Applications of Fuzzy Systems

Fuzzy Classification Rule Mining:

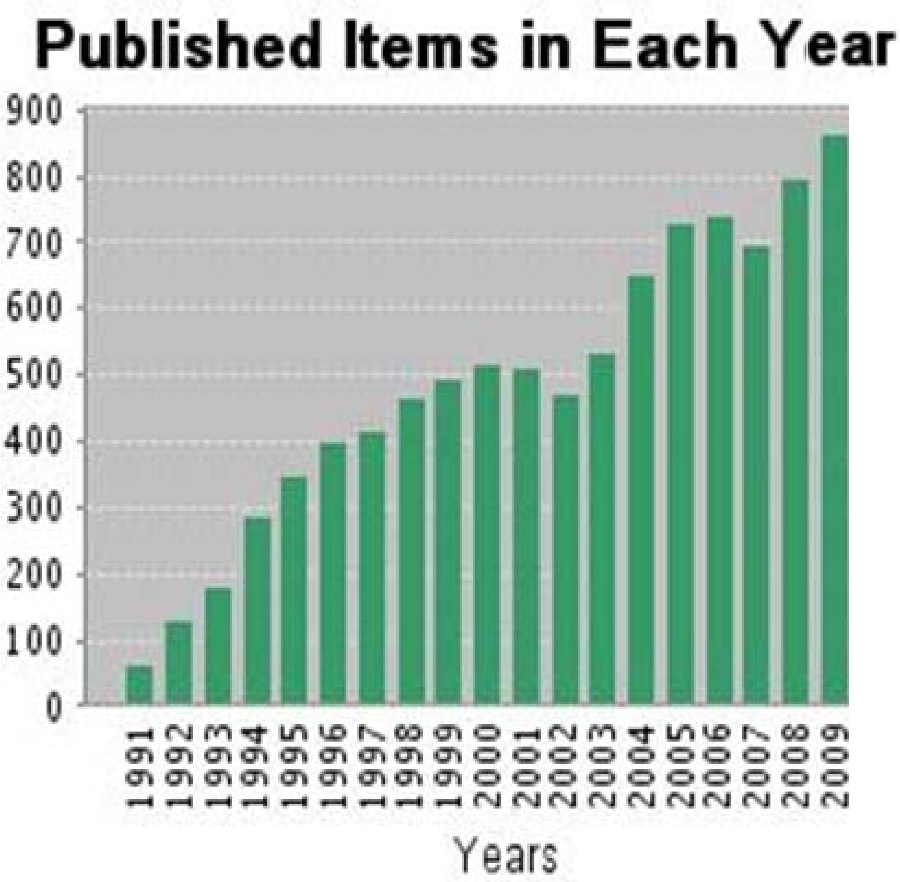
A case study

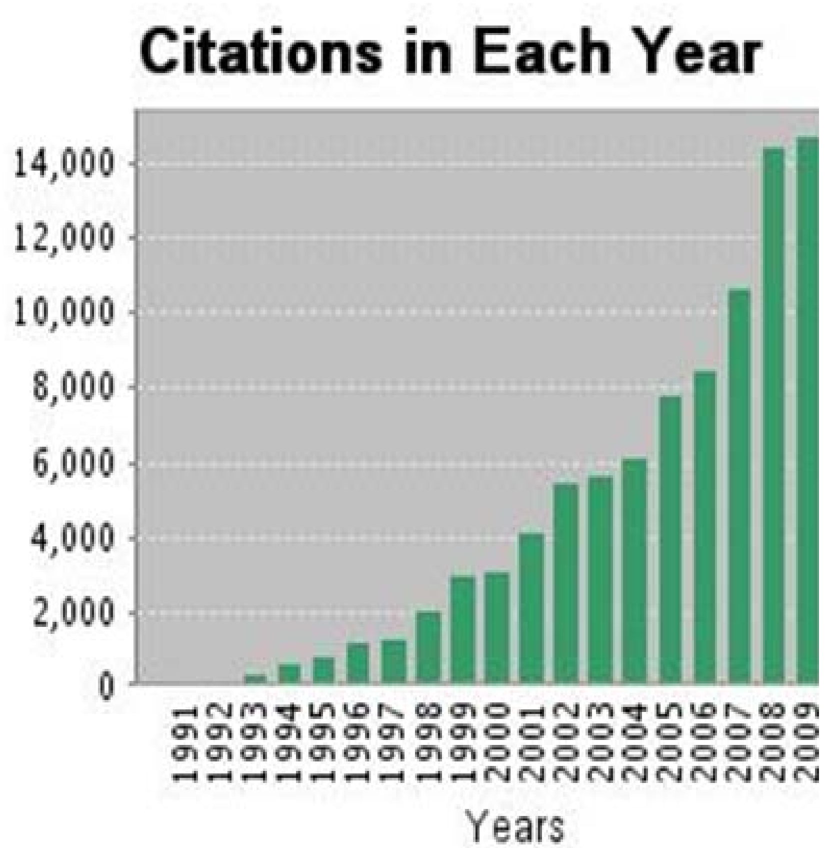
B. B. Misra

Application Areas of Fuzzy Systems

- Fuzzy Control
- Fuzzy Clustering
- Fuzzy Classification

Fuzzy Control

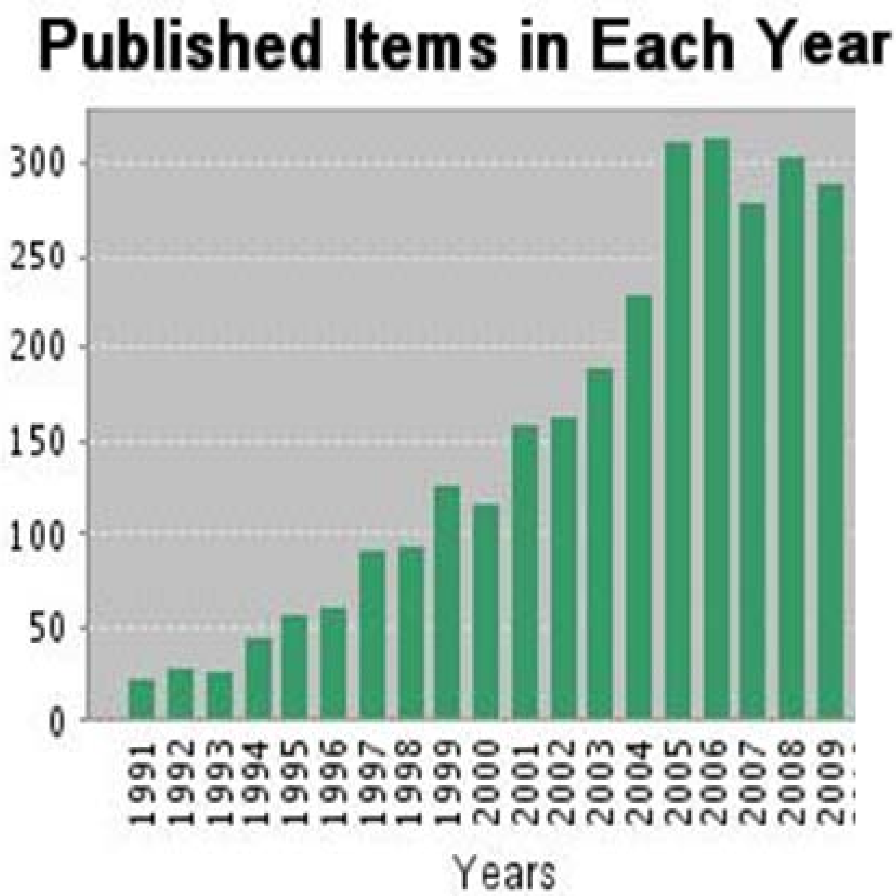


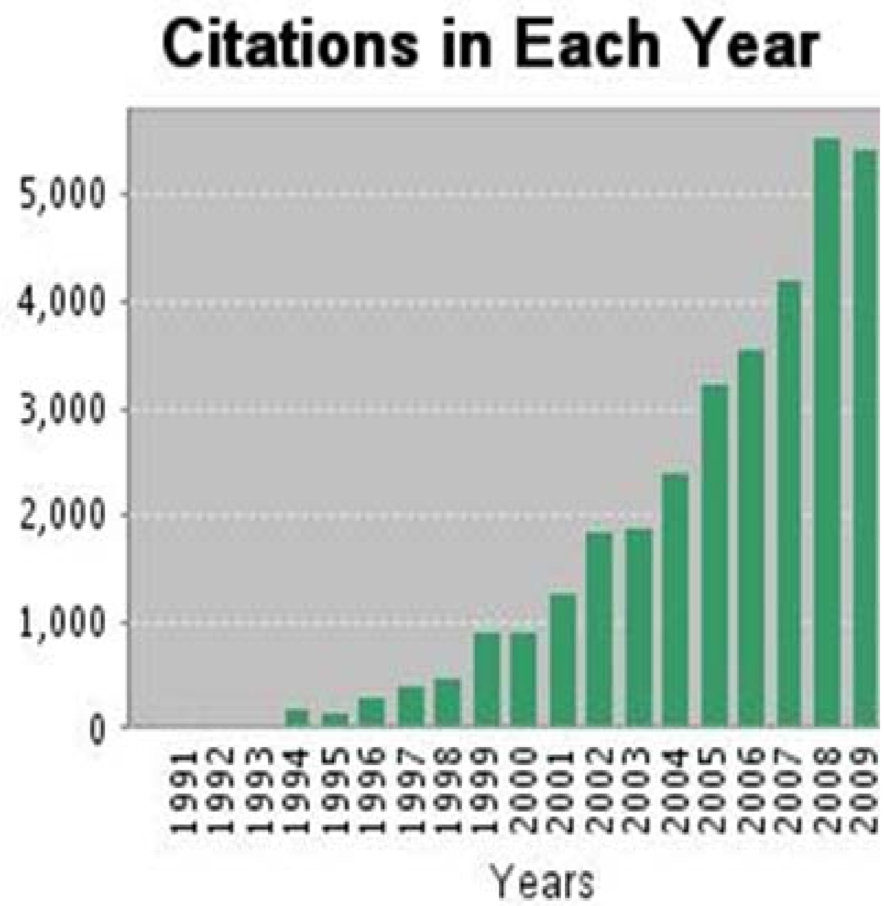


Publications: 9,421

Citations: 90,485

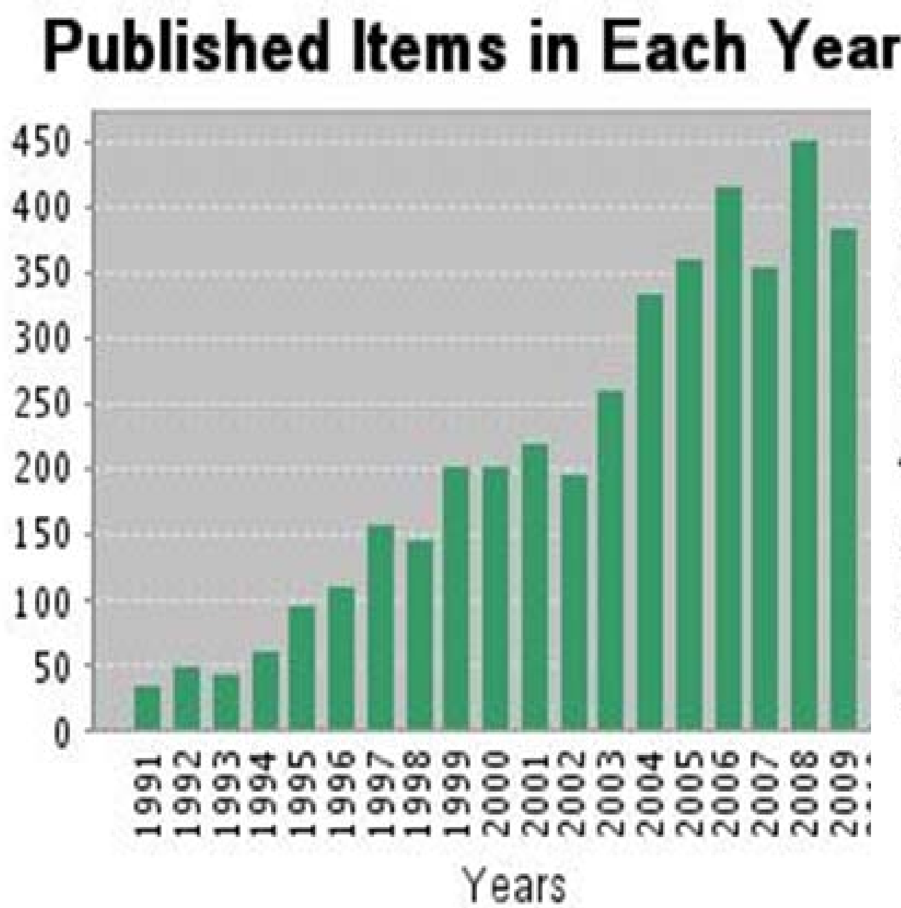
Fuzzy Clustering

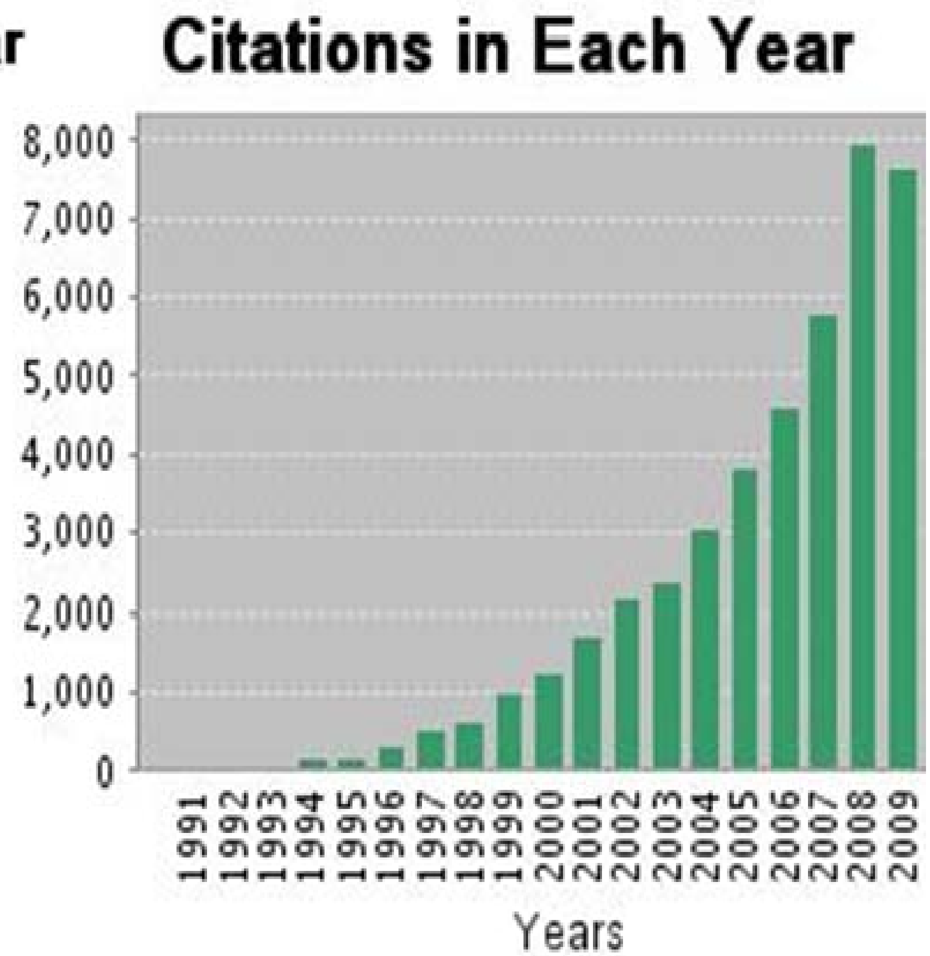




Publications: 2,968
Citations: 32,977

Fuzzy Classification





Publications: 4,144

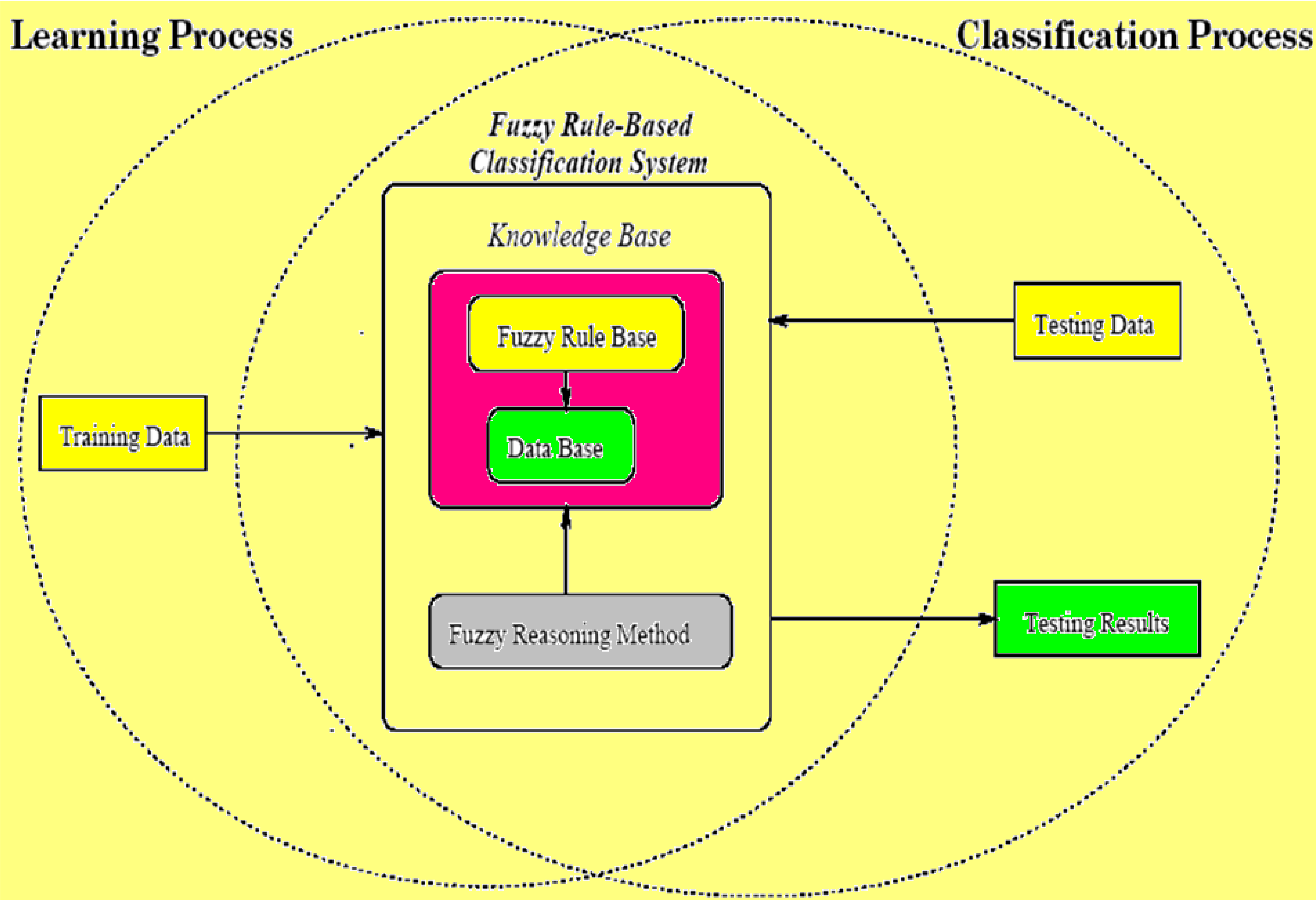
Citations: 43,485

The Classification Task

- Given a database $D=\{t_1, t_2, \ldots, t_n\}$ of tuples and a set of classes $C=\{C_1, C_2, \ldots, C_m\}$, the classification problem is to define a mapping $f:D \rightarrow C$ where each t_i is assigned to one class. A class C_j contains precisely those tuples mapped to it, that is

$$C_j = \{t_i \mid f(t_i) = C_j, 1 \leq i \leq n, \text{ and } t_i \in D\}$$

Fuzzy Rule-Based Classification

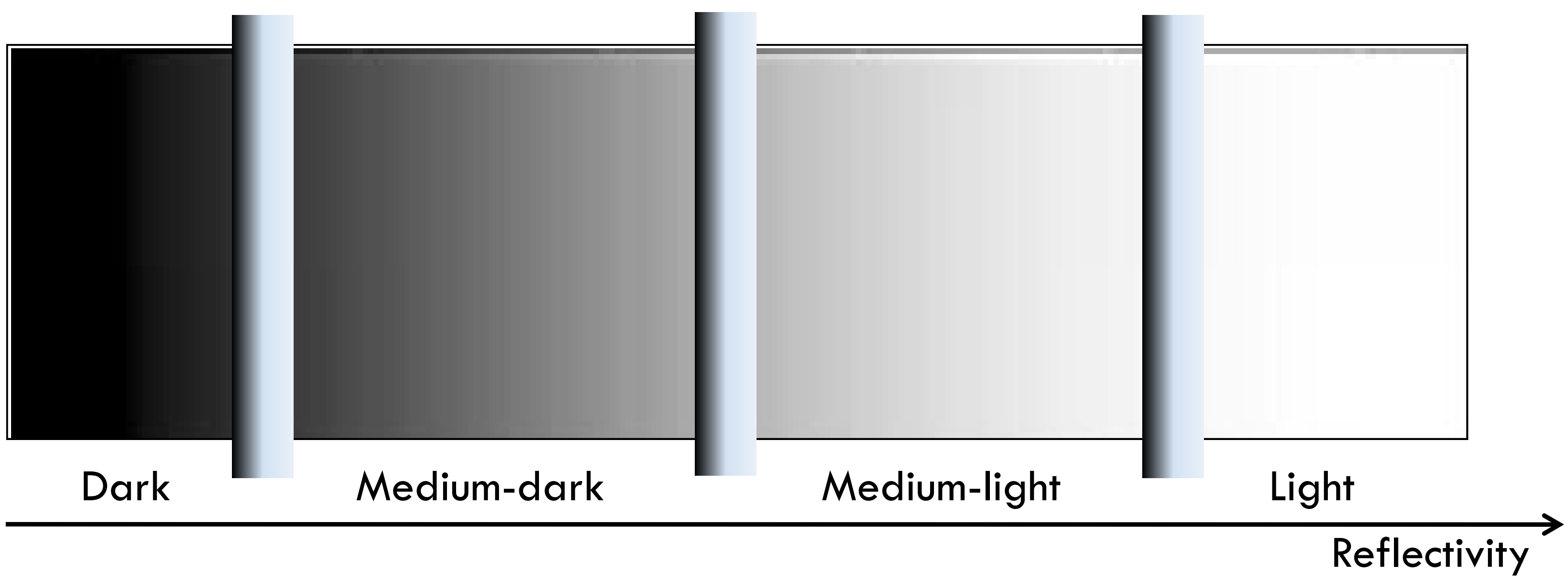


Fuzzy Classification

- Informal knowledge about problem domain used for classification
- Example:
  - Adult salmon is oblong and light in color
  - Sea bass is stouter and dark
- Goal of fuzzy classification
  - Create fuzzy "category memberships" function
    - To convert objectively measurable parameters to "category memberships"
- Which are then used for classification

Categories

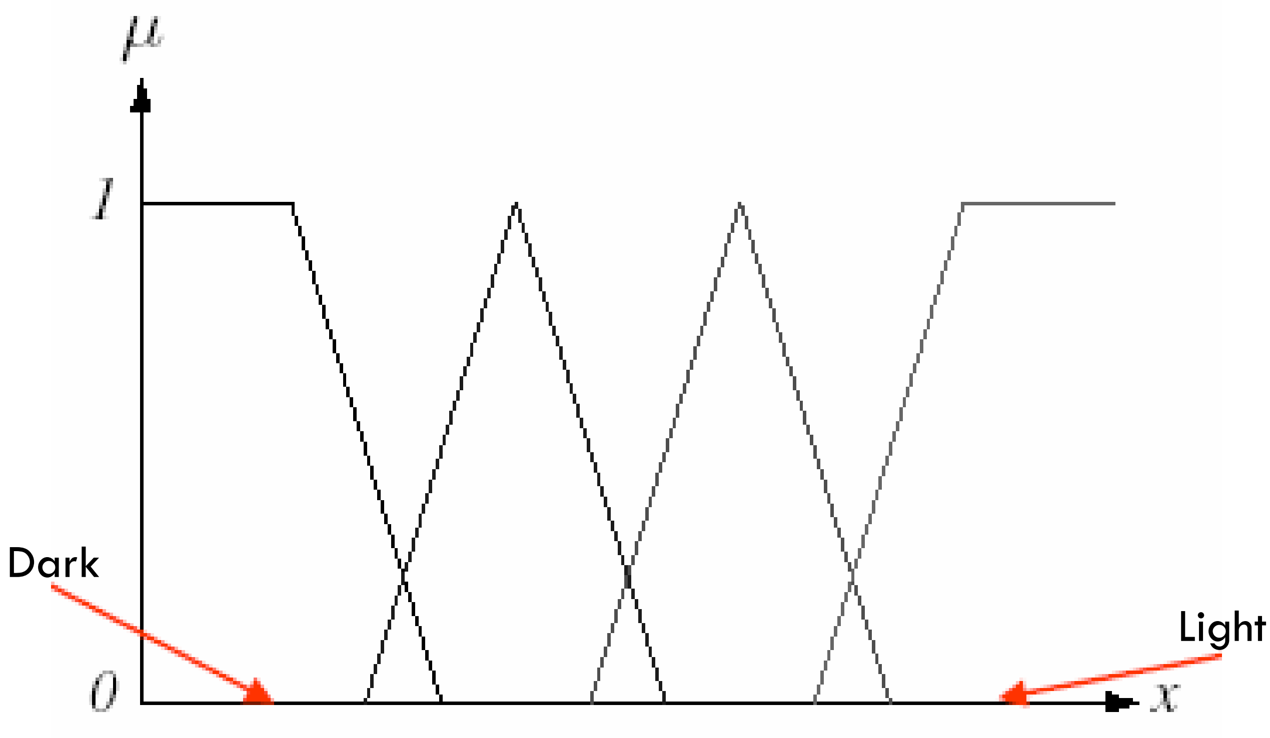
- **Does not refer to final classes**
- **Refer to overlapping ranges of feature values**
- **Example:**
  - **Lightness is divided into four categories**
  - **Dark, medium-dark, medium-light, light**



Category membership functions

- It is derived from the designer's prior knowledge with conjunction rule lead to discriminants.
- Here x represents objectively measureable value i.e. reflectivity of a fish's skin.
- Designer feels four categories for the relativity feature i.e. dark, medium-dark, medium-light, light.

The categories of the feature are not the same as the true categories or classes for the pattern.



Conjunction Rule

- Merging several category functions corresponding to different features to yield a number to make the final decision
- Example: two category membership functions can be merged using

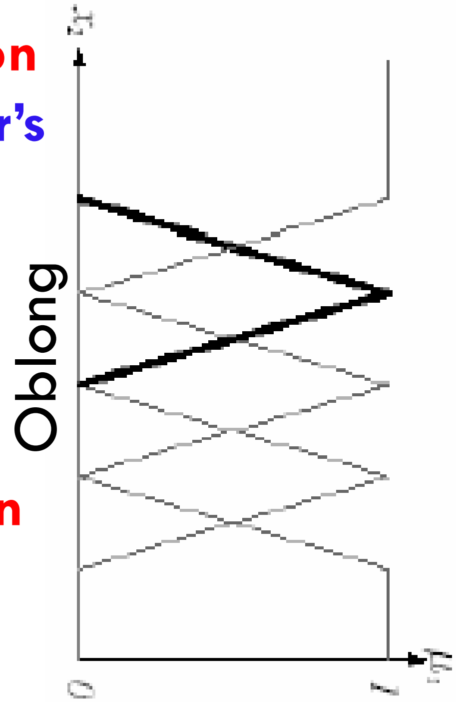
$$\mu_x(X).\ \mu_y(Y)$$

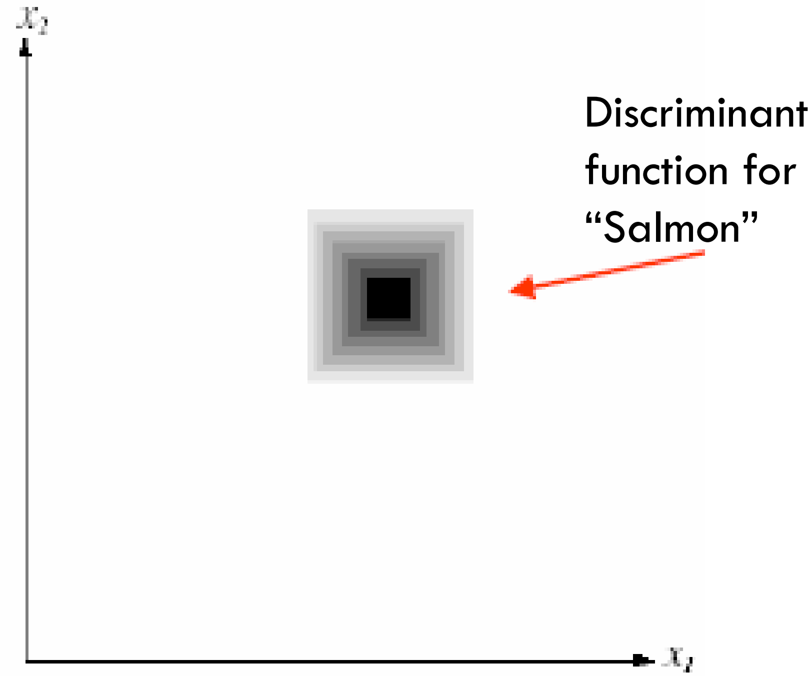
Specifies category function for x

Measured value of feature

Discriminant function based on category membership functions

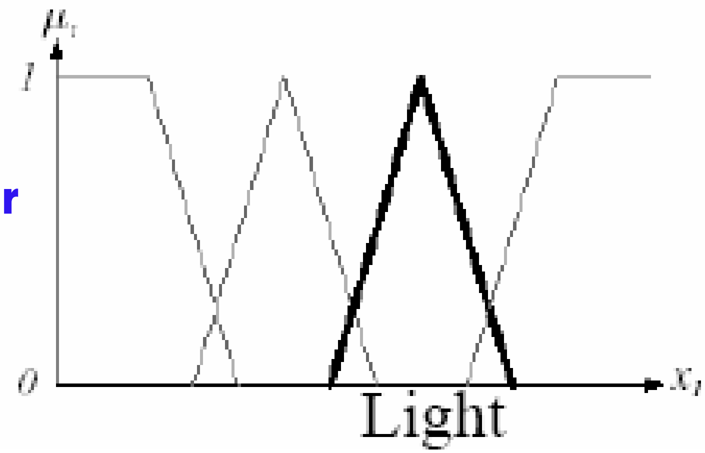
"Category membership" functions and a conjunction rule based on the designer's prior knowledge lead to discriminant functions.





It is expected that a particular class can be described as the conjuction of two "category memberships"

Here the conjuction rule gives the discriminant function.
Similarly construct discrriminant function for other categories.
For classification, maximum discriminant function is taken.



(From R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification copyright © 2001 by John Wiley & Sons, Inc.)

Representation of fuzzy rules

Usually the fuzzy if-then rules are represented in three forms.

1. Fuzzy rules with a class in the consequent

Rule R_k : If x_1 is $A^k{}_1$ and … and x_n is $A^k{}_n$, then Y is C_j

Where x_1,…,x_n are selected features for classification problem, $A^k{}_1$,…, $A^k{}_n$ are linguistic labels used to discretize the continuous domain of the variables, Y in the class C_j to which pattern belongs.

2. Fuzzy rules with a class and a certain degree in the consequent

Rule R_k : If x_1 is A^k_1 and … and x_n is A^k_n, then Y is C_j with r^k

Where r^k is the certainty degree of the classification in the class C_j for a pattern belonging to the fuzzy subspace delimited by the antecedent.

This certainty degree is determined by the ratio,

$$r^k = S^k_j / S^k$$

where S^k_j is the sum of the matching degrees for the class C_j patterns belonging to the fuzzy region delimited by the antecedent, and

S^k is the sum of the matching degrees for all the patterns belonging to this fuzzy subspace, regardless its associated class.

3. Fuzzy rules with certain degree for all classes in the consequent

Rule R_k : If x_1 is A^k_1 and … and x_n is A^k_n, then Y is C_j with $r^k_1, \ldots, r^k_m$

Where r^k_j is the soundness degree for the rule k to predict the class C_j for a pattern belonging to the fuzzy region represented by the antecedent of the rule.

The degree of certainty is determined as in the previous case.

Phases of rule generation

- **Generation of fuzzy if-then rules from numerical data consists of two phases:**
  - fuzzy partition of a pattern space into fuzzy subspaces and
  - determination of a fuzzy if-then rule for each fuzzy subspace.

Design of the Classifier

- **Let the pattern space is the unit square [0, 1] [0, 1] for the simplicity of notation.**
- **Suppose that m patterns $x_p = (x_{p1}, x_{p2})$, p=1,2,…,m are given as training patterns from M (M « m) classes: Class 1 (C1), Class 2 (C2), …, Class M (CM).**
- **Let each axis of the pattern space is partitioned into k fuzzy subsets $\{A^k_1, A^k_2, …, A^k_k\}$ where A^k_i is the i^{th} fuzzy subset.**
- **Then, let us use the following fuzzy if-then rule**

Rule R^k_{ij} : If x_{p1} is A^k_i and x_{p2} is A^k_j, then x_p belongs to C^k_{ij}

with CF = CF^k_{ij}

where R^k_{ij} is the label of the fuzzy if-then rule, A^k_i and A^k_j are fuzzy subsets in the unit interval [0, 1], C^k_{ij} is the consequent classes and CF^k_{ij} is the grade of certainty of the fuzzy if-then rule.

Generation of fuzzy if-then rules

Step1: Calculate β_{CT} for each class T(T=1,2,...,M) as

$$\beta_{CT} = \sum_{x_p \in CT} \mu_i^k(x_{p1}).\mu_j^k(x_{p2})$$

where β_{CT} is the sum of the compatibility of x_p's in class T to the fuzzy if-then rule R^k_{ij}.

Step 2: Find class X(CX) such that $\beta_{CT} = \max\{\beta_{C1}, \beta_{C2},\ldots, \beta_{CM}\}$.

If two or more classes take the maximum value or all the β_{CT}'s are zero, the consequent C^k_{ij} of the fuzzy if-then rule corresponding to the fuzzy subspace $A^k_i \times A^k_j$ can not be determined uniquely.

In this case, let $C^k_{ij} = \emptyset$.

If a single class takes the maximum value C^k_{ij} is determined as CX.

Generation of fuzzy if-then rules cntd.

Step 3: If a single class takes the maximum value in step 2, CF^k_{ij} is determined as

$$CF_{ij}^{k} = \frac{\beta_{CX} - \beta}{\sum_{T=1}^{M} \beta_{CT}} \qquad \textbf{where} \qquad \beta = \sum_{\substack{T=1 \\ T \neq x}}^{M} \frac{\beta_{CT}}{M-1}$$

==

Let us denote the set of the generated K^2 fuzzy if-then rules by S^K.

$$S^K = \{R^k_{ij} \mid i=1,2,\ldots,K; j=1,2,\ldots,K\}.$$

That is, S^K is the rule set corresponding to the $K \times K$ fuzzy rule table.

Classification of a new pattern

When a rule set S is given, a new pattern $x_p = (x_{p1}, x_{p2})$ is classified by the following procedure based on the fuzzy if-then rules in S.

Step 1: Calculate α_{CT} for each class T (T = 1,2,…,M) as

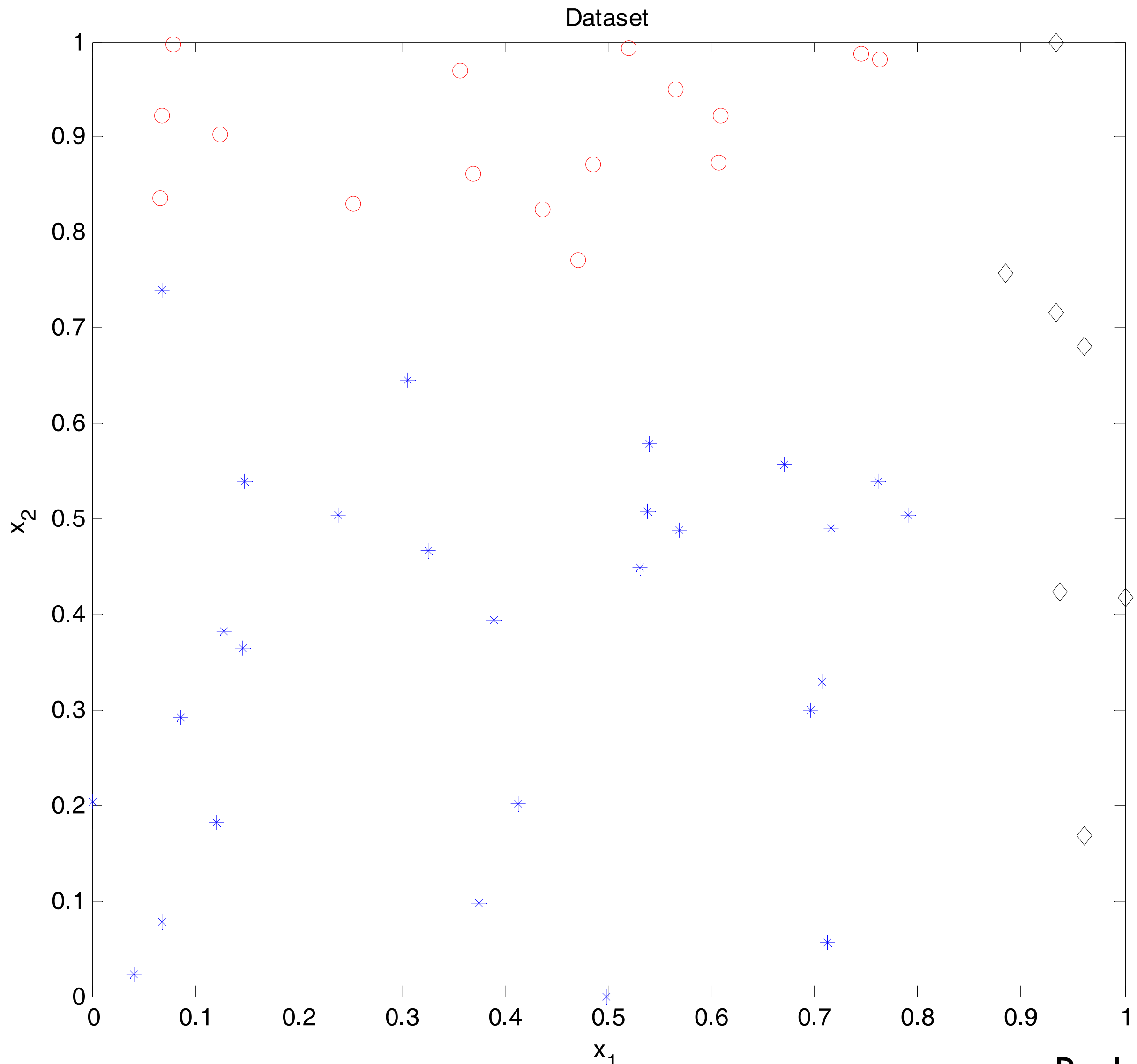
$$\alpha_{CT} = \max\left\{\mu_{\mathrm{i}}^{\mathrm{k}}\left(x_{p1}\right).\mu_{\mathrm{j}}^{\mathrm{k}}\left(x_{p2}\right).CF_{ij}^{k} \mid CF_{ij}^{k} = CT \text{ and } \mathrm{R}_{\mathrm{ij}}^{\mathrm{k}} \in S\right\}$$

Step 2: Find class X(CX) such that $\alpha_{CX} = \max\left\{\alpha_{C1}, \alpha_{C2}, \ldots, \alpha_{CM}\right\}$

If two or more classes take the maximum value or all the α_{CT}'s are zero, then pattern x_p is considered unclassifiable,

otherwise assign x_p to class X(CX).

Sample Dataset



Red circle : class1

Blue Star: class2

Black Diamond: class3

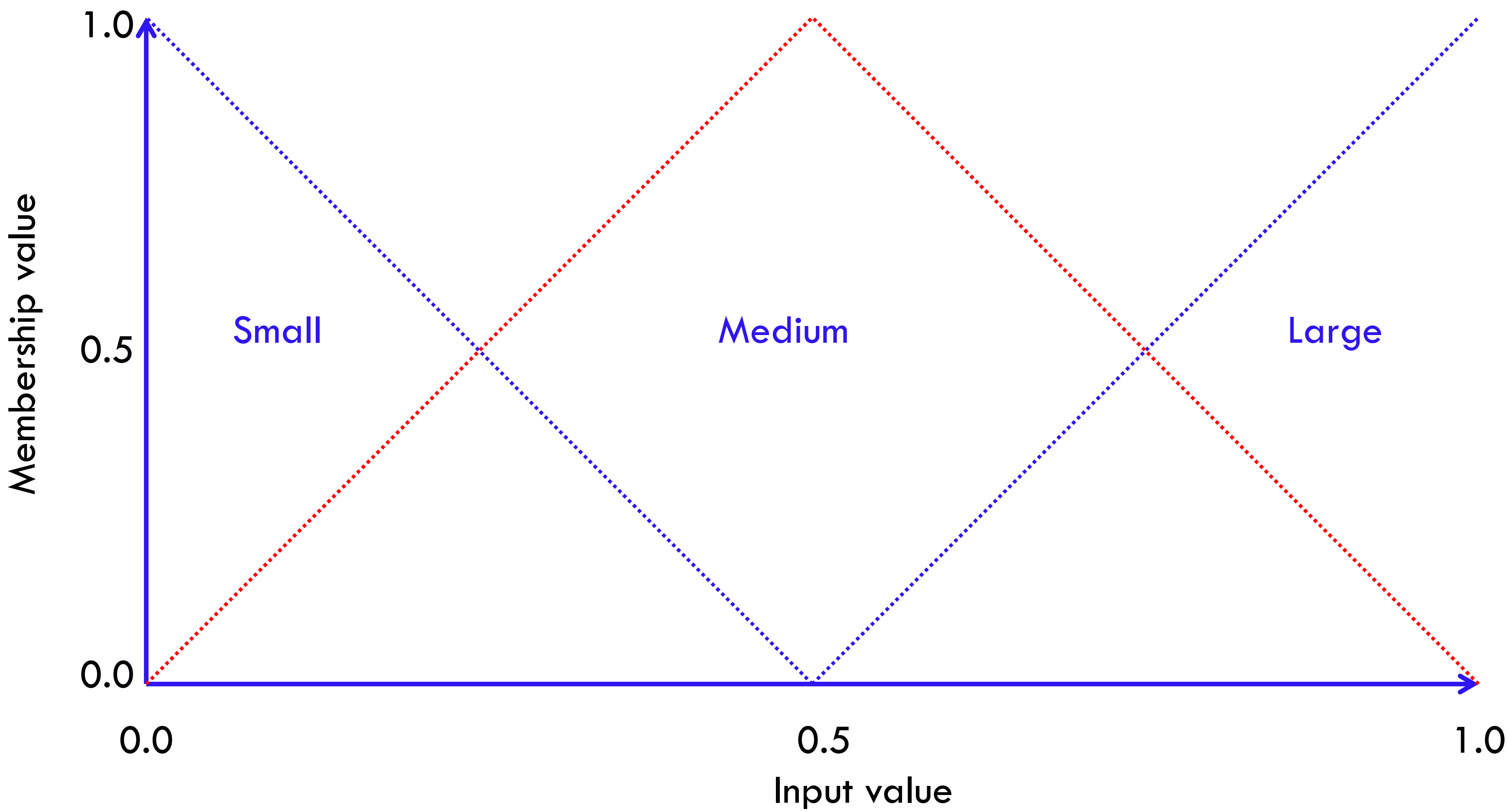
Let triangular fuzzy membership function is used.

$$triangle(x;a,b,c)=\begin{cases}0, & x \le a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x\end{cases}$$

Let the linguistic variable considered are small (S), medium(M) and large (L)

Table shows the parameters for each linguistic variable.

	a	b	c
Small	0	0	0.5
Medium	0	0.5	1
Large	0.5	1	1



List of all possible rules

- R1: If x1 is small and x2 is small
- R2: If x1 is small and x2 is medium
- R3: If x1 is small and x2 is large
- R4: If x1 is medium and x2 is small
- R5: If x1 is medium and x2 is medium
- R6: If x1 is medium and x2 is large
- R7: If x1 is large and x2 is small
- R8: If x1 is large and x2 is medium
- R9: If x1 is large and x2 is large

- Then let us find out which rule belongs to which class

Calculate for Rule 1:

If x1 is small and x2 is small

$$triangel(x;a,b,c)=\begin{cases}0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x\end{cases}$$

- **Calculate for the class1 patterns**

  $\mu_{small}(x1)*\mu_{small}(x2)$

	a	b	c
Small	0	0	0.5
Medium	0	0.5	1
Large	0.5	1	1

- **P1: (0.5-0.06)/(0.5-0.0) * 0=0**
- **P2: 0* 0=0**
- **P3: (0.5-0.34)/0.5 * 0=0**

#	x1	x2	class
P1	0.06	0.99	1
P2	0.59	0.76	1
P3	0.34	0.88	1
P4	0.17	0.02	2
P5	0.63	0.69	2
P6	0.01	0.27	2
P7	0.46	0.05	3
P8	0.93	0.77	3
p9	0.86	0.97	3

- **The sum of compatibility in class1 to rule1,** $\beta_{C1} = 0 + 0 + 0 = 0$

Calculate for Rule 1: cntd.

If x1 is small and x2 is small

$$triangel(x;a,b,c)=\begin{cases}0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x\end{cases}$$

	a	b	c
Small	0	0	0.5
Medium	0	0.5	1
Large	0.5	1	1

- Calculate for the class2 patterns

  $\mu_{small}(x1)*\mu_{small}(x2)$

- P4: (0.5-0.17)/0.5 * (0.5-0.02)/0.5

  =0.66 * 0.96=0.63

- P5: 0 * 0=0
- p6: (0.5-0.01)/0.5 * (0.5-0.27)/0.5

  =0.98 * 0.46=0.45

The sum of compatibility in class2 to rule1,

β_{C2} = 0.63 +0+0.45=1.08

#	x1	x2	class
P1	0.06	0.99	1
P2	0.59	0.76	1
P3	0.34	0.88	1
P4	0.17	0.02	2
P5	0.63	0.69	2
P6	0.01	0.27	2
P7	0.46	0.05	3
P8	0.93	0.77	3
p9	0.86	0.97	3

Calculate for Rule 1: cntd.

If x1 is small and x2 is small

- Calculate for the class3 patterns

$$\mu_{small}(x1)*\mu_{small}(x2)$$

- P7: (0.5-0.46)/0.5 * (0.5-0.05)/0.5

  = 0.08 * 0.90 = 0.07

- P8: 0*0=0
- P9: 0*0=0

The sum of compatibility in class3 to rule1, β_{C3} = 0.07

Then, the maximum compatible class =

max{β_{C1}, β_{C2}, β_{C3}}=max{0, 1.08, 0.07}= β_{C2}

R1: If x1 is small and x2 is small then class2

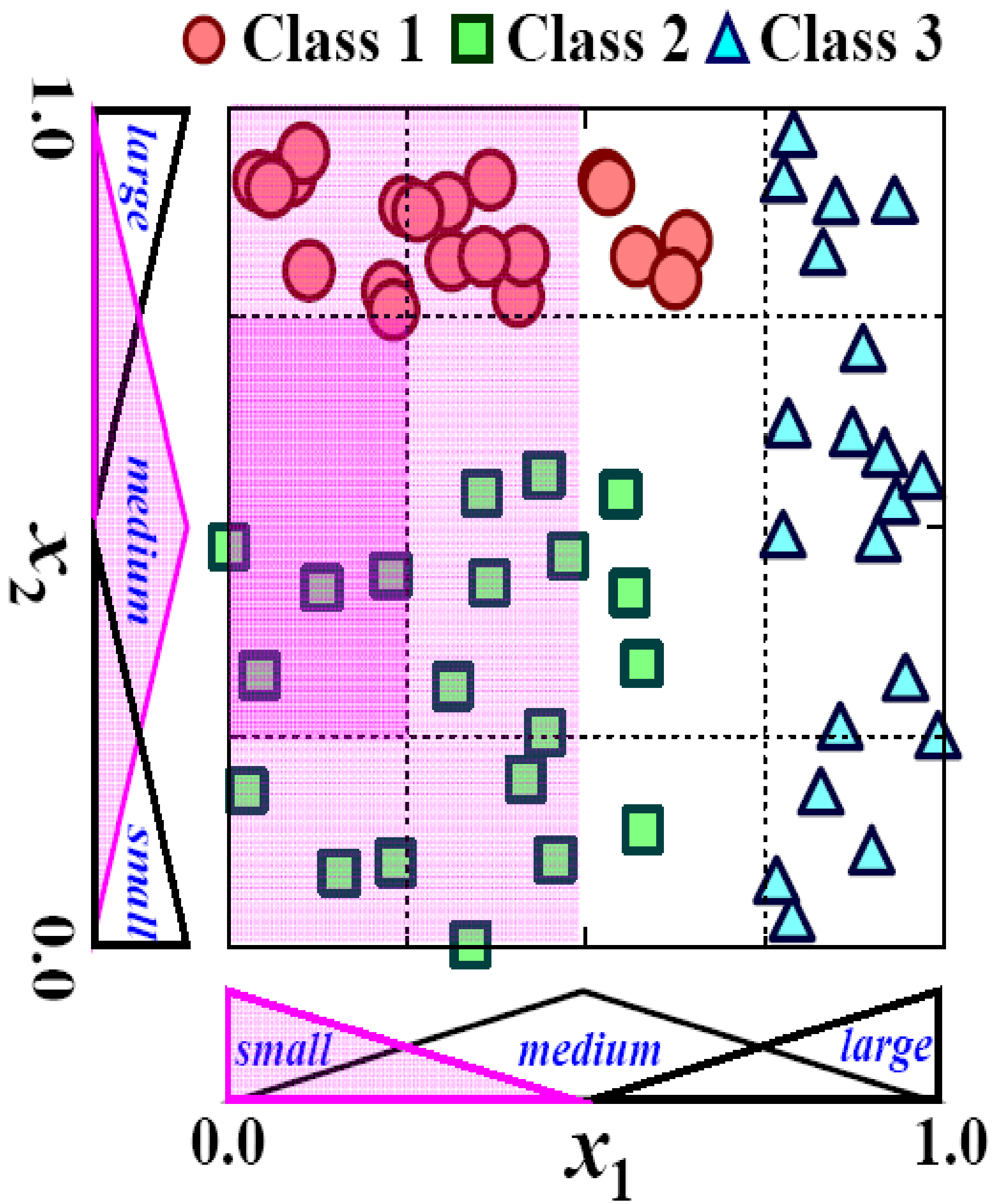
Though R1 is assigned C2, R1 may not classify all patterns of C2. Again R1 may misclassify patterns of other class as C2.

$$triangel(x;a,b,c)=\begin{cases}0, & x\le a\\ \frac{x-a}{b-a}, & a\le x\le b\\ \frac{c-x}{c-b}, & b\le x\le c\\ 0, & c\le x\end{cases}$$

	a	b	c
Small	0	0	0.5
Medium	0	0.5	1
Large	0.5	1	1

#	x1	x2	class
P1	0.06	0.99	1
P2	0.59	0.76	1
P3	0.34	0.88	1
P4	0.17	0.02	2
P5	0.63	0.69	2
P6	0.01	0.27	2
P7	0.46	0.05	3
P8	0.93	0.77	3
p9	0.86	0.97	3

Fuzzy Rule-Based Classifier Design

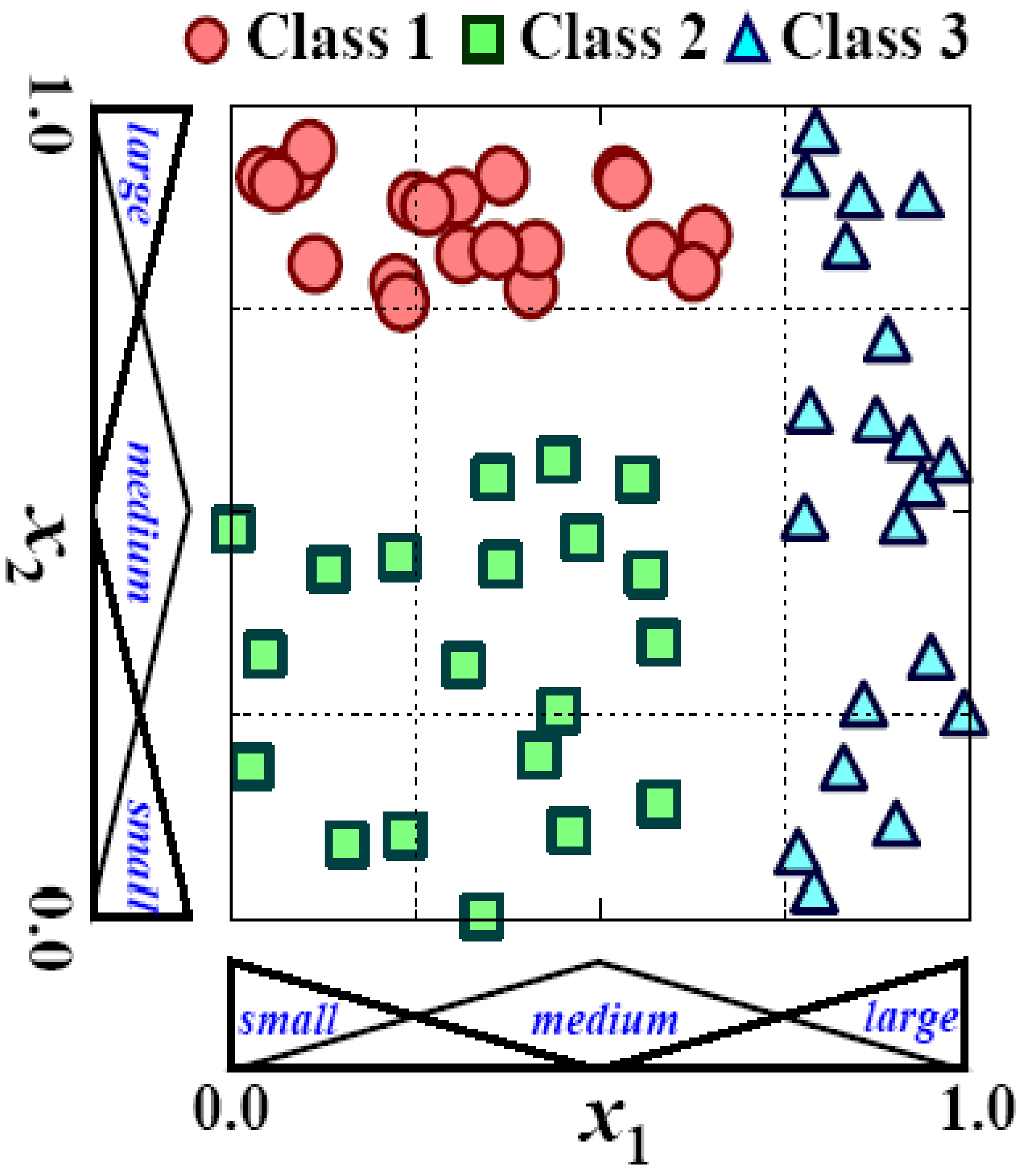


Basic Form

If x_1 is *small* and x_2 is *small*
then Class 2

If x_1 is *small* and x_2 is *medium*
then Class 2

Fuzzy Rule-Based Classifier Design cntd.



Basic Form

If x_1 is *small* and x_2 is *small* then Class 2

If x_1 is *small* and x_2 is *medium* then Class 2

If x_1 is *small* and x_2 is *large* then Class 1

. . .

If x_1 is *large* and x_2 is *large* then Class 3

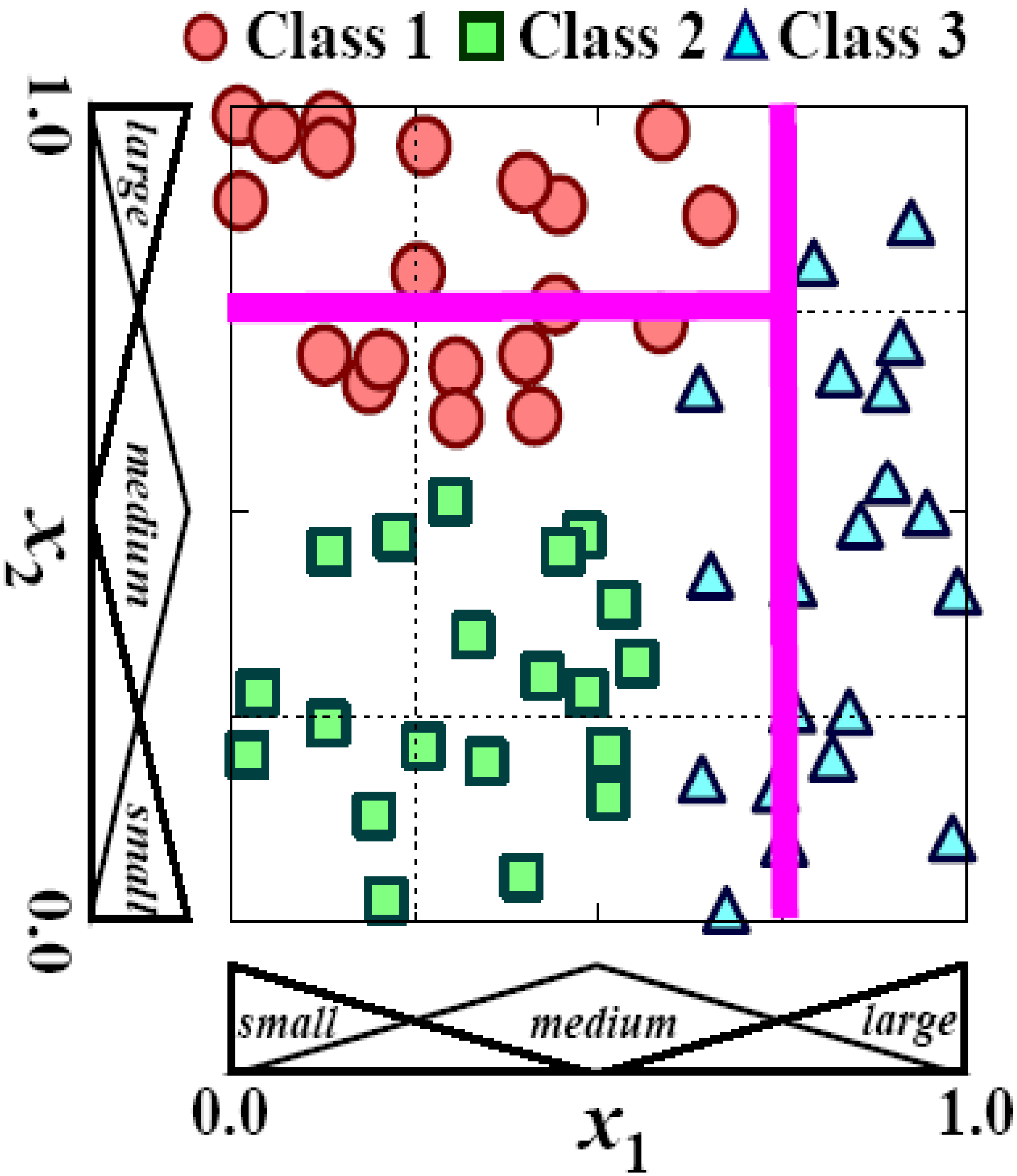
High Interpretability

Easy to Understand !

Fuzzy partition by a simple fuzzy grid cntd.

- *The performance* of a fuzzy classification system based on fuzzy if-then rules depends on the choice of a fuzzy partition.
- If a fuzzy partition is too coarse, the performance may be low.
- If a fuzzy partition is too fine, many fuzzy if-then rules cannot be generated because of the lack of training patterns in the corresponding fuzzy subspaces.
- Therefore the choice of a fuzzy partition is very important.

Basic form does not always have high accuracy



Basic Form

If x_1 is *small* and x_2 is *small*
then Class 2

If x_1 is *small* and x_2 is *medium*
then Class 2

If x_1 is *small* and x_2 is *large*
then Class 1

. . .

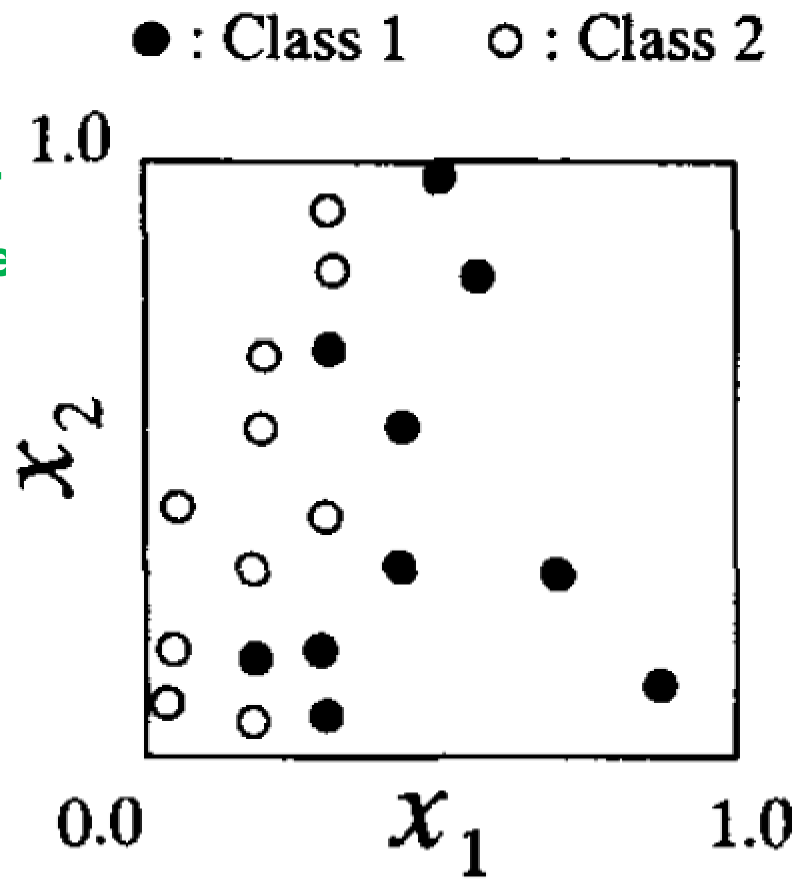
If x_1 is *large* and x_2 is *large*
then Class 3

High Interpretability

Low Accuracy

Difficulties in grid based partitioning

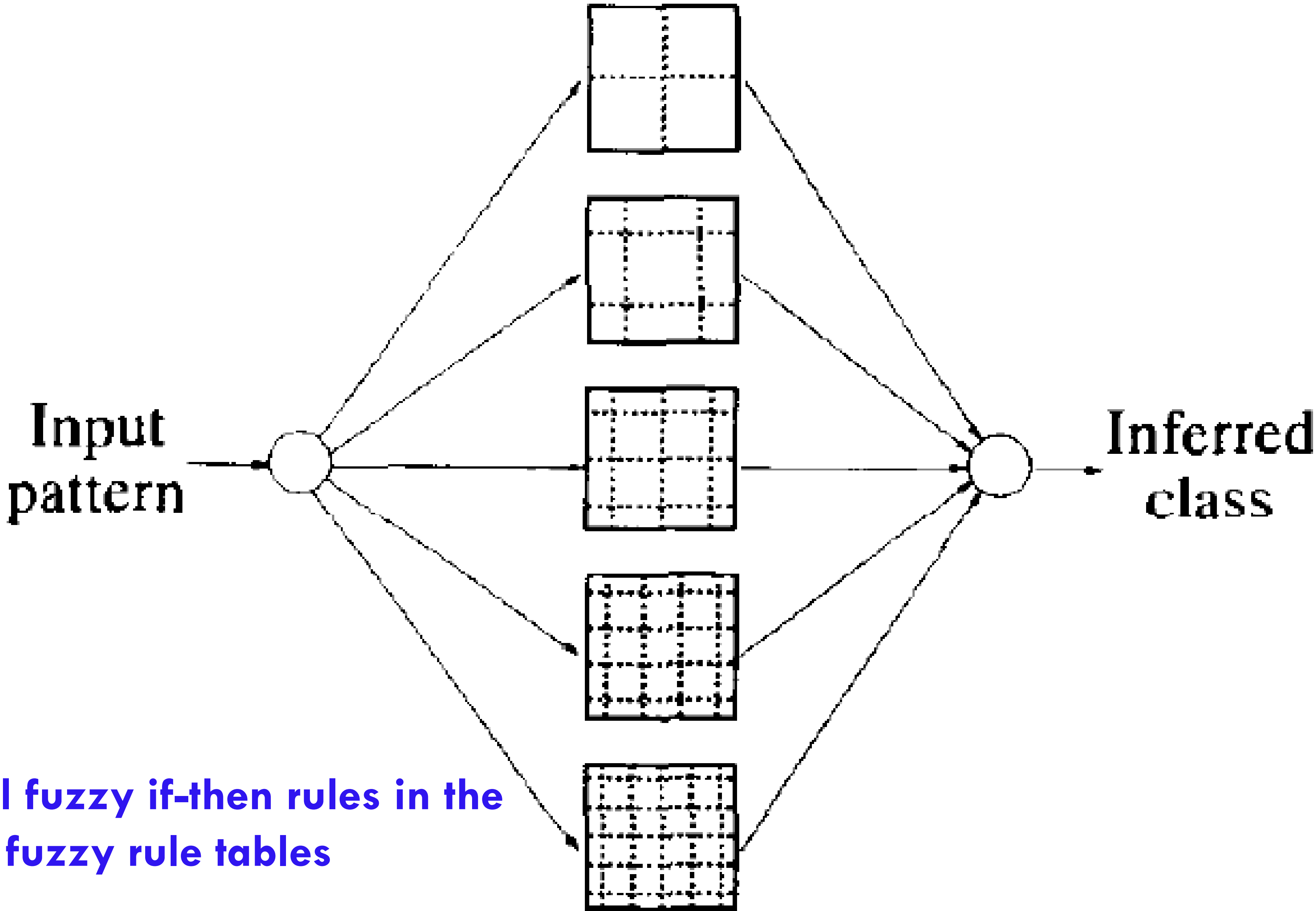
- Consider a two-class classification problem.
- For this classification problem, a fine fuzzy partition is required for the left half of the pattern space but a coarse fuzzy partition is appropriate for the right half.
- Therefore the choice of an appropriate fuzzy partition based on a simple fuzzy grid is difficult for such a classification problem.



An approach to overcome difficulties in grid based partitioning

- To cope with this difficulty, the concept of distributed fuzzy if-then rules is used, where all fuzzy if-then rules corresponding to several fuzzy partitions were simultaneously employed in fuzzy inference.
- That is, multiple fuzzy rule tables were simultaneously employed in a single fuzzy classification system.

Fuzzy classification system based on multiple fuzzy rule tables



Total fuzzy if-then rules in the five fuzzy rule tables

$= 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = 90$

Multiple fuzzy rule tables cntd.

- The fuzzy if-then rules corresponding to coarse fuzzy partitions as well as fine fuzzy partitions are simultaneously employed in a single fuzzy classification system, this approach remedies the difficulty in choosing an appropriate fuzzy partition.
- The main drawback of this approach is that the number of fuzzy if-then rules becomes enormous especially for classification problems with high-dimensional pattern spaces.

Need of reduction of number of rules

- **Unnecessary/redundant/less significant fuzzy if-then rules should be removed and relevant fuzzy if-then rules should be selected , to have better performance with few selected rule set.**
- **A compact fuzzy classification system based on a small number of fuzzy if-then rules has the following advantages:**
  1. It does not require a lot of storage.
  2. The inference speed for new patterns is high.
  3. Each fuzzy if-then rule can be carefully examined by user.

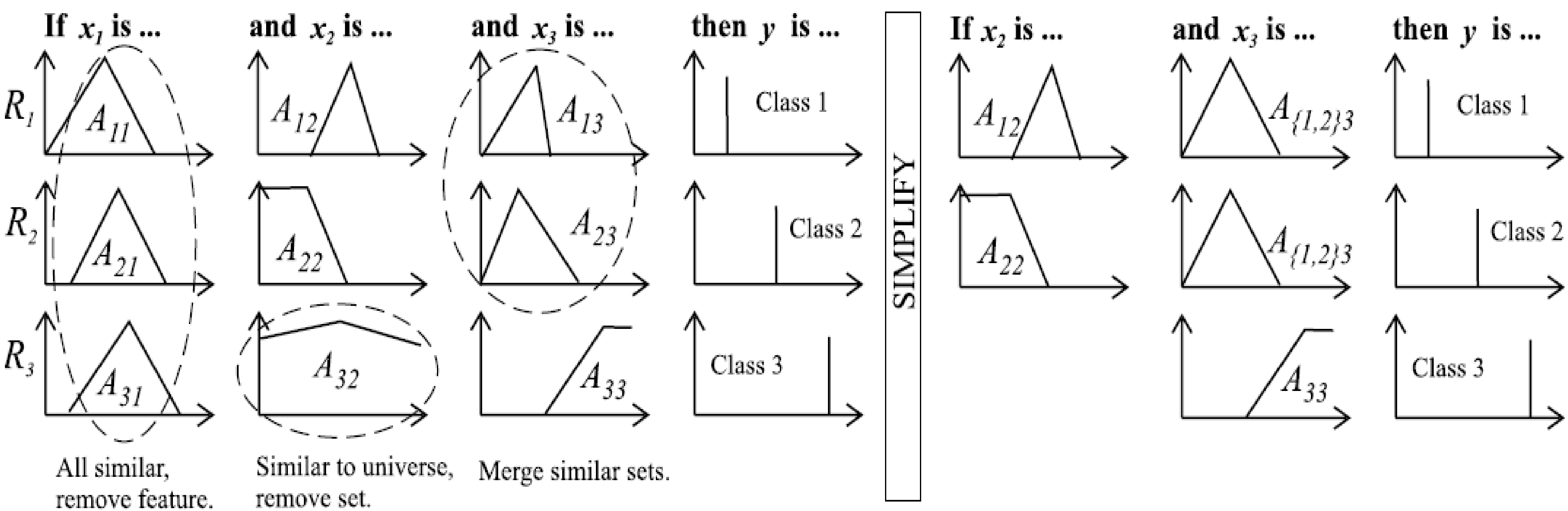
Model reduction

- **Similarity-driven rule-base simplification** $S(A,B) = \frac{|A \cap B|}{|A \cup B|}$

  If S(A,B) =1, the two membership functions are equal.

  If S(A,B) =0, the two membership functions are non-overlapping.

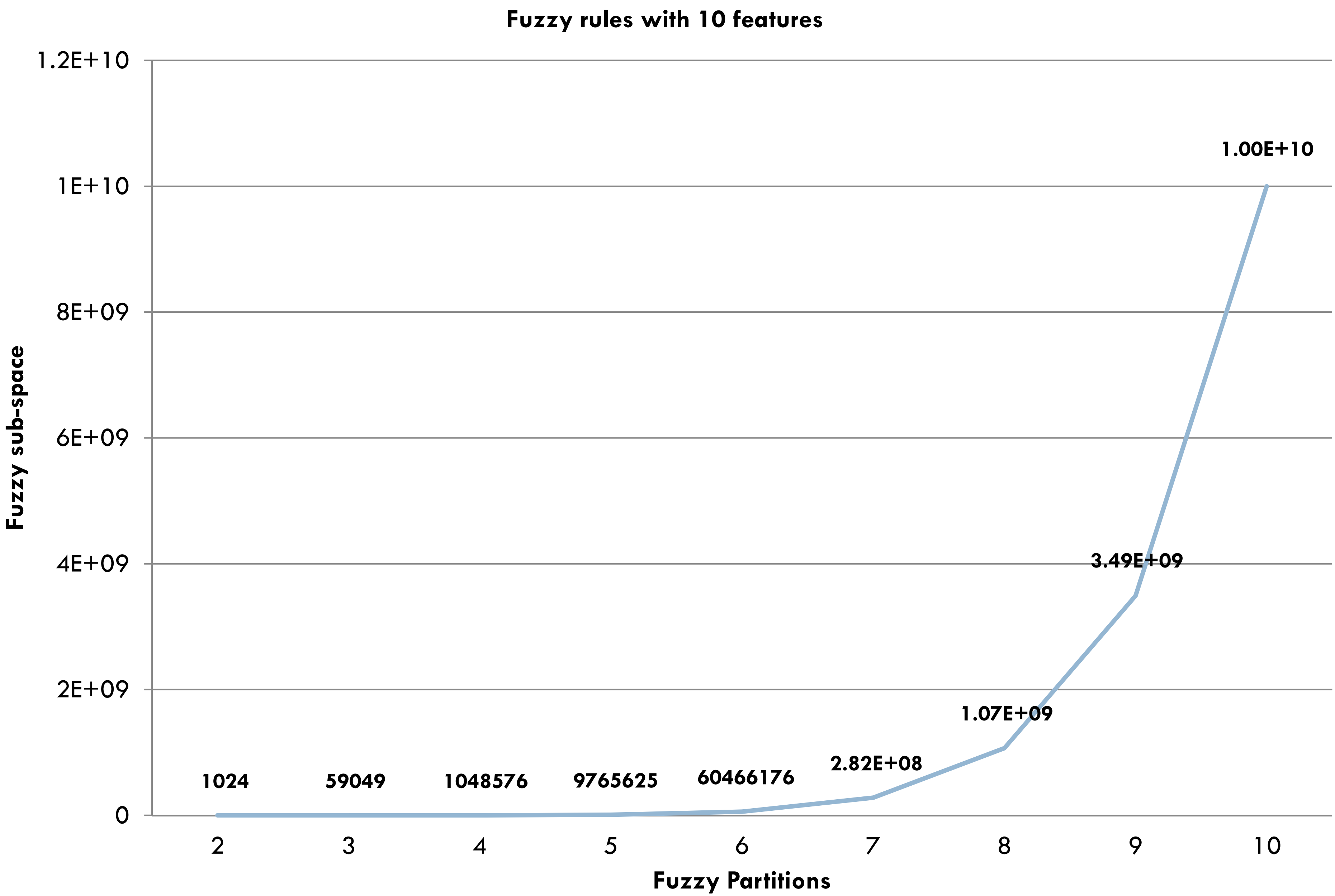
  - Fuzzy sets are merged when their similarity exceeds a user defined threshold
  - If all the fuzzy sets for a feature are similar to the universal set U, then this feature is eliminated



Fuzzy sