MASR - 2011

ESTIMATION OF THE EFFICIENCY OF KNOWLEDGE ACQUISITION TECHNIQUES USING CLUSTERING

ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДОВ ИЗВЛЕЧЕНИЯ ЗНАНИЙ С ПОМОЩЬЮ КЛАСТЕРИЗАЦИИ

Pēteris Grabusts

Rezekne Higher Educational Institution

Inese Polaka

Institute of Information Technology, Riga Technical University

June 29, 2011

RULE EXTRACTION METHODS

The motivation for studying rule extraction methods is the following:

- **the amount of multidimensional data to be analysed becomes too large for the potentialities of statistical analysis;**
- **popular neural network methods operate by the black box principle that complicates interpretation of the results for the user;**
- **previously unknown regularities are present in the data;**
- **the regularities found can be represented in a way that is easy to perceive and understand for the user.**

RULE EXTRACTION METHODS

The main requirement which is put forward to the results of data analysis is that the results must always be interpreted as correctly as possible. The rules that represent the regularities found have to be stated as simple and easy to understand logical expressions. Namely, they must look as these logical rules:

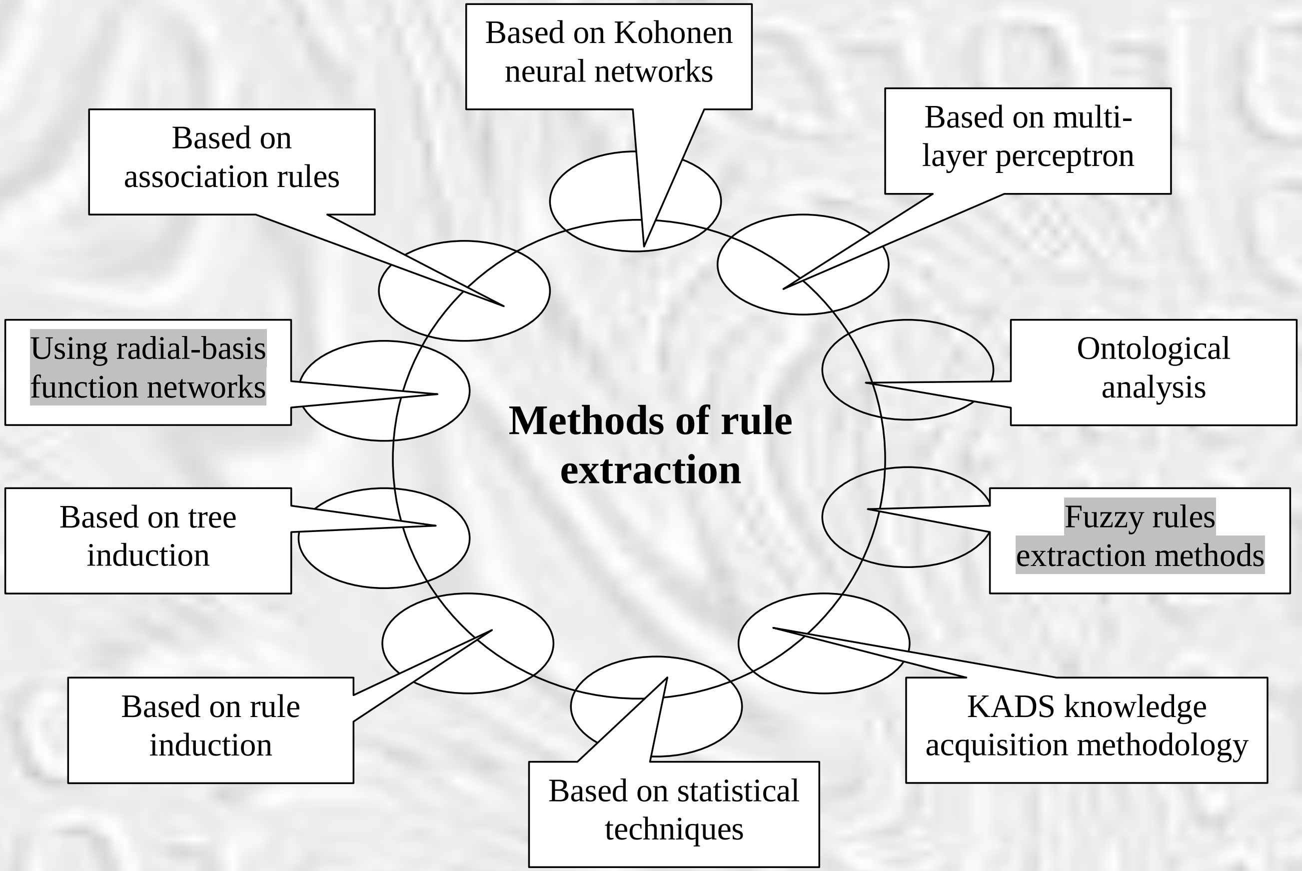
IF {(Event 1) AND (Event 2) AND ... (Event N)} THEN ...

In what follows, the author will employ logical conditional rules (production rules) of this kind:

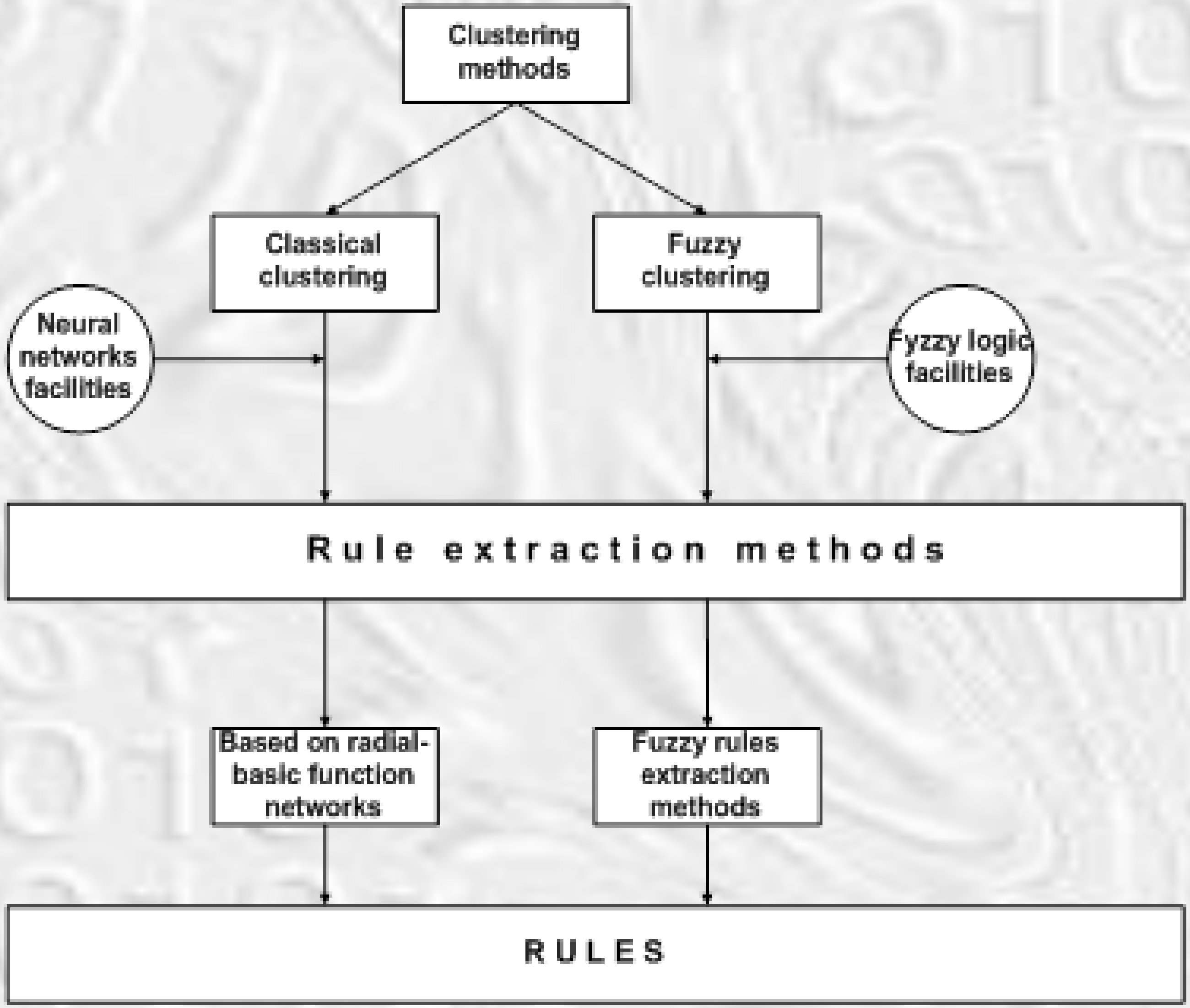
$$IF\ \underbrace{(Antecedent\,1)\ and\ (Antecedent\ 2)\ and \ldots (Antecedent\ N)}_{A}\ THEN\ \underbrace{(Consequent)}_{B}.$$

RULE EXTRACTION METHODS

- Expert Systems
- Rule-based Classifier
- Classifier based on fuzzy rules
- Knowledge-based Agent Control
- Pattern Recognition
- Rule-based Forecasting
- Rule-based Prediction
- Tool to detect Patterns in databases
- Tool to detect Regularities in databases
- Tool for extraction of useful knowledge from raw data
- Genetic-based Machine Learning



RULE EXTRACTION METHODS



- **FUZZY RULE BASE DESIGN METHOD**

- **EXTRACTING RULES FROM TRAINED RBF NEURAL NETWORK**

I. FUZZY RULE BASE DESIGN

FCM

The fuzzy c-means algorithm allows each data point to belong to a cluster to a degree specified by a membership grade, and thus each point may belong to several clusters.

- Membership matrix m_{ik} - [0,1] :

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{2/(q-1)}}$$

- Objective function is :

$$J(M, c_1, c_2, \ldots c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{k=1}^{K} m_{ik}^q d_{ik}^2$$

- Optimal center:

$$c_i = \frac{\sum_{k=1}^{K} m_{ik}^q u_k}{\sum_{k=1}^{K} m_{ik}^q}$$

FCM ALGORITHM

FCM algorithm

(1) Initialise the membership matrix M with random values between 0 and 1.
(2) Calculate cluster centres c_i (i=1,2,..., c).
(3) Compute the objective function. Stop if either it is below a certain threshold level or its improvement over the previous iteration is below a certain tolerance.
(4) Compute a new M.
(5) Go to step 2.

Fuzzy Rule Base

Fuzzy classifier

The fuzzy classifier is based on the set of final rules R for which the following holds:

R: If x_1 is $\mu^{(1)}_R$ and … and x_p is $\mu^{(p)}_R$ Then class is C_R

Distribution of rules:

$$R(x_1, \ldots, x_p) = \begin{cases} C, & if \ \mu_C^{(R)}(x_1,\ldots,x_p) > \mu_D^{(R)}(x_1,\ldots,x_p) \ \ for\, all \ D \in C,\, D \neq C \\ \notin C, & otherwise \end{cases}$$

