xi_1^+ = 1$ $\xi_2^+ = 1$, $\xi_3^+ = 1$, $\xi_4^- = 1$ $c_1 = 0.4991$, $c_2 = 0$, $\varepsilon = 0.001$

$d = 0.2060Z_1 + 0.0100Z_2 + 0.0100Z_3 - 0.7740Z_4 \geq 0.4991$ for group 1 (*Setosa*)

$$d = 0.2060Z_1 + 0.0100Z_2 + 0.0100Z_3 - 0.7740Z_4 \leq 0.4991 - \varepsilon$$
$$d = 0.2060Z_1 + 0.0100Z_2 + 0.0100Z_3 - 0.7740Z_4 \geq 0$$
for group 1 (*Versicolor*)

$$d = 0.2060Z_1 + 0.0100Z_2 + 0.0100Z_3 - 0.7740Z_4 \leq 0 - \varepsilon$$ for group 1 (*Virginica*)

%100 performance in training set and %90.47 performance in test set.

● Multi-layer artificial neural network structure, trained with real-coded genetic algorithm, has 100% correctly classification performance in training set and 92.06% correctly classification performance in test set.

●*Proposed multi-group mathematical programming model*;

$$d_1 = 0.0140 + 0.00595Z_1 + 0.22640Z_2 + 0.020066Z_3 + 0.021456Z_4$$ for group 1 (*Setosa*)

$$d_2 = 0.00697 + 0.057056Z_1 + 0.01876Z_2 + 0.17352Z_3 + 0.01621Z_4$$ for group 2 (*Versicolor*)

$$d_3 = 0.012126 + 0.008932Z_1 + 0.001485Z_2 + 0.1601Z_3 + 0.25698Z_4$$ for group 3 (*Virginica*)

%100 performance in training set and %92.06 performance in test set.

Multi-layer structure networks trained with back-propagation and our proposed mathematical programming model are observed to have high level of accurate classification performances when training and test data set results are analyzed.

6. CONCLUSION AND FUTURE WORK

In this study, a new multi-group classification method is developed in solving multi-group classification problems. Our new mathematical programming method does not need in a particular ordering arrangement of data as well as the fact that the model does not have any integer structure.

According to the computational results of well-known IRIS data, it is seen that our new mathematical programming method is capable of solving multi-group classification problems.

For a further study, the performance of proposed method may be also investigated by using the other real-world and simulation data for multi-group classification problems.

Table 2. IRIS data test (holdout) set.

	Setosa				*Versicolor*				*Virginica*			
Unit	Z_1	Z_2	Z_3	Z_4	Z_1	Z_2	Z_3	Z_4	Z_1	Z_2	Z_3	Z_4
1	4.5	2.3	1.3	0.3	4.9	2.4	3.3	1	6.7	3.3	5.7	2.1
2	5	3.5	1.6	0.6	6.2	2.2	4.3	1.5	7.3	2.9	6.3	1.8
3	4.3	3	1.1	0.1	5.5	2.6	4.4	1.2	4.9	2.5	4.5	1.7
4	5	3.5	1.3	0.3	6	3.4	4.5	1.6	6.7	3.1	5.6	2.4
5	5.1	3.8	1.9	0.4	5.8	2.7	3.9	1.2	5.8	2.8	5.1	2.4
6	5	3.4	1.5	0.2	5.6	3	4.1	1.3	6.5	3	5.5	1.8
7	5.1	3.7	1.5	0.4	5.7	2.9	4.2	1.3	7.7	3.8	6.7	2.2
8	5.1	3.8	1.5	0.3	5.9	3	4.2	1.5	5.7	2.5	5	2
9	4.6	3.4	1.4	0.3	6.9	3.1	4.9	1.5	6.8	3	5.5	2.1
10	5.4	3.4	1.7	0.2	5.2	2.7	3.9	1.4	7.7	3	6.1	2.3
11	5.8	4	1.2	0.2	7	3.2	4.7	1.4	6.9	3.2	5.7	2.3
12	4.9	3	1.4	0.2	5.7	2.6	3.5	1	7.4	2.8	6.1	1.9
13	5	3.2	1.2	0.2	6.6	2.9	4.6	1.3	7.2	3.2	6	1.8
14	5	3	1.6	0.2	6	2.9	4.5	1.5	6.4	2.7	5.3	1.9
15	5.1	3.8	1.6	0.2	6.6	3	4.4	1.4	7.9	3.8	6.4	2
16	4.7	3.2	1.3	0.2	6.1	2.8	4	1.3	6.2	2.8	4.8	1.8
17	5.4	3.4	1.5	0.4	6.4	3.2	4.5	1.5	6.4	3.2	5.3	2.3
18	4.6	3.2	1.4	0.2	5.5	2.5	4	1.3	6.7	3.3	5.7	2.5
19	4.9	3.1	1.5	0.2	5.8	2.6	4	1.2	6.4	2.8	5.6	2.1
20	5	3.4	1.6	0.4	6.4	2.9	4.3	1.3	6.3	2.8	5.1	1.5

