

## Providing a Hybrid Model Based on Genetic Algorithm and Artificial Neural Network to Measure Credit Risk of Tejarat Bank's Real Customers (Branches of Tehran Province)

<sup>1</sup>Omid Aghamoradi Bisetooni

<sup>1</sup> Islamic Azad University Arak Branch, Faculty of Engineering, Iran

Received: 29<sup>th</sup> June 2017, Accepted: 26<sup>th</sup> July 2017, Published: 1<sup>st</sup> September 2017

### Abstract

It has been initially mentioned in this research that because of increasing diversity of credit institutions and increasing fields of activity and their relationship with people, these institutions require updated and useful methods to manage their credit risk in a scientific method. Research questions include how to measure credit risk of the customers and methods of credit risk measurement. General concepts on credit risk and methods of credit rating have been expressed and have been selected by using genetic algorithm of existing, limited data, and 11 ultimate features. Then the results have been expressed in tables and diagrams in the results section by putting the data in two Mamdani and Sugeno neuro-fuzzy networks, and training network in Matlab software. These results indicate that firstly, rate of interest and facilities has the greatest effect on customers' credit risk measurement and whatever the number of features is fewer, network accuracy is higher. They also indicate that Sugeno neuro-fuzzy network has more accuracy despite closeness of the results of the two networks' analysis.

**Key words:** risk, credit risk measurement, neuro-fuzzy network, Mamdani structure, Sugeno structure

### Introduction

Today credit industry plays an important role in the countries' economy. Increase of credit demand has created more competition for the creditor institutions. Banks are considered as one of the fundamental pillars of this system. When a bank can develop a desirable system to evaluate and assess specific credits of each customer can be considered successful (Gupta, 2000). Parametric, nonparametric, and data mining methods can be mentioned among credit rating methods (Alborzi, Pourzandi, and Khanbabaee, 2010).

Fuzzy model, support vector, and hybrid model of neural networks and genetic algorithm can be also mentioned among hybrid methods (Janu et al, 2011; Kuzni, 2015). Each of these methods has advantages and disadvantages compared to each other. For example, the old methods such as statistical logistic regression are not suitable for large and small inputs. This assumption is often considered that the relationship between variables is linear, while

Nonlinear relationship is practically between variables in most cases (Khales and Shekuhi, 2010). Neuro-fuzzy networks method has been used in this research to examine all aspects of the existing methods. Simplicity and intelligibility of fuzzy systems are great advantages of them that make them very simple to use and understand. Approximate inference capability and adjustable accuracy in dealing with the problems are considered of other their advantages. But these systems do not have learning capability, so they encounter with problem when facing with complex systems (Zhang et al., 1997).

The aim of conducting this research is to provide an appropriate model for measuring credit risk of customers as well as identifying and assorting the credit characteristics of customers by using genetic algorithm.

### Research Background

Risk management in banks has higher sensitivity and complexity compared to other economic units (Mahdavi Najmabadi, 2002). So that paying no attention to the concepts of risk and returns in the modern world of modern banking can be led to irrecoverable risks for banks. The subject of this research among the different types of risks in the banking industry is to evaluate credit risk. According to the definition provided by Switzerland Basel Committee: "Credit risk is the potential possibility that the borrower from bank and / or on behalf of his account becomes incapable in the implementation of his obligations to the bank within a specific period." crises observed in the country banking system had been mainly because of inefficiency and credit risk management. Credit risk management is a process in which losses resulting from loan default during business cycles are calculated quantitatively and includes third steps: calculation of likelihood of customers' loans default by using quantitative models; calculation of losses resulting from default by using statistical models; and evaluation of the model by using data from the past.

In fact credit scoring is a statistical model that the creditor as the result of its implementation will be able to determine the risk of granting facilities to credit applicant based on his features. This technique is considered one of the most reliable and most

consistent techniques in the field of risk assessment of potential or existing customers. In recent decades, credit scoring techniques has been changed to one of the most popular modern techniques. In general, the advantages of using credit rating models can be summarized in six following cases:

1. Reduction of evaluation time
2. Reduction of evaluation cost
3. Ability to trust and rely on credit decisions
4. Reduction of non-performing loans
5. Expansion of financial services
6. Expansion of spreading credit culture at macro level

The following methods are used to solve the problem of credit rating:

1. Linear probability model
2. Decision Trees
3. Genetic algorithm
4. Certified Systems
5. Support Vector Machine
6. Discriminant analysis
7. nearest neighbor
8. Logistic regression
9. Neural network
10. Neuro-fuzzy networks

Continuing to deal to domestic and foreign researches conducted in the field of credit rating:

Boggs (1967) tried for the first time to improve accurate credit scoring models by using complex statistical multivariable tool.

Elmer and Brufski (1988) used Multilayer Perceptron Neural Networks Model to predict the ability to repay loans.

Jensen (1992) used neural network for credit scoring. The accuracy of this network varied between 76 to 80 percent despite being small statistical population (125 loan applicants) of the research.

Trnkil in 2006, in his PhD thesis compared the power of artificial neural network and traditional statistical models in scoring. He had two hypotheses: first, the power of artificial neural networks' classification is higher than traditional statistical methods, and second, credit scoring will have different results by changing the methods of interpreting the neural network weights. He proved the results obtained from both hypotheses.

**We have briefly expressed some of the domestic conducted researches in following Table:**

Researcher	Topic	Year of research
Mansouri	Designing and explaining mathematical model of banking facilities with classic models and neural network approach	2004
Arab Mazad and Roeintan	Factors affecting the credit risk of banking customers, the Agricultural Bank Case Study	2005
Faghih	Credit rating model design for legal customers of commercial banks by using statistical method and logistic regression and multidimensional discriminant analysis,	2005
Ansari	The use of neural network to validate the Parsian bank customers, results: variables of assets ratio to total assets, profit before taxes to total debt, sales to total debt and total debt to equity have more importance and higher weight	2009

**Methodology**

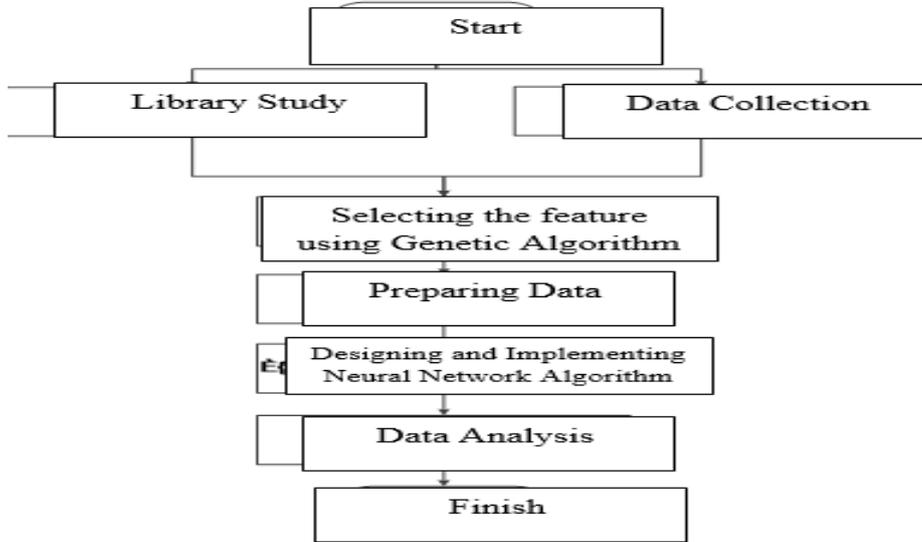
The research type is applied and is descriptive in terms of nature and method. The research statistical population is all of the real customers of the bank between 2012 and 2014, who have received more than 700 million Rials facilities from the bank. We have only chosen the contracts concluded in Tehran province to limit the statistical population. Accordingly, the total number of contracts is 4327 cases. Age, gender, credit history, type of activity, facilities interest rate, type of bond, the loan amount, type of facilities, number of installments, repayment duration, and type of installment were considered as

model inputs based on the previous conducted studies and the existing information.

Chromosomes with binary genes have been intended in this research to select GA feature. The length of each chromosome (d) is equal to the number of features. Zero number represents feature removal and 1 represents the existence of feature in the assortment. Each of chromosomes has been considered as a transformation matrix of  $H_i (i = 1, 2, \dots, p)$ . A new subset of features has been formed by multiplying input matrix in each of these chromosomes.

The base of feature selection in this research is on accuracy of the sorter that has shown the power of selected features in data separation.

This research steps have been also shown in Figure 1



**Results and Discussion**

**Data processing**

The dataset used in this study had been related to bank customers' credit information. 20 input variables, including current account status, account duration, credit history, purpose of loan request, debt status of credit applicant to other banks, monthly income of credit customer, credit amount, savings account status, occupational status, percentage rate of installments, personal profile and gender, guarantor status, residence duration at the current location, assets, age, number of installments,

installment type, number of credits in this bank, occupational status, repayment duration have been firstly considered for each banking facilities applicant.

Some of these variables have declined according to the information database of that bank and lack of access to accurate and credible information. Final number of the required features has been obtained experimentally through genetic algorithm method in the final step.

**Table 4.1: Variables examined in research background**

Row	Variable name	Reference
1	Age, credit history	Rajabzadeh ghotri and et al (2009)
2	Age, gender, type of activity	Keshavarz Haddad and Ayati Gazar (2007)
3	Age, gender, loan interest rate, credit history, collateral type, loan amount	Sharifi Ranani and et al (2011)
4	Type of activity, credit history, collateral type, loan amount	Ekrami and Ski Rahnama (2009)
5	Type of collateral, the loan amount	Rahmani and Esmaeili (2010)
6	Type of facilities, number of installments, repayment duration, type of split	Bank credit experts

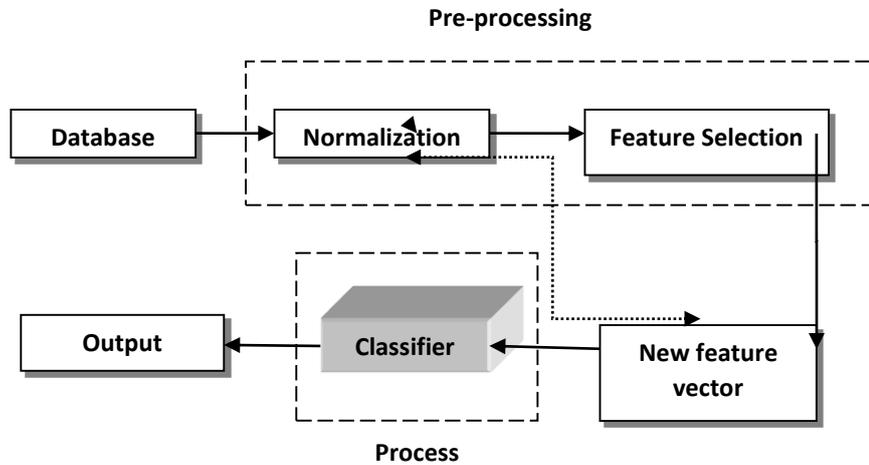
The used variables have been obtained by gathering the above variables

**Table 4.2: Variables examined in research**

Row	Variable name
1	Age, gender, credit history, type of activity, loan interest rates, collateral type, loan amount, Type of facilities, number of installments, repayment duration, Type of split

As seen in the figure below, first step is pre-processing and second step is processing. A series of measures have been performed on the real data in the first step to prepare data for assortment. These measures include placement of non-existent data, normalization, and feature selection. Designing neuro-fuzzy sorter is in the second step.

Figure 4.1: Overall diagram of assortment algorithm



The results of algorithm simulation have been examined in this section. It has been tried in the simulation to bring the best result of assortment of training and testing with fewer errors as a result of the simulation in this chapter by changing algorithm rate and the number of repetition training. Data used for simulation has two classes (default and non-default). The collected data that included both classes have 1574 samples and 11 features. We have considered features of this data as  $x_1, x_2, \dots, x_{11}$  according to Table (4.3).

Table 4.3: Data features name

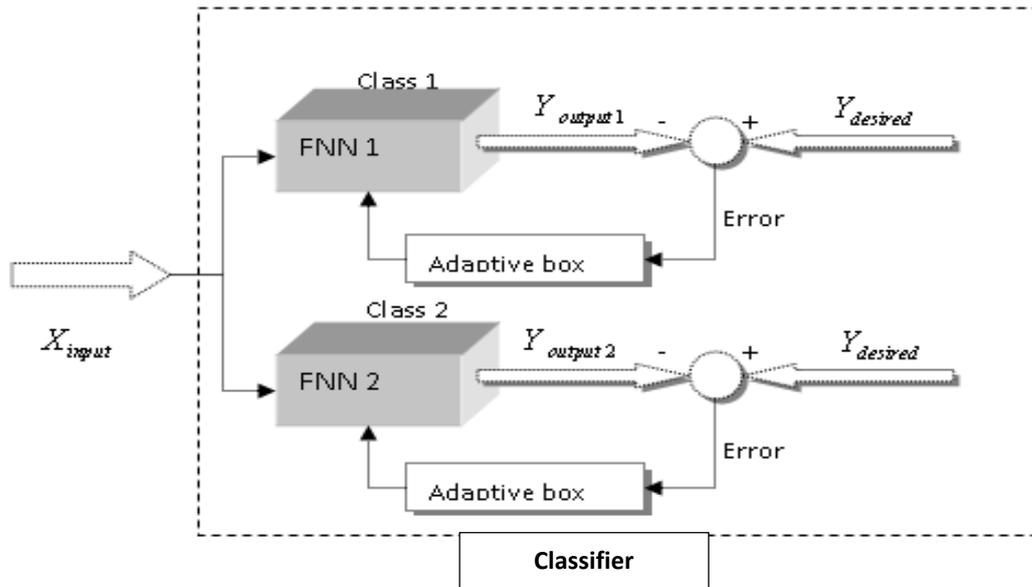
Row	Feature name
$x_1$	type of activity
$x_2$	Credit history
$x_3$	Type of facilities
$x_4$	Facilities amount
$x_5$	interest rate
$x_6$	Number of installments
$x_7$	Bail
$x_8$	Type of split
$x_9$	Duration of repayment
$x_{10}$	Age
$x_{11}$	Sex

Here 90% of total data has been used for fuzzy neural network training and the other 10% for network testing.

**Sorter**

Neuro-fuzzy sorter is the main part of the suggested algorithm. A series of considerations that should be considered in designing the sorter and during simulation has been dealt in this section. Since the used data have two classes, therefore a neuro-fuzzy network has been designed for each class. Its diagram has been shown in the following figure.

Figure 4.2: Overall diagram of sorter



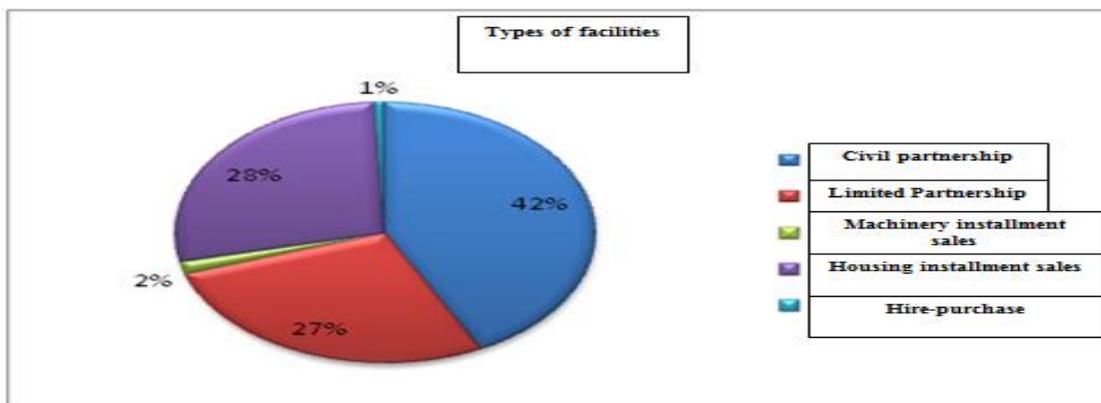
**Research data**

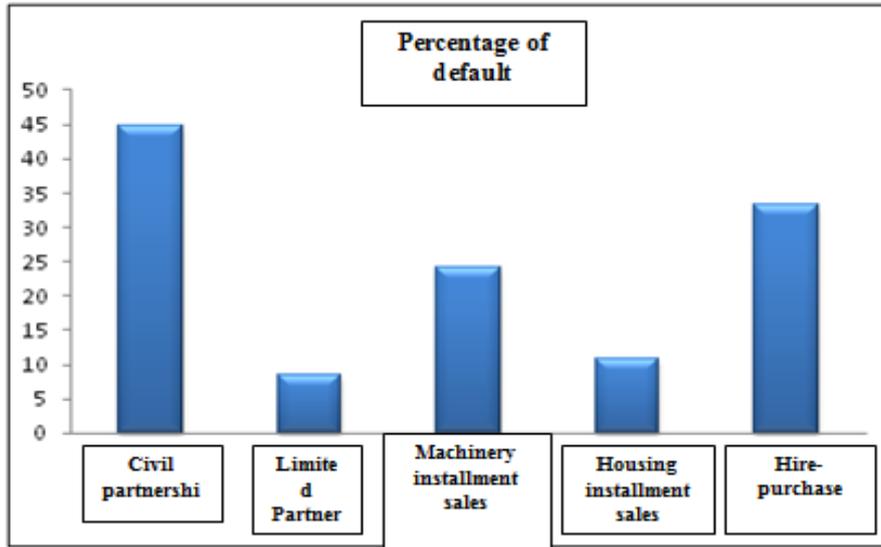
Bank has divided type of the contract concluded with the customer contract to be able to give several kinds of loans to the people as follows: interest-free loan, civic participation, limited partnership, self, machinery installment sale, housing installment sale, raw materials installment sale, hire-purchase, reward.

Contract of an individual among those who have received loans had been as interest-free loan and one as self and contracts of three people as raw materials installment sale in this research data. They have paid loan on time. Four people had been as reward that two people have not paid on time. They have been ignored in the calculations and drawing tables because they are little compared to the overall sample.

Table 4.4: Research data summary

Types of facilities	civil partnership	Limited Partnership	Machinery installment sales	Housing installment sales	Hire-purchase	Presented
Total number	665	429	25	434	12	4
Number of default	297	36	6	47	4	2
Amount (million rials)	788507	701973	58555	326611	59733	57199
Percentage of default	44.66%	8.39%	24%	10.83%	33.33%	50%





It can be concluded according to the obtained information that the likelihood of facilities repayment increase significantly if the bank registers contract with applicants as limited partnership.

**Applying Mamdani and Sugeno neuro-fuzzy network**

given that these two structures of neuro-fuzzy and feature selection algorithm have been used, this network simulation in several steps and the results all together have been shown in table of each subdivision.

**Neuro-fuzzy network with GA feature selection algorithm**

A genetic algorithm has been used here for feature selection. In this manner the best subset of features that has higher accuracy and least number of features has been selected in any kind of data by placing a series of restrictions on the feature selection algorithm. The number of steps of features selection step is 10 steps. 40 chromosomes have been produced in each step by the selected chromosome mutation and cut-off given that the initial population of chromosomes is 20 chromosomes.

It should be noted that the length of the chromosomes to be equal to the number of features. A subset of features has been selected in the feature selection step after applying each chromosome to features vector. 50 courses with selected features have been trained to evaluate the accuracy of the chromosome in selecting features, network. The number of training repetition after selecting features is 200 courses. But number of training repetition at the feature selection step is 200 courses. Network structure also varies according to the number of inputs.

Feature selection algorithm has been firstly asked to select the best subset without limitation for feature selection. Features of  $x_3, x_5$  were selected among all the features. It has been tried to obtain the best binary, ternary and so on subsets of features for each structure for further evaluation and effect of an increase in the number of features. Then the result to be recorded by network training and testing. The number of training course after feature selection is 200 courses. The results of data applied to the neuro-fuzzy network algorithm with Mamdani and Sugeno inference engine along with genetic feature selection algorithm have been shown in Tables (4.5) and (4.6), respectively.

**Table 4.5: Results of applying neuro-fuzzy network algorithm with Mamdani and Sugeno inference engine and genetic algorithm for feature selection**

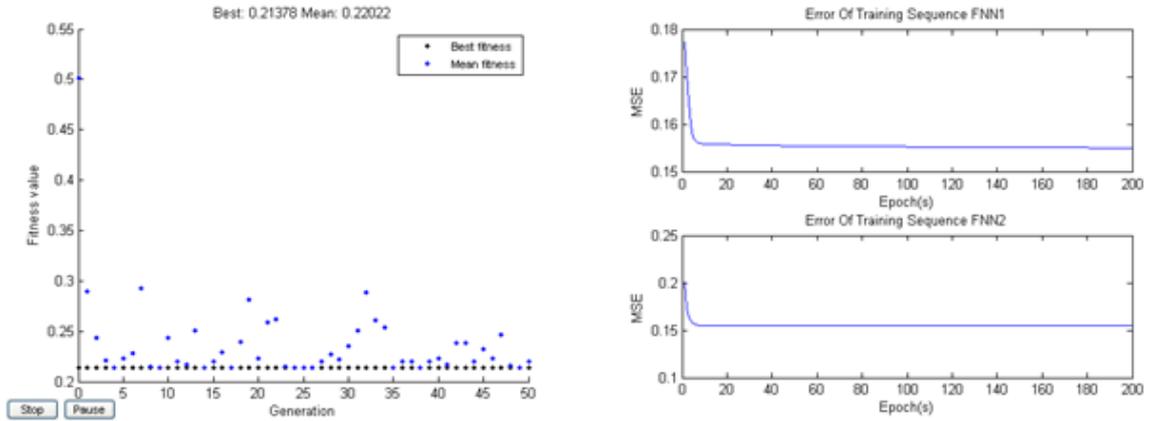
Number of entry	Selected feature	Precision of Training	Precision of test	Threshold 1	Threshold 2
2	$x_3, x_5$	78.6217	76.9737	0.4289	0.6038
3	$x_1, x_3, x_5$	78.6217	76.9737	0.4289	0.6038
4	$x_2, x_3, x_4, x_5$	78.8326	76.3158	0.4154	0.5828
5	$x_1, x_3, x_5, x_6, x_9$	76.6526	75.6579	0.6164	0.3759
11	All features	74.7352	74.2792	0.6671	0.3412

We can found according to the values of above table that increase in selected features causes a little change in values of training and testing accuracy. But increase in features enhances the complexity of network structure and prolongs processing time. So the process of increasing the number of features is continued to 5 features. It

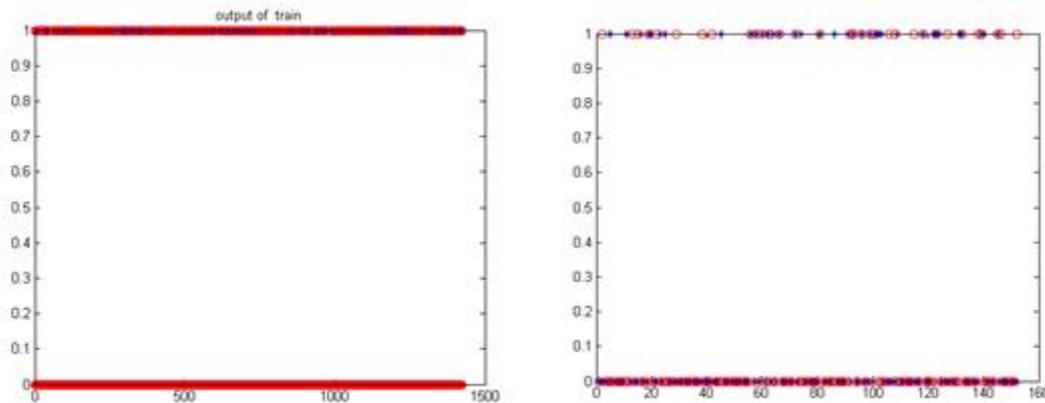
can be seen among the selected features that facilities type and interest rate features have the most repetition frequency. This indicates that these two features have great importance and other features have been added to these two features by increasing the number of features. Diagrams related to ternary and quintuple feature selection with Mamdani structure have been shown below

**Figure 4.3:** Diagrams of training error and thresholding and diagrams of test output and network training that to be related to neuro-fuzzy network with Mamdani structure with three features have been shown in Figures A to D, respectively

**A. Diagram of MSE error reduction during neuro-fuzzy network training with two Mamdani structures**



**B. Diagram of assortment error reduction during determining threshold in genetic algorithm**



C. Desired output diagram (in blue) and neuro-fuzzy network output (in red) with test data  
 D. Desired output diagram (in blue) and neuro-fuzzy network output (in red) with training data

**Table 4.6: Results of applying data to neuro-fuzzy network algorithm with Sugeno inference engine and genetic algorithm**

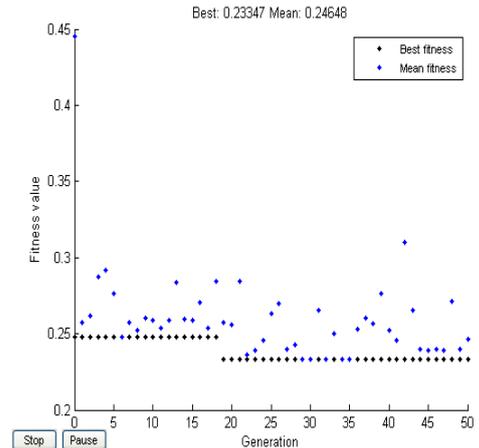
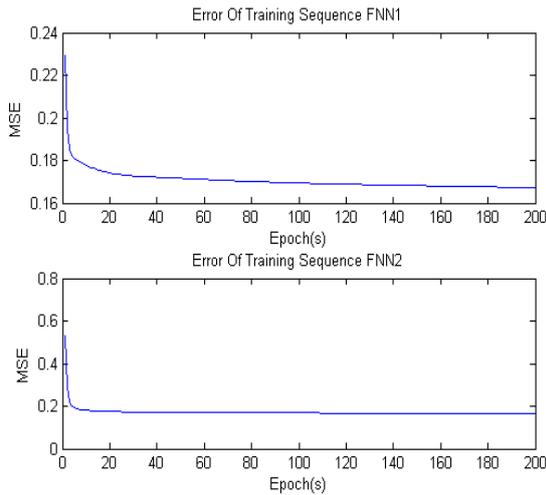
Hm for feature selection					
Number of entry	Selected feature	Precision of Training	Precision of test	Threshold 1	Threshold 2
2	$x_3, x_5$	78.7623	76.9737	0.4289	0.5163
3	$x_2, x_3, x_5$	78.6217	75.6579	0.4044	0.6029
4	$x_2, x_3, x_5, x_7$	78.9030	76.3158	0.5221	0.5131
5	$x_2, x_3, x_5, x_6, x_7$	79.1842	76.3158	0.5312	0.4689
11	All features	78.8351	76.4247	0.4821	0.4256

We can here also see that increase in selected features causes a little change in values of accuracy in addition to enhance the complexity of network structure and prolong processing time according to

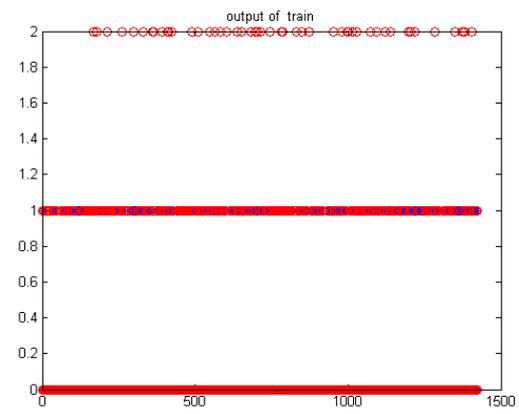
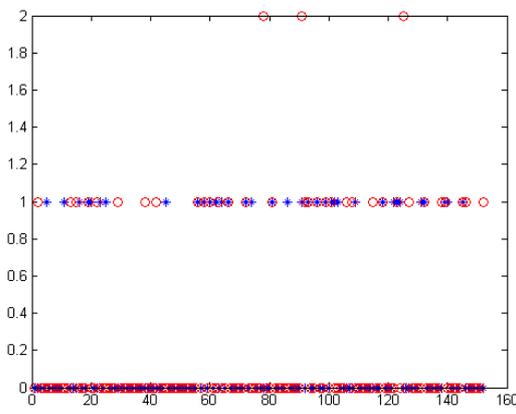
the values of above table. So, it can be seen among the selected features that facilities type and interest rate features have the most repetition frequency like as Mamdani structure. This indicates that these two

**Figure 4.6: Diagrams of training error and threshold and diagrams of test output and network training that are related to neuro-fuzzy network with Sugeno structure with five features have been shown in Figures A to D, respectively**

- A. Diagram of MSE error reduction during neuro-fuzzy network training with two Mamdani structures**
- B. Diagram of assortment error reduction during determining threshold in genetic algorithm**



- C. Desired output diagram (in blue) and neuro-fuzzy network output (in red) with test data**
- D. Desired output diagram (in blue) and neuro-fuzzy network output (in red) with training data**



It can be seen by comparing diagrams of Mamdani and Sugeno structures testing and training that a number of samples in the diagrams related to Sugeno structure has been assorted in a class that is not of two default and non-default classes. This indicates the concurrency of Sugeno structure in determining input samples' class. Since wrong assortment of default class in non-default class and vice versa will cause bank officials' mistakes in decision-making, so none-assortment of these samples will be better than their assortment. These are of Sugeno structure's strengths. It can be seen by comparing values of accuracy of both neuro-fuzzy network structures testing and training that their difference is negligible but Sugeno structure has been with better performance.

Precision of education and testing the network are changed after applying the data to the network according to the type and number of selected features and type of fuzzy neural structure. Sometimes adding new feature to increase the precision of classifier and in such cases also adding new feature reduces the classification precision or has the little change. Due to the data properties used is type of discrete and continuous data, the features selected, mostly are continuous.

By viewing the classification accuracy we see that in some cases the classification accuracy is very low and classification accuracy changes according to the type of feature selected and fuzzy neural structure. The reasons for the low accuracy of the classifier is that data distribution within each class is not normal and the number of samples is related to default class 75.1% of total sample and 24.9% of samples is related to other class (non-default). Here, also, given that the data properties used is the type of discrete and continuous, features selected, are often two features of type of facilities and the interest rate

**Table 4-7 best results obtained for data classification after applying genetic algorithm as selection algorithm of feature**

Type of structure	Selected properties	Precision of test	Precision of education
Mamdani	$x_3, x_5$	76.9737	78.6217
Sugeno	$x_3, x_5$	78.7623	76.9737

**Conclusions**

✓ We see when a genetic algorithm with binary chromosome is used for feature selection that this algorithm has selected the main feature and applies it to sorter network. We can found from the percentage of accuracy obtained for training and testing the network that facilities type and the interest rate features have a high frequency in terms of repetition. Since inputs are used again in the result

of Sugeno structure and its number of training parameters is more than Mamdani structure, so it has relatively better performance.

✓ Revision of conventional techniques of customers' credit assessment, which are based on subjective judgments, seems essential due to increasing expansion of variety and number of facilities provided by banks, as well as intense competition among financial institutions to attract customers. So, turning to the mentioned advantages in this study, implementation of credit rating models in banking system make faster, more accurate and with lower cost decision-making possible and meet also increasing trend of credit demand.

✓ Neural networks do not require any distribution assumptions for input data. For this reason they can be used for a wider range of problems compared to statistical techniques and regression and discriminant analysis.

✓ Neural networks have this property to allow new data to be added to training neural network, and to update the previous trained resources, while new and old data in statistical techniques should be entered together into a single class to yield new results.

✓ These networks are an estimator without the need for model. This feature allows them to be able to record data interactive effect on each other without need to formulate an explicit model

**Reference**

- Gupta tabucanon mt (200) fuzzy multiple criteria decision making
- Bill Russell, Jackson, Tom, Albers, Mahmoud (Translator) (17 December, 2010) Introduction to Neural Networks Sharif University of Technology, Academic Press
- Ekrami, Mahmoud, ski Rahnama, A. (2009) "Assessing Factors Affecting the past and the outstanding bank debt maturing" for a bank, Number Six
- Ansari, Sara. (2009). Legal Customer Credit Bank Corp using artificial neural network. Master's thesis.
- Rajabzadeh, Ali, Mirzai, Arash and Prvyzahmdy (2009) "intelligent system design combines customer credit rating bank using a combination of fuzzy reasoning models, Journal of Business Research, Issue 53
- Rahmani, Ali, Esmail, Gharibeh, performance neural networks, logistic regression, and discriminant analysis in predicting default, Journal of Quantitative Economics, Volume 7, Number 4, Winter 2010
- Sharifi Renani , Hossein, Ranjbar, Homayoun and Ebrahim Fuladi (2011) "Review of factors contributing to the failure to repay facilitate the

farmers' agricultural economics and development,  
Vol XIX, No. 74

8. Arabmazar, Abbas and rointan, Puneh. (2006).  
Credit risk banking customers Bank study  
Kshavrzy.dv Journal - Research Economic Essays
9. Lee, Mansoori, Adel Azar, designed to explain the  
efficient allocation of bank loans - neural network  
approach, logistic regression Vkhity Two-Journal -  
Research Lecturer, Volume 6, Number 3. Autumn  
2002
10. Faghih, M. (2004) "credit risk management and  
its policies", Journal of banks and the economy,  
Issue 46.
11. Kshavrz hdad, GR, Ayati Gazar, Hossein  
(Winter, 2007) compared the function of the logit  
model and methods of classification and regression  
trees in the validation process of genuine applicants  
for the use of bank loans, Journal of Economic  
Research, Year Issue IV
12. Mahdavi Najmabadi, Seyyed Hossein, (2002)  
"differences in the distribution of risk in Islamic and  
traditional banking system", Proceedings of the  
Conference on Islamic Banking, Tehran: Iranian  
banking institution.
13. Elmer, Peter J., and David M. Borowski (1988).  
an Expert System and Neural Networks Approach  
To Financial Analysis, Financial Management
14. Jenson. Herbert, 1992; "Using Neural Network  
for Credit Scoring", Managerial Finance, 18 (6), pp.  
15-26
15. J.S.R.Jang, C.T.Sun, and E. Mizutani, "Neuro-  
Fuzzy and Soft Computing," Prentice-Hall Inc.  
1997.
16. Trinkle. B. S, 2006; Interpretable Credit Model  
Development via Artificial Neural Network, PhD  
thesis, University of Alaba