2 D example:

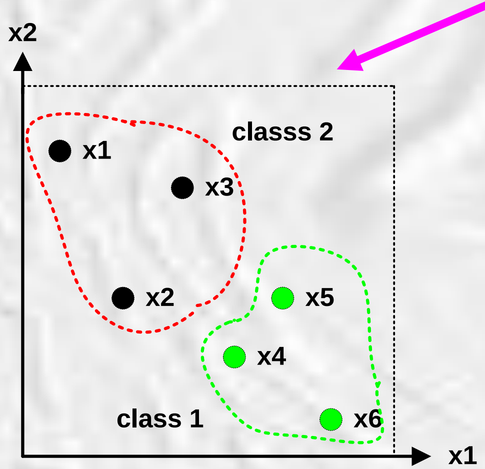
IF x is μ1 and y is v1 THEN class is A

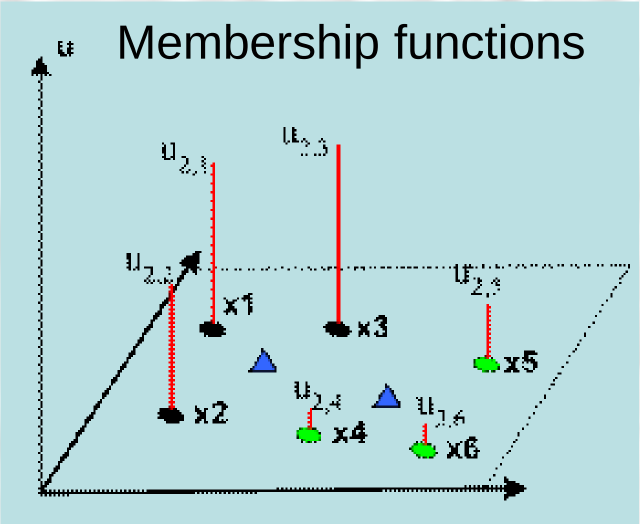
IF x is μ2 and y is v2 THEN class is B

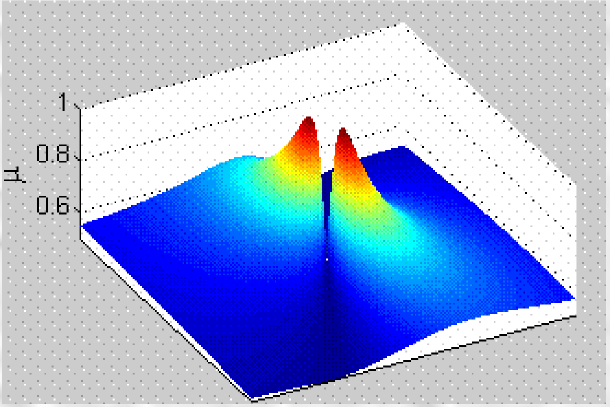
5 Stages

Example

$$X = \{(0.14, 0.85), (0.28, 0.42), (0.42, 0.71), (0.57, 0.28), (0.71, 0.57), (0.85, 0.14)\}$$





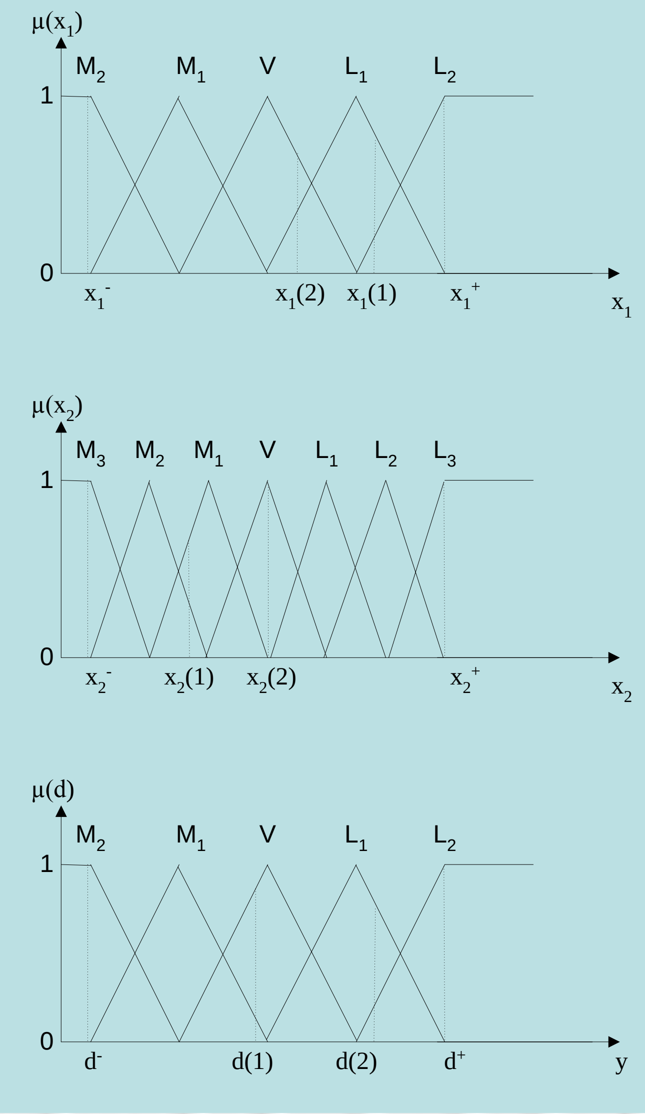


After FCM-
Centers= {(0.6910, 0.2991), (0.2991, 0.6908)}

$X_1(MF_1)$	$X_1(MF_2)$	$X_2(MF_1)$	$X_2(MF_2)$	Klases	
1.0000	0.0000	0.0000	1.0000	0	1
0.8028	0.1972	0.6056	0.3944	0	1
0.6056	0.3944	0.1972	0.8028	0	1
0.3944	0.6056	0.8028	0.1972	1	0
0.1972	0.8028	0.3944	0.6056	1	0
0.0000	1.0000	1.0000	0.0000	1	0

Stage 1. Separation of input and output data

Each of the intervals is divided into (2N+1) parts



X1 – 5 intervals (N=2)
X2 – 7 intervals (N=3)
Y - 5 intervals (N=2)

Stage 2. Construction of fuzzy rules using learning set data

For each pair of learning data a single rule can be set, for example, in this way:

$(x_1(1), x_2(1); d(1)) \rightarrow$

$\{x_1(1)[\text{max:}0.8 \text{ in domain } L_1],$

$x_2(1)[\text{max:}0.6 \text{ in domain } S_1];$

$d(1)[\text{max:}0.9 \text{ in domain } M]\}$

⇩

R^1: If (x_1 is L_1 and x_2 is S_1) Then y is M.

Stage 3. Determination of confidence degree for each rule

For the rule of the form

R: If (x_1 is A_1 and x_2 is A_2) Then (y is B)

the confidence degree will be defined as follows:

$$SP(R) = \mu_{A_1}(x_1) \cdot \mu_{A_2}(x_2) \cdot \mu_B(y)$$

For example:

$$SP(R^1) = \mu_{L_1}(x_1) \cdot \mu_{M_1}(x_2) \cdot \mu_V(y) = 0.8 * 0.6 * 0.9 = 0.432$$
$$SP(R^2) = \mu_V(x_1) \cdot \mu_V(x_2) \cdot \mu_{L_1}(y) = 0.7 * 1.0 * 0.7 = 0.49$$

As a result, not only the problem of rule contradiction would be solved but also the total number of rules would decrease essentially.

Stage 4. Formation of fuzzy rule base

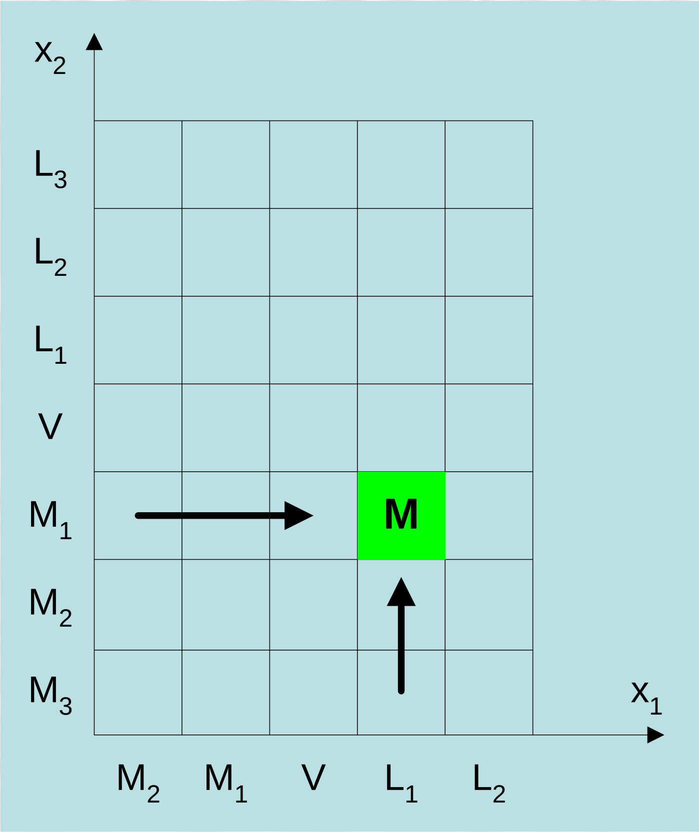
If a rule is given in the form

R^1: If (x_1 is L_1 and x_2 is S_1), Then y is M

the value of a fuzzy set that is contained in the Then part of the rule, i.e. the value M in this example, is recorded in the point of intersection of column L_1 and row S_1.

In case if various rules with the same condition exist, a rule with the highest confidence degree is selected of them.

The rule base is set in the form of a table:



Stage 5. Defuzzification

At this stage, mapping $f:(x_1,x_2)\to\bar{y}$, where $\bar{y}$ is the output value of the fuzzy system, has to be derived using the obtained knowledge base.