21	4.4	2.9	1.4	0.2	5.5	2.4	3.8	1.1	6.5	3	5.2	2
22	4.6	3.6	1	0.2	5.1	2.5	3	1.1	6.8	3.2	5.9	2.3
23	5.1	3.3	1.7	0.5	5.6	2.5	3.9	1.1	7.7	2.8	6.7	2
24	5	3.3	1.4	0.2	6.3	3.3	4.7	1.6	6.1	3	4.9	1.8
25	5.1	3.5	1.4	0.2	5.4	3	4.5	1.5	6.1	2.6	5.6	1.4
26	5	3.6	1.4	0.2	6.7	3.1	4.4	1.4	7.1	3	5.9	2.1
27	5.4	3.9	1.3	0.4	5.6	3	4.5	1.5	5.6	2.8	4.9	2
28	5.1	3.3	1.7	0.5	5.6	2.7	4.2	1.3	6.3	3.3	6	2.5
29	5.5	3.5	1.3	0.2	6	2.2	4	1	6.4	2.8	5.6	2.2
30	4.8	3.4	1.9	0.2	6.7	3.1	4.7	1.5	6.3	2.7	4.9	1.8
31	4.6	3.1	1.5	0.2	6.3	2.5	4.9	1.5	6.4	3.1	5.5	1.8
32	5.1	3.5	1.4	0.3	5.7	2.8	4.1	1.3	7.7	2.6	6.9	2.3
33	5.3	3.7	1.5	5.6	5.6	2.9	3.6	1.3	6.9	3.1	5.1	2.3
34	4.8	3.1	1.6	6.5	6.5	2.8	4.6	1.5	6.7	3	5.2	2.3
35	4.7	3.2	1.6	6.1	6.1	3	4.6	1.4	6.3	3.4	5.6	2.4
36	4.8	3	1.4	5.7	5.7	2.8	4.5	1.3	6	2.2	5	1.5
37	5.1	3.4	1.5	5.5	5.5	2.4	3.7	1	7.2	3.6	6.1	2.5
38	5.7	3.8	1.7	6.1	6.1	2.8	4.7	1.2	5.8	2.7	5.1	1.9
39	5.2	3.4	1.4	6.2	6.8	2.8	4.8	1.4	6.2	3.4	5.4	2.3
40	4.4	3.2	1.3	6.2	6.2	2.9	4.3	1.3	7.6	3	6.6	2.1
41	4.4	3	1.3	6.7	6.7	3	5	1.7	5.9	3	5.1	1.8
42	5.4	3.9	1.7	6.3	6.3	2.3	4.4	1.3	6.7	3	5	1.7

Z_1 : *sepal* length, Z_2 : *sepal* width, Z_3 : *petal* length, Z_4 : *petal* width

REFERENCES

[1] Freed, N., Glover, F., "A linear programming approach to discriminant problem", ***Decision Sciences***, 12: 68–74 (1981a).

