International Conference on Fuzzy Sets and Soft Computing in Economics and Finance
FSSCEF 2004

Proceedings

Volume I

Saint-Petersburg, Russia
June 17-20, 2004

ISBN 968-489-028-1
968-489-029-X (Printed version)
968-489-030-3 (Electronic version)

Instituto Mexicano del Petróleo
Eje Central Lázaro Cárdenas No. 152
Col. San Bartolo Atepehuacan
C.P. 07730 México D.F.

Russian Fuzzy Systems Association
Vavilova 40, GSP-1, Moscow 119991
Russia

http://sedok.narod.ru/fsef.html

Editors :
Ildar Batyrshin
Janusz Kacprzyk
Leonid Sheremetov

Design and Format:
Edgar J. Larios

Edited and Formatted in Mexico
Printed in Russia
(500)

Tuning the Fuzzy Classification Models with Various Learning Criteria: the Case of Credit Data Classification

Serhiy SHTOVBA, Olga PANKEVICH, and Georgios DOUNIAS

Department of Computer Based Information & Management Systems,
Vinnitsa National Technical University, Vinnitsa, Ukraine
shtovba@ksu.vstu.vinnica.ua

Department of Financial and Management Engineering,
University of the Aegean, Chios, Greece, *g.dounias@aegean.gr*

Abstract. In this paper we study the efficiency of various learning criteria for the proper tuning of a fuzzy classifier. Different cases of crisp and noisy class borders are considered, and a specific credit-risk application is discussed.

1 Introduction

Fuzzy rule–based systems are powerful tools which perform adequately in classification tasks related to various financial and economic decision-making problems, such as customer segmentation, credit-risk prediction, project evaluation, fraud detection, etc. [3, 4]. Fuzzy classifiers usually provide a good balance between decision accuracy and model transparency. In this paper we study the efficiency of various learning criteria for the tuning process of a fuzzy classifier. Tuning corresponds to the search-process of weights of fuzzy if-then rules and parameters of the membership functions that minimize the difference between actual and inferred decisions. Cases of crisp and noisy class-separating curves with credit-risk assessing application are considered.

2 The Fuzzy Rule – Based Classifier

Let us consider a classifying system with n-inputs $(x_1, x_2, \ldots, x_n)$ and one output y. The classification can be considered as the mapping:

$$X = (x_1, x_2, \ldots, x_n) \rightarrow y \in \{d_1, d_2, ..., d_m\},$$

where $d_1, d_2, ..., d_m$ are decisions. The classification is performed with the aid of the following fuzzy knowledge base [1]:

$$\bigcup_{p=1}^{k_j} \left(\bigcap_{i=1}^{n} x_i = a_{i,jp} \text{ with weight } w_{jp} \right) \rightarrow y = d_j, \; j = \overline{1,m}, \tag{2.1}$$

where $\bigcap$ is logical operation AND, $\bigcup$ is logical operation OR, $a_{i,jp}$ denotes a fuzzy term for the evaluation of input x_i in rule with number jp, $w_{jp} \in [0,1]$ is a subjective degree of the expert's confidence in rule with number jp, k_j is the number of rules corresponding to decision d_j.

The membership degrees of an object $X = (x_1, x_2, \ldots, x_n)$ to decisions d_j ($j = \overline{1,m}$) are calculated as follows [1]:

$$\mu_{d_j}(X) = \bigvee_{p=\overline{1,k_j}} w_{jp} \cdot \bigwedge_{i=\overline{1,n}} \left[\mu_{jp}(x_i)\right], \quad j = \overline{1,m}, \tag{2.2}$$

where $\mu_{jp}(x_i)$ denotes the membership function of the fuzzy term $a_{i,jp}$, and $\vee(\wedge)$ is the max (min) operation. The decision with the maximal fulfillment degree corresponds to object X:

$$y = \underset{\{d_1, d_2, ..., d_m\}}{\arg} \max \left(\mu_{d_1}(X), \mu_{d_2}(X), ..., \mu_{d_m}(X)\right). \tag{2.3}$$

3 Learning Criteria for Tuning the Fuzzy Classifier

Let us denote a fuzzy classifier by

$$y = F(X, P, W) \tag{3.1}$$

where X is an input vector, P is a vector of the membership functions' parameters in the knowledge base (2.1), W is a vector of the rule-weights in (2.1), and F is an input-output operator corresponding to (2.2) - (2.3).

We denote the training set by

$$(X_r, y_r), \; r = \overline{1,M}, \tag{3.2}$$

where $X_r = (x_{r,1}, x_{r,2}, ..., x_{r,n})$ and y_r are the input vector and its corresponding output for the input-output pair with number r.

The tuning corresponds to searching a vector (P, W) that minimizes the difference between actual (3.2) and inferred (3.1) decisions. This difference may be defined in various ways.

Criterion 1. The percentage of misclassification is widely used as a learning criterion for diverse pattern recognition tasks. For this case, the tuning is equivalent to the following minimization:

$$\frac{1}{M}\sum_{r=\overline{1,M}} \Delta_r \rightarrow \min, \quad \text{where} \quad \Delta_r = \begin{cases} 1, & y_r \neq F(X_r, P, W) \\ 0, & y_r = F(X_r, P, W) \end{cases}. \tag{3.3}$$

An advantage of this criterion is its simplicity and the clear interpretation of values. The main drawback is the optimization difficulty related with plateau-shaped objective functions. It is often very hard to guess the suitable parameters of gradient optimizing routines, for example, the change in variables for finite difference gradient calculation.

Criterion 2. Let us apply fuzzification of the output variable in the training set (3.2) as follows [1]:

$$\left.\begin{array}{ll} \tilde{y} = (1/d_1, 0/d_2, ..., 0/d_m), & \text{if} \quad y = d_1 \\ \tilde{y} = (0/d_1, 1/d_2, ..., 0/d_m), & \text{if} \quad y = d_2 \\ ... & \\ \tilde{y} = (0/d_1, 0/d_2, ..., 1/d_m), & \text{if} \quad y = d_m \end{array}\right\}. \tag{3.4}$$

The desirable values of the inferred membership grades (2.2) are equal to (3.4). Hence, the tuning problem may be now formulated as in the following minimization [1]:

$$\frac{1}{M}\cdot\sum_{r=1}^{M}\sum_{j=1}^{m}\left[\mu_{d_j}(y^r) - \mu_{d_j}(X^r, P, W)\right]^2 \rightarrow \min, \tag{3.5}$$

where $\mu_{d_j}(y^r)$ denotes the desirable membership degree according to (3.4), and $\mu_{d_j}(X^r, P, W)$ denotes the inferred membership degree of object X^r by formula (2.2).

The objective function in (3.5) does not have large plateaus, allowing the use of gradient-based optimization methods. However, the optimal vector for (3.5), sometimes does not obtain a minimal misclassification level as well, due to the presence of objects, laying close to the class-separating curves, which almost equally contribute to correct and error classification.

Criterion 3. A combination of the advantages of the abovementioned criteria is proposed below. The main idea is to increase the contributions into (3.5), for the misclassified objects. The fuzzy classifier tuning process, is now formulated according to the following minimization problem:

$$\frac{1}{M}\cdot\sum_{r=1}^{M}\left((\Delta_r\cdot R+1)\cdot\sum_{j=1}^{m}\left[\mu_{d_j}(y^r)-\mu_{d_j}(X^r,P,W)\right]^2\right)\to \min, \qquad (3.6)$$

where R>0 is a penalty value.

Let us refer to (3.3) as criterion I, (3.5) as criterion II, and (3.6) as criterion III. The efficiency of the described learning criteria is studied below.

4 Experiment for the Case of Crisp Separating Curves

We consider a classification task with 2 inputs ($x_1, x_2 \in [0,1]$) and 3 decisions (d_1, d_2, d_3). Fig. 1 shows the data sets and class separating curves. The training set consists of 80 objects and the test set consists of 5000 objects. The input data in the sets were generated randomly.

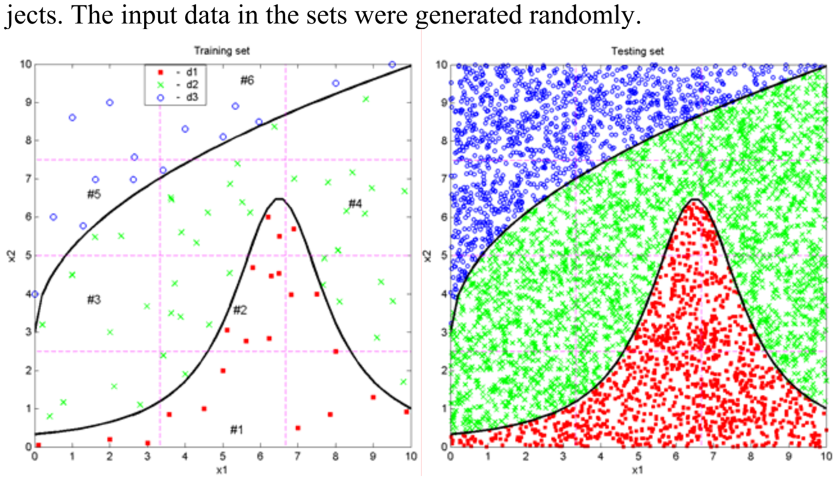
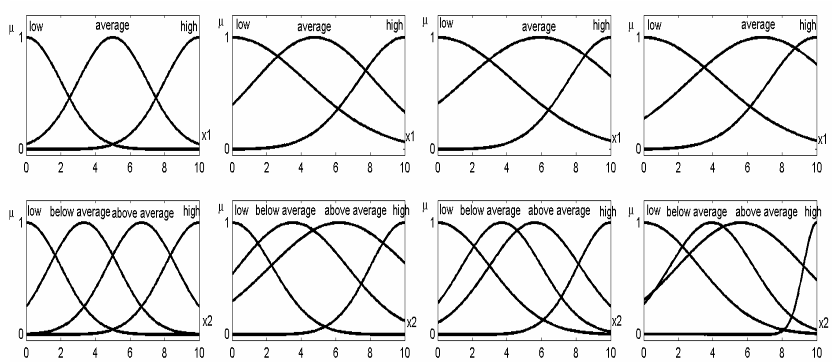


Fig. 1. Training and test sets

Table 1 shows the expert fuzzy knowledge base. The source and optimal (by various criteria) membership functions are shown on Fig. 2.

Table 1. Fuzzy knowledge base

x_1	x_2	y	w			
			source classifier	classifier I	classifier II	classifier III
average	low	d_1	1	0.62	0.75	0.71
average	below average	d_1	1	0.41	0.39	0.49
low	below average	d_2	1	0.81	1	0.90
high	higher average	d_2	1	0.46	1	0.71
low	higher average	d_3	1	0.66	0.49	0.65
average	high	d_3	1	0.02	0.02	0.91



a) source classifier b) classifier I c) classifier II d) classifier III

Fig. 2. Membership functions

Table 2 compares the results achieved by the tuning process according to various learning criteria. As an alternative classification tool we used the following decision tree (stated below as a set of equivalent decision if-then rules):

if ((x1>1.2929) & (x2≤1)) then y=d1,
if ((x1>4.6335) & (x2>1) & (x1≤7.5) & (x2≤6)) then y=d1,
if ((x2>8.3607) & (x1>5.3301)) then y=d3,
if ((x1≤5.3301) & (x2>6.9107)) then y=d3,
if ((x1≤1.2929) & (x2>3.4988) & (x2≤6.9107)) then y=d3,
otherwise, y=d2.

Table 2 shows that the usage of criterion III provides the best classification accuracy (error of 9.28%). Criterion II performs worse of all criteria. In fact, there is a small difference between the accuracy levels of Classifiers I and III but the matching of the feasible parameters for the gradient-based optimization according to criterion I, is often a time-consuming process.

Table 2. Testing the classification models

Model	Criterion I	Criterion II	Criterion III (R=9)	Misclassification on test set
Source classifier	32.5%	0.53	3.99	25.92%
Classifier I	6.25%	0.52	1.03	9.78%
Classifier II	18.75%	0.44	1.88	16.42%
Classifier III	6.25%	0.46	0.92	9.28%
Decision tree	7.5%	n/a	n/a	15.24%

5 An Experiment with Noisy Separating Curves: Credit Risk Assessment

The task of credit-risk assessment corresponds to the differential decision-making process, for the acceptance or rejection of customers' request for issuing a credit card, based on 15 customer parameters $x_1 ... x_{15}$. The data sets are available upon request from [2].

Let us create a fuzzy classifier with three inputs: x_8 - mean time of occupation at a workplace, x_{11} - years of collaboration with the bank, and x_{15} -savings account balance. Fig. 3 shows the data sets for this case.

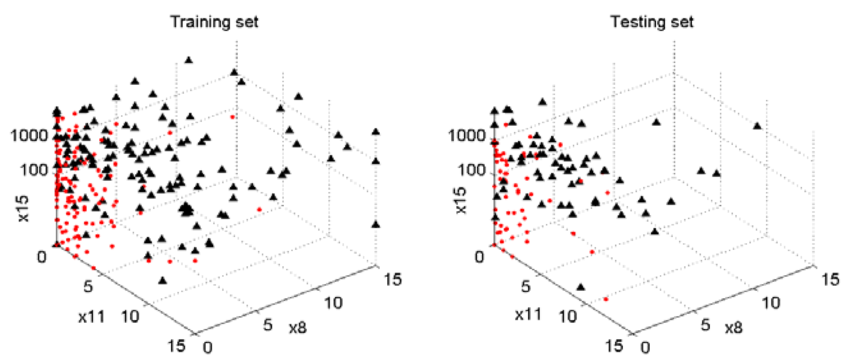


Fig. 3. Credit data sets (● - reject; ▲ - accept)

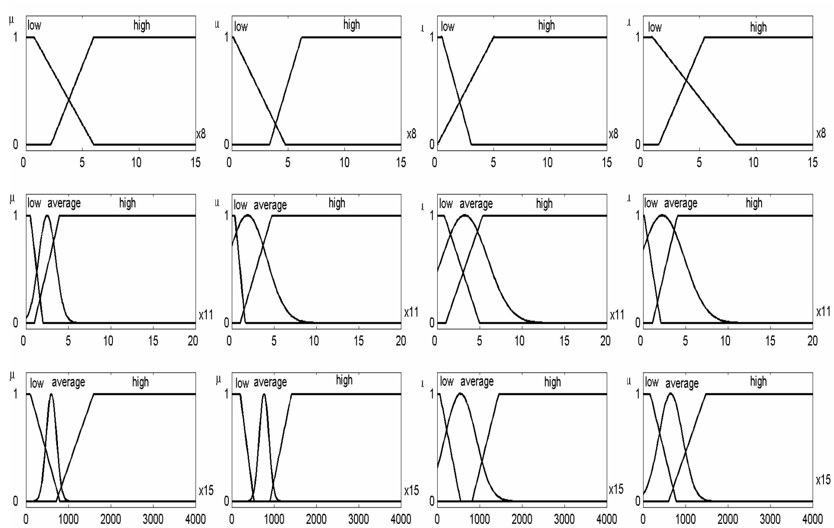
Table 3 shows the fuzzy knowledge base, created from the training data distribution with the aid of an expert. Source and optimal membership functions are shown in Fig. 4. Table 4 compares results achieved via tuning according to various learning criteria. As an alternative classification tool we used the following decision tree:

if x11>2, then y="accept";
elseif x15≤141, then y="reject";

elseif x8>1.1,then y=”accept”;
else y=”reject”.

Table 3. Fuzzy knowledge base

x_8	x_{11}	x_{15}	y	w			
				source classifier	classifier I	classifier II	classifier III
any	low	low	reject	1	1	0.63	1
low	low	any	reject	1	0.64	0.80	0.77
low	any	low	reject	1	1	0.50	0.63
any	high	any	accept	1	1	0.92	1
any	any	high	accept	1	1	0.61	1
high	average	any	accept	1	1	0.87	1
high	any	average	accept	1	0.38	0.05	0.12



a) source classifier b) classifier I c) classifier II d) classifier III

Fig. 4. Membership functions

Table 4 shows that the usage of criterion III provides the best classification accuracy (error of 21%). The difference of accuracy levels among all the classifiers is rather small, perhaps due to the noisy nature of the data.

Fig. 5 shows the minimal value of savings account balance, which provides a positive credit card issuing decision. The surface is created on the basis of fuzzy classifier III.

Table 4. Testing the classification models

Model	Crite-rion I	Crite-rion II	Crite-rion III (R=4)	Misclas-sification on test set
Source classifier	23.88%	0.46	2.34	23.5%
Classifier I	23.47%	0.43	2.16	22.0%
Classifier II	23.27%	0.35	1.49	22.5%
Classifier III	23.88%	0.42	2.00	21.0%
Decision tree	21.43%	n/a	n/a	22.5%

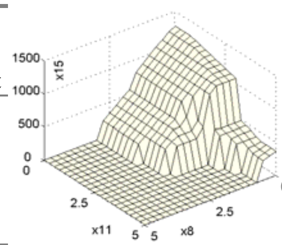


Fig. 5. Minimal value of the saving account for receiving a positive credit decision

5 Conclusion

We have studied three learning criteria for the tuning of a fuzzy classifier: 1) percentage of misclassification, 2) mean squared memberships' difference, and 3) penalized mean squared memberships' difference. The abovementioned criteria were tested in 2 classification tasks: (a) a simple classification problem with nonlinear separating curves and (b) a real credit-risk assessment problem. The experiments suggest that the third criterion provides the best classification accuracy, especially in the case of crisp separating curves. This allows us to recommend the penalized mean squared memberships' difference, as a learning criterion of choice, for the proper tuning of fuzzy rule-based classification systems.

References

1. Rotshtein A (1998) Design and Tuning of Fuzzy Rule-Based Systems for Medical Diagnosis. In: Teodorescu NH, Kandel A, Jain LC (eds) Fuzzy and Neuro-Fuzzy Systems in Medicine. CRC-Press, Boca-Raton, pp 243-289
2. Sample Data Sets for the See5-Demo Product, Release 1.09a (2000). www.rulequest.com
3. Zimmermann H-J (1996) Fuzzy Set Theory and its Applications, 3rd ed.. Kluwer Academic Publishers, Norwell, Dordrecht
4. Zopounidis C, Pardalos PM, Baourakis G (eds., 2001) Fuzzy Sets in Management, Economics and Marketing. World Scientific, Singapore