1. The defuzzification is considered completed if a specific value for each linguistic variable is obtained. To accomplish that, the activity degree of the k-th rule is calculated using formula:

$$\tau^{(k)} = \mu_{A_1^{(k)}}(x_1)\cdot\mu_{A_2^{(k)}}(x_2)$$

2. Defuzzification by the gravity centre method (COGS - Centre of Gravity for Singleton) can be employed to calculate the output value $\bar{y}$:

$$\bar{y} = \frac{\sum_{k=1}^{N}\tau^{(k)}\bar{y}^{(k)}}{\sum_{k=1}^{N}\tau^{(k)}}$$

Fuzzy rule base is generated

Example

Membership functions:

$X_1(MF_1)$	$X_1(MF_2)$	$X_2(MF_1)$	$X_2(MF_2)$	Klases	
1.0000	0.0000	0.0000	1.0000	0	1
0.8028	0.1972	0.6056	0.3944	0	1
0.6056	0.3944	0.1972	0.8028	0	1
0.3944	0.6056	0.8028	0.1972	1	0
0.1972	0.8028	0.3944	0.6056	1	0
0.0000	1.0000	1.0000	0.0000	1	0

Rules

Rule1: IF X1 is MF1 to degree 1 AND X2 is MF2 to degree 1 THEN CLASS is 2 to degree 0.95

Rule2: IF X1 is MF1 to degree 0.80282 AND X2 is MF1 to degree 0.60563 THEN CLASS is 2 to degree 0.4619

Rule3: IF X1 is MF2 to degree 1 AND X2 is MF1 to degree 1 THEN CLASS ir 1 to degree 0.95

Activities of rules:

X_1	X_2	Rule 1		Rule 2		Rule 3		Classes	
0.14	0.85	1	1	1	0	0	0	0	1
0.28	0.42	0.8	0.4	0.8	0.6	0.2	0.6	0	1
0.42	0.71	0.6	0.8	0.6	0.2	0.4	0.2	0	1
0.57	0.28	0.4	0.2	0.4	0.8	0.6	0.8	1	0
0.71	0.57	0.2	0.6	0.2	0.4	0.8	0.4	1	0
0.85	0.14	0	0	0	1	1	1	1	0

(0.14; 0.85)

Activity for Rule1:

$$\frac{0.14*1+0.85*1}{1+1}=0.495$$

Activity for Rule2:

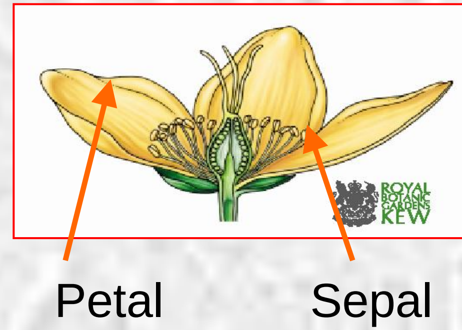
$$\frac{0.14*1+0.85*0}{1+0}=0.14$$

Activity for Rule3:

$$\frac{0.14*0+0.85*0}{0+0}=0$$

Application example: IRIS data set

setosa, versicolor un virginica.



4 parameters:
SL – *sepal length*
SW – *sepal width*
PL - *petal length*
PW – *petal width*

Setosa			
SL	SW	PL	PW
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
....	....	....	....

Versicolor			
SL	SW	PL	PW
7.0	3.2	4.7	1.4
6.4	3.2	4.5	1.5
6.9	3.1	4.9	1.5
5.5	2.3	4.0	1.3
6.5	2.8	4.6	1.5
....	....	....	....

Virginica			
SL	SW	PL	PW
6.3	3.3	6.0	2.5
5.8	2.7	5.1	1.9
7.1	3.0	5.9	2.1
6.3	2.9	5.6	1.8
6.5	3.0	5.8	2.2
....	....	....	....

Application example

The objective of the experiments was:

- To acquire rules from IRIS data base using the FCM algorithm;
- To ascertain the effect of membership function number on the count of acquired rules.
- To check the quality of the rules obtained.

First part

In the first part of the experiments:

3 membership functions were calculated for 3 clusters.

4 rules were acquired for Class 1,

3 rules – for Class 2,

11 rules for Class 3.

Rule 11: if X1 is MF2 to degree 0.94 and X2 is MF2 to degree 0.75 and X3 is MF2 to degree 0.97 and X4 is MF2 to degree 0.88 then Class is 3 to degree 0.57

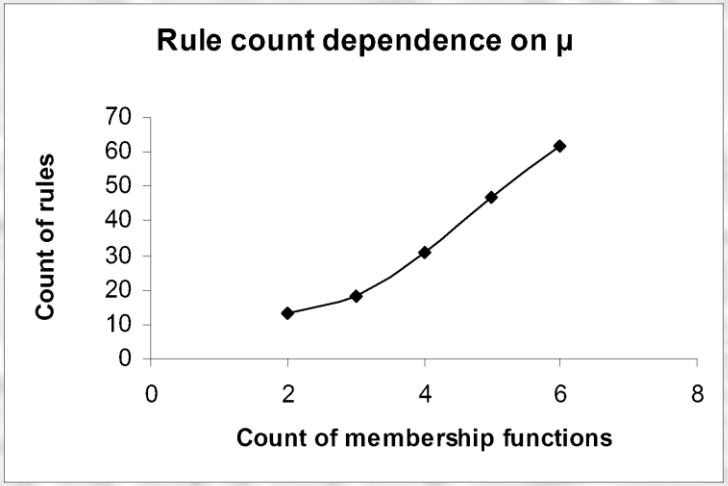
Rule 12: if X1 is MF1 to degree 0.83 and X2 is MF1 to degree 0.79 and X3 is MF2 to degree 0.59 and X4 is MF2 to degree 0.67 then Class is 3 to degree 0.25

Rule 13: if X1 is MF2 to degree 0.94 and X2 is MF1 to degree 0.75 and X3 is MF2 to degree 1 and X4 is MF2 to degree 0.92 then Class is 3 to degree 0.62

Second part

Dependence of the obtained rule number on the count of initially set membership functions

Number of memberships	Class 1	Class 2	Class 3	Count
2	2	8	3	**13**
3	4	3	11	**18**
4	7	11	13	**31**
5	13	16	18	**47**
6	21	19	22	**62**



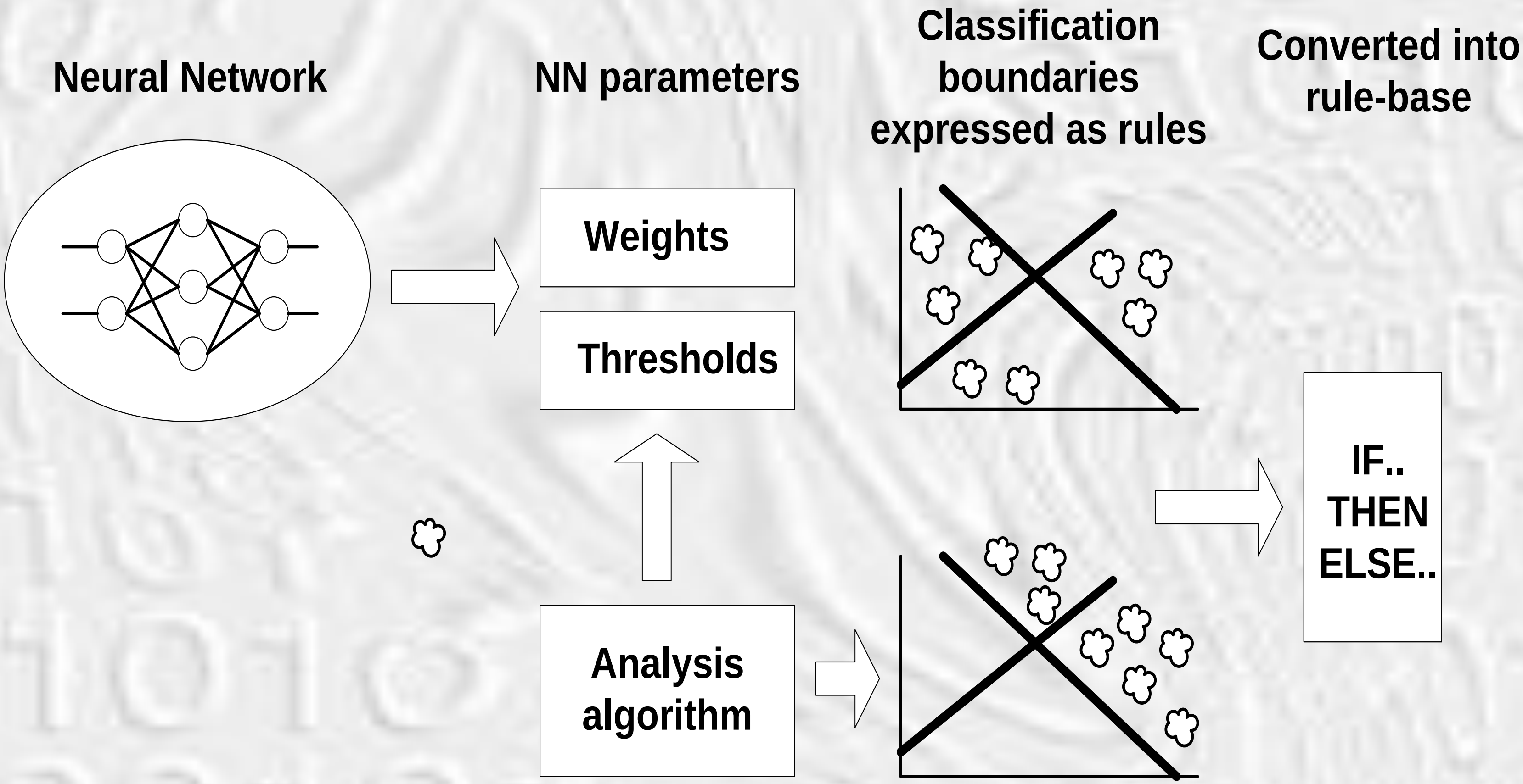
Conclusions for rule base design

Methods of fuzzy rule base design are widely used in different control processes. They, however, can be adapted to rule extraction from the numerical data.

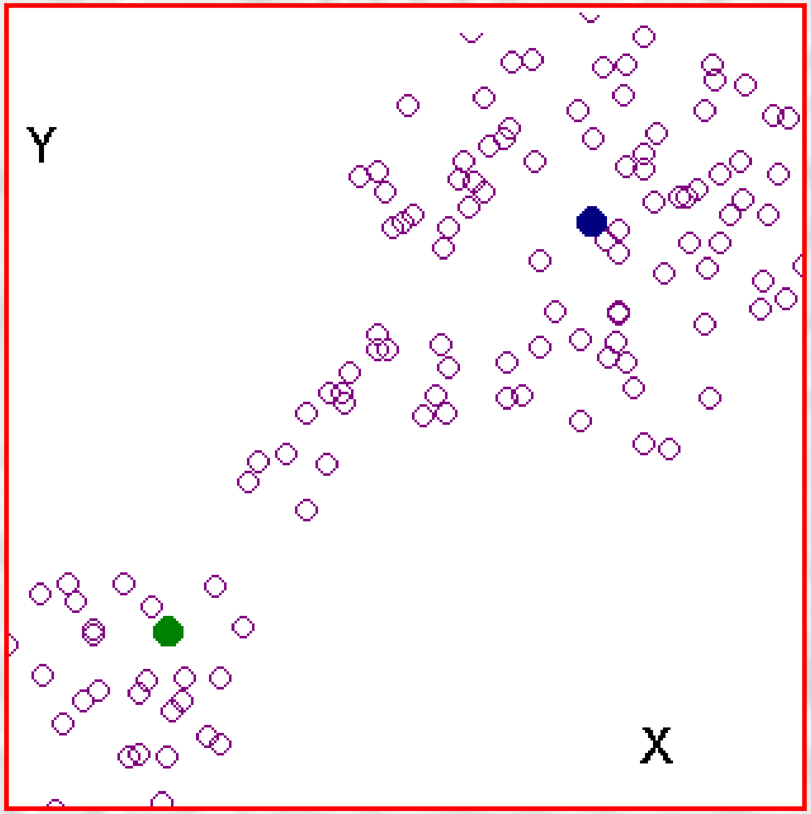
The extracted rules can help discover and then analyse the hidden knowledge in data sets.

II. EXTRACTING RULES FROM TRAINED RBF NEURAL NETWORK

Rule extraction process from NN

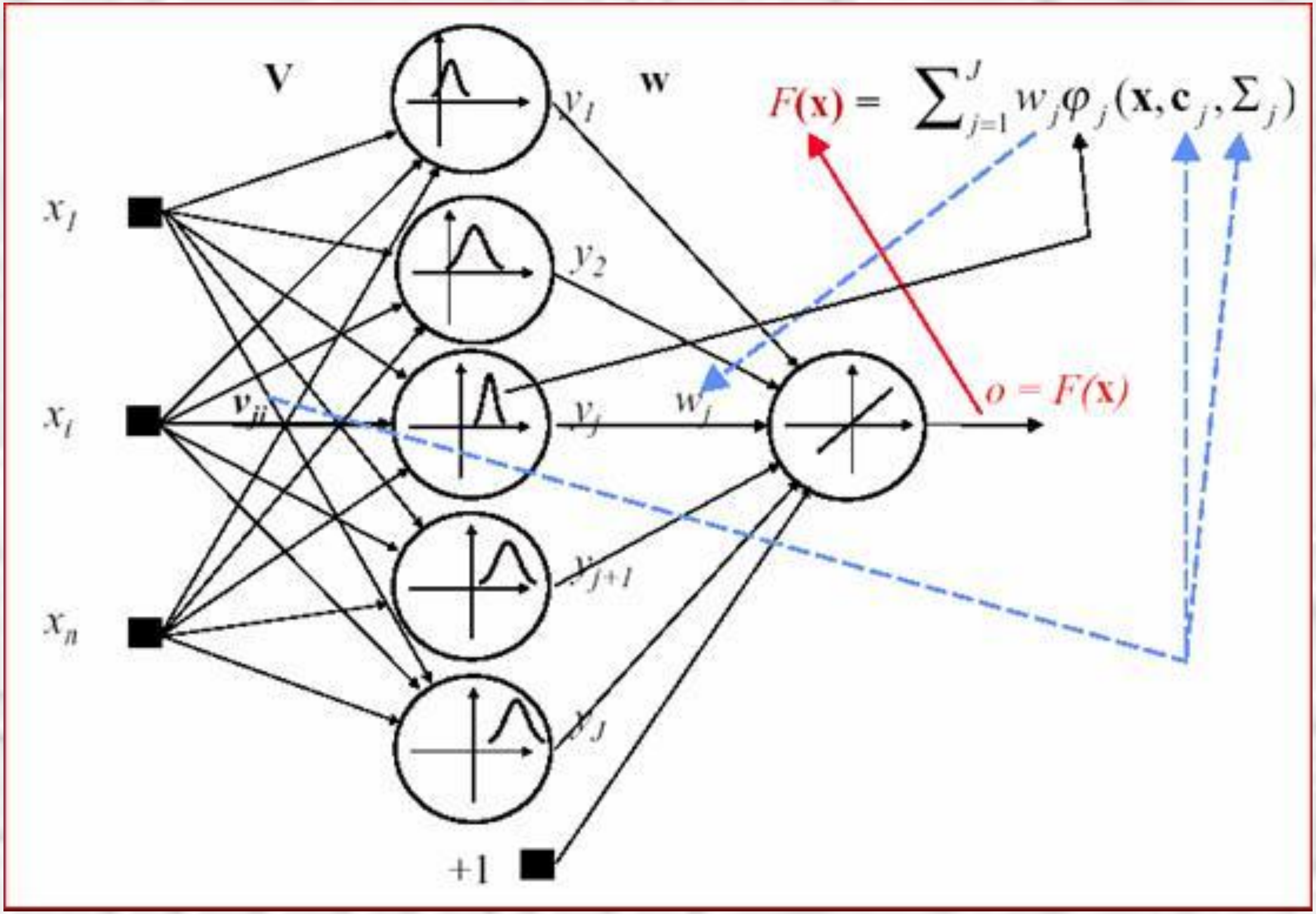


Simple idea

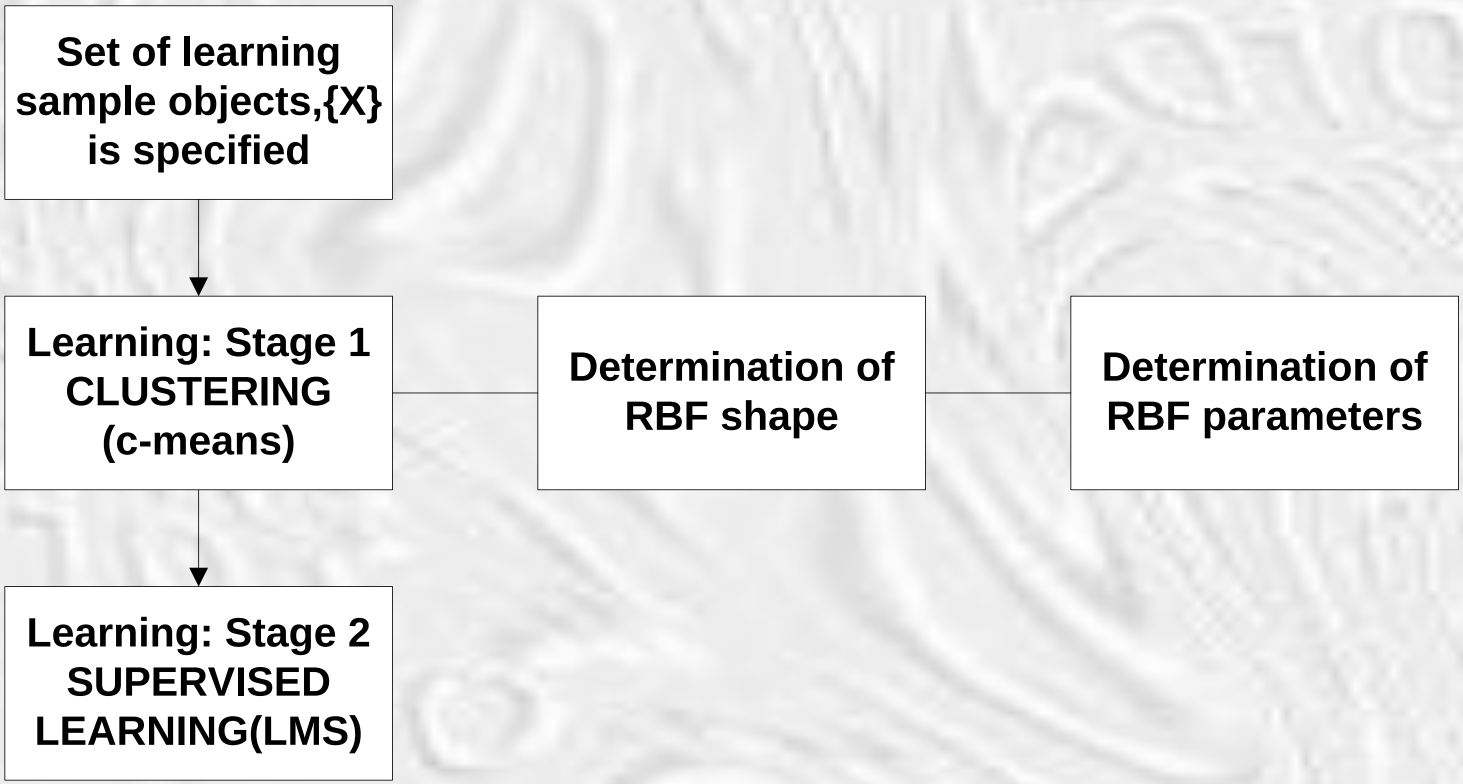


IF x is high THEN y is high
IF x is low THEN y is low

RBF Network Architecture



Clustering – the first stage of the RBF network learning



RULEX algorithm

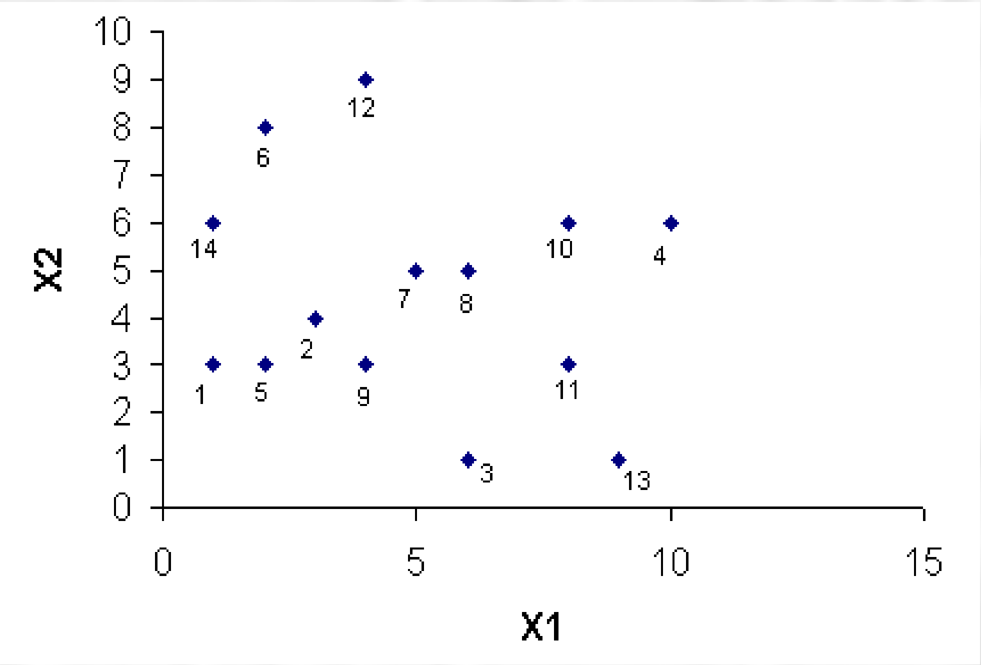
The local nature of each RBF hidden unit enables a simple translation into a single rule:

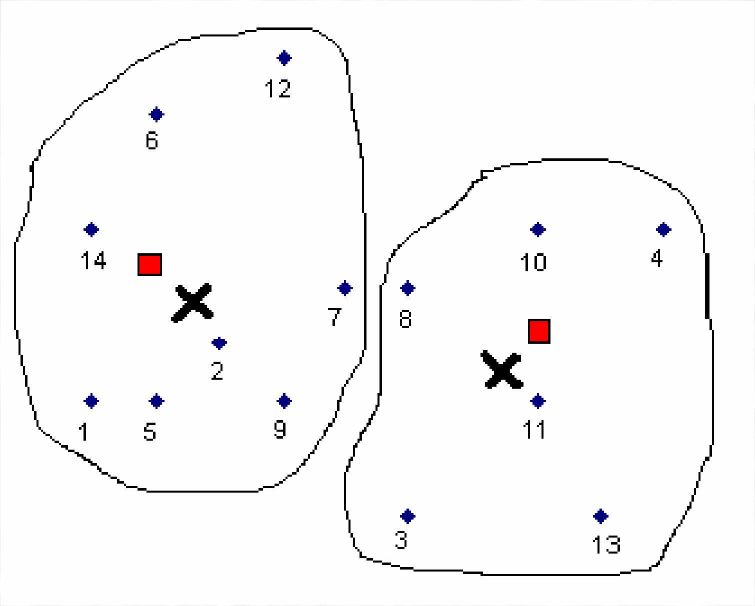
IF $Feature_1$ is TRUE AND IF $Feature_2$ is TRUE AND IF $Feature_n$ is TRUE THEN $Class_x$,

Input:	**Hidden weights μ (centre positions)** **Gaussian radius spread σ** **Steepness S**
Output:	**One rule per hidden unit**

Procedure:	**Train RBF network on data set**
	For each hidden unit:
	    For each μ_i
	        $X_{lower} = \mu_i - \sigma_i + S$
	        $X_{upper} = \mu_i + \sigma_i - S$
	Build rule by:
	    antecedent=[X_{lower}, X_{upper}]
	    Join antecedents with AND
	    Add class label
	Write rule

Simple example



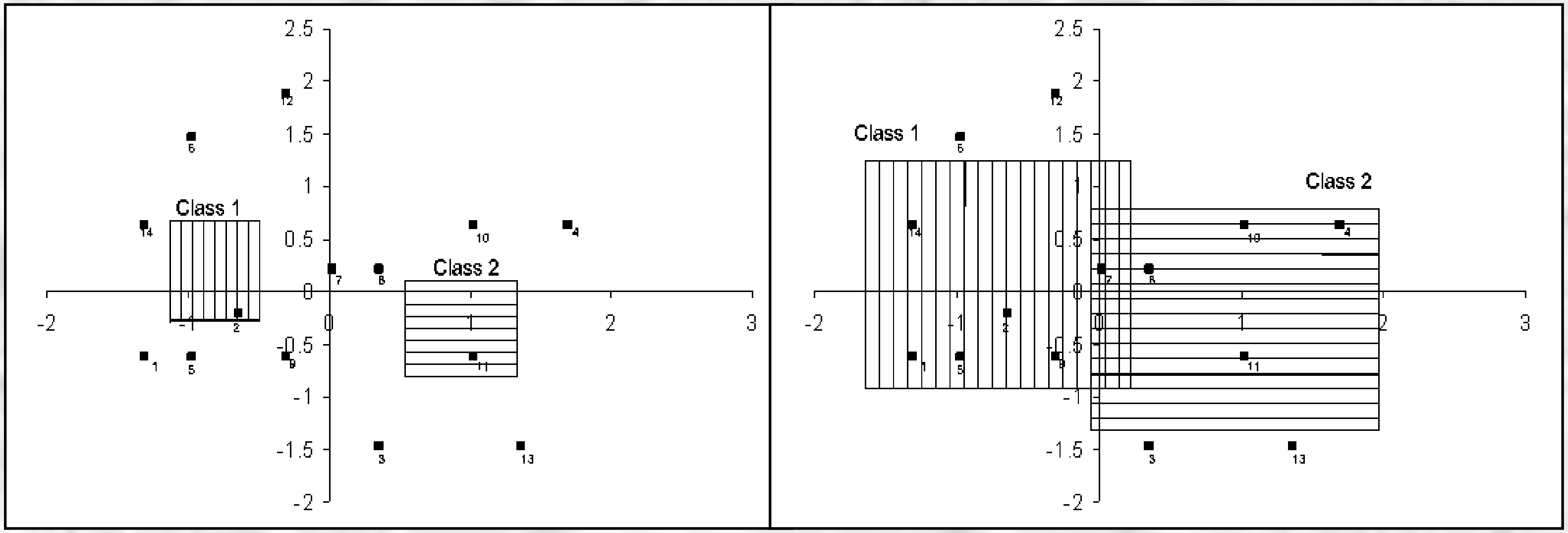


Two clusters with centers at points (-0.73; 0.26) and (0.97; -0.35)

radius values σ_1^2 = 1.07 and σ_2^2 = 1.04

Results

Regions of rules are represented in Figure.



a) b)

Region of rules: steepness=0.6 (a) and steepness=0 (b)

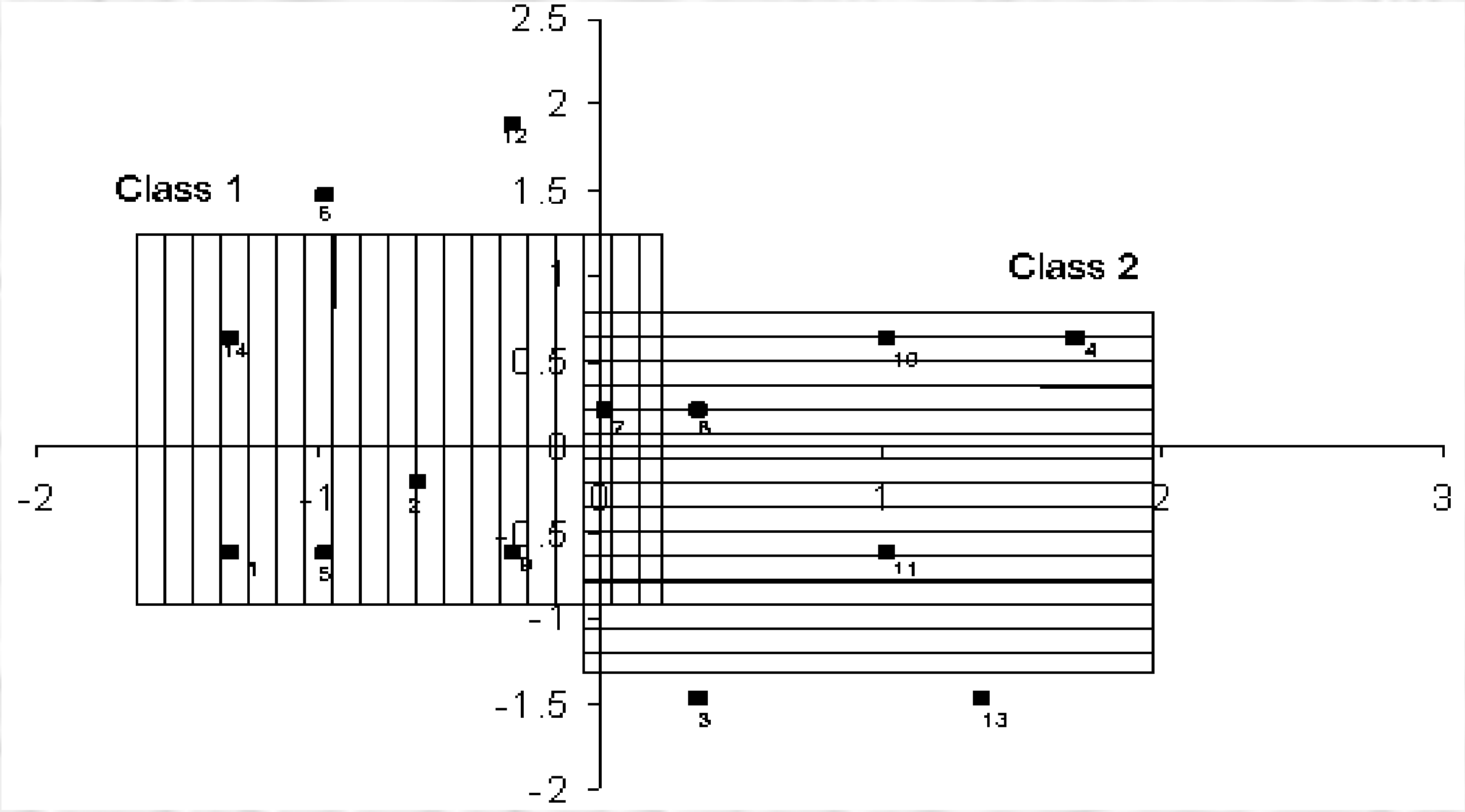
Results

Errors: 28% (4 input vectors out of 14 - Points 3, 6, 12 and 13 in fig.

RULES: (Steepness=0)

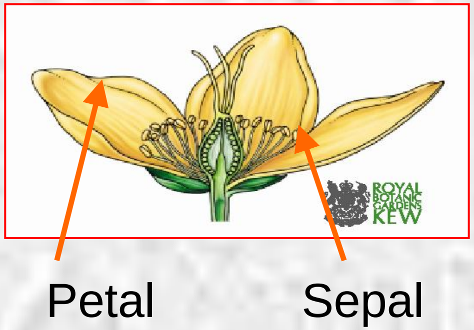
IF (x1 ≥ -1.76 AND ≤ 0.3) AND IF (x2 ≥ - 0.77 AND ≤ 1.29) THEN CLASS 1

IF (x1 ≥ -0.04 AND ≤ 1.98) AND IF (x2 ≥ - 1.36 AND ≤ 0.66) THEN CLASS 2.



Application example: IRIS data set

setosa, versicolor un virginica.



4 parameters:

SL – *sepal length*

SW – *sepal width*

PL - *petal length*

PW – *petal width*

Setosa			
SL	SW	PL	PW
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
....	....	....	....

Versicolor			
SL	SW	PL	PW
7.0	3.2	4.7	1.4
6.4	3.2	4.5	1.5
6.9	3.1	4.9	1.5
5.5	2.3	4.0	1.3
6.5	2.8	4.6	1.5
....	....	....	....

Virginica			
SL	SW	PL	PW
6.3	3.3	6.0	2.5
5.8	2.7	5.1	1.9
7.1	3.0	5.9	2.1
6.3	2.9	5.6	1.8
6.5	3.0	5.8	2.2
....	....	....	....

Results

	Parameter S=-0.9	Parameter S=0
Values of centers and radii	Class 1 = 5.01 3.42 1.46 0.24 Class 2 = 5.94 2.77 4.26 1.33 Class 3 = 6.59 2.97 5.55 2.03 Values of radii = 0.30 0.61 0.87	Class 1 = 5.01 3.42 1.46 0.24 Class 2 = 5.94 2.77 4.26 1.33 Class 3 = 6.59 2.97 5.55 2.03 Values of radii = 0.30 0.61 0.87
Rules correctly describe elements of classes (%)	100	58.7

Rule of Class 1	IF (X1>= 3.80 AND < 6.21) AND
	IF (X2>= 2.21 AND < 4.62) AND
	IF (X3>= 0.26 AND < 2.67) AND
	IF (X4>= -0.96 AND < 1.45)
	THEN SETOSA
		IF (X1>= 4.70 AND < 5.31) AND
		IF (X2>= 3.11 AND < 3.72) AND
		IF (X3>= 1.16 AND < 1.77) AND
		IF (X4>= -0.06 AND < 0.55)
		THEN SETOSA

Rule of Class 2	IF (X1>= 4.42 AND < 7.45) AND
	IF (X2>= 1.26 AND < 4.28) AND
	IF (X3>= 2.75 AND < 5.77) AND
	IF (X4>= -0.19 AND < 2.84)
	THEN VERSICOLOR
		IF (X1>= 5.32 AND < 6.55) AND
		IF (X2>= 2.16 AND < 3.38) AND
		IF (X3>= 3.65 AND < 4.87) AND
		IF (X4>= 0.71 AND < 1.94)
		THEN VERSICOLOR

Rule of Class 3	IF (X1>= 4.82 AND < 8.36) AND
	IF (X2>= 1.20 AND < 4.74) AND
	IF (X3>= 3.78 AND < 7.32) AND
	IF (X4>= 0.26 AND < 3.80)
	THEN VIRGINICA
		IF (X1>= 5.72 AND < 7.46) AND
		IF (X2>= 2.10 AND < 3.84) AND
		IF (X3>= 4.68 AND < 6.42) AND
		IF (X4>= 1.16 AND < 2.90)
		THEN VIRGINICA

Results of training set B (arbitrary 20 elements of every class)

Correct	Values of parameter S											
	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0	0.1	0.2
Class 1	49	49	48	48	45	40	39	27	14	9	2	0
Class 2	50	49	49	48	45	44	40	36	28	20	10	3
Class 3	49	49	48	47	45	43	43	42	39	35	29	23
%	98.7	98	96.7	95.3	90	84.7	81.3	70	54	42.7	27.3	17.3

Conclusions for NN

1) After training the RBF classifier, the rules will be extracted through analyzing the parameters of the classifier.

2) One hidden unit corresponds to one rule.

3) It is desirable to reduce the number of hidden units of RBF neural networks while maintaining high classification accuracy.

4) The extracted rules can help discover and analyze the hidden knowledges in data sets further.

Conclusions

Such rule extraction technique is shown through IRIS data set experimental results.

The extracted rules can help discover and analyze the hidden knowledge in data sets further.

The experiments have shown that these methods can be viewed as alternatives to traditional data analysis methods.

The correct adjustment of parameters in both methods proposed will allow minimizing data processing risks in the analysis of economic data.

Thanks !