[2] Freed, N., Glover, F., "Simple but powerful goal programming models for discriminant problems", ***European Journal of Operational Research***, 7: 44–60 (1981b).

[3] Fisher, R.A, "The use of multiple measurements in taxonomy problems", ***Annals of Eugenics***, 7: 179–188 (1936).

[4] Smith, C.A.B., "Some examples of discrimination", ***Annals of Eugenics***, 13: 272–282 (1947).

[5] Lam, K.F., Choo, E.U., Moy, J.W., "Minimizing deviations from the group mean: A new linear programming approach for the two-group classification problem", ***European Journal of Operational Research***, 88: 358–367 (1996).

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

[7] Sueyoshi, T., "DEA-Discriminant Analysis: Extended DEA-Discriminant analysis", ***European Journal of Operational Research***, 131: 324–351 (2001).

[8] Sueyoshi, T., "Mixed integer programming approach of extended DEA-Discriminant Analysis", ***European Journal of Operational Research***, 152: 45–55 (2004).

[9] Sueyoshi, T., "DEA-Discriminant Analysis: Methodological comparison among eight discriminant analysis approaches", ***European Journal of Operational Research***, 169: 247–272 (2006).

[10] Glen, J.J., "A comparison of standard and two-stage mathematical programming discriminant analysis method", ***European Journal of Operational Research***, 171: 496–515 (2006).

[11] Bal, H., Örkcü, H.H., Çelebioğlu S., "An experimental comparison of the new goal programming and linear programming approaches in the two-group discriminant problems", ***Computers&Industrial Engineering***, 50(3): 296–311 (2006).

[12] Bal, H., Örkcü, H.H., Çelebioğlu S., "An alternative model to Fisher linear programming approaches in two-group classification problem: Minimizing deviations from the group median", ***G.U. Journal of Science***, 19(1): 49–55 (2006).

[13] Patwo, E., Hu, M.Y., Hung, M.S., "Two-group classification using neural networks", ***Decision Sciences***, 24(4): 825–845 (1993).

[14] Holmstrom, L., Koistinen, P., Laaksonen, J., Oja, E., "Neural and statistical classifiers taxonomy and two case studies", ***IEEE Trans. Neural Networks***, 8: 5–17 (1997).

[15] Mangiameli, P., West, D., "An improved neural classification network for the two-group problem", ***Computers and Operations Research***, 26: 443–460 (1999).

[16] Pendharkar, P.C., “A threshold varying artificial neural network approach for classification and its application to bankruptcy prediction problem”, ***Computers and Operations Research***, 32: 2561–2582 (2005).

[17] Pavur, R., Loucopoulos, C., “Examining optimal criterion weights in mixed integer programming approaches to the multi group classification problem”, ***Journal of Operational Research Society***, 46: 626–640 (1995).

[18] Loucopoulos, C., Pavur, R., “Computational characteristics of a new mathematical programming model for the three-group discriminant problem”, ***Computers and Operations Research***, 2: 179–191 (1997).

[19] Gehrlein, W.V., “General mathematical programming formulations for the statistical classification problem”, ***Operations Research Letters***, 5: 299–304 (1986).

[20] Gochet, W., Stam, A., Srinivisan, V., Chen, Shaoxiang, C., “Multigroup discriminant analysis using linear programming”, ***Operations Research***, 45(2): 213–225 (1997).

[21] Östermark, R., Höglund, R., “Addressing the multigroup discriminant problem using multivariate statistics and mathematical programming”, ***European Journal of Operational Research***, 108: 224–237 (1998).

[22] Patterson, D.W., Artificial Neural Networks, ***Prentice Hall***, Singapore, (1996).

[23] Sexton, S.R., Dorsey, R.E., “Reliable classification using neural networks: a genetic algorithm and backpropagation comparison”, ***Decision Support Systems***, 30: 11–22 (2000).

[24] Rumelhart, D.E., Hinton, G., Williams, R., “Learning representation by back-propagation errors”, ***Nature***, 323(9): 533-536 (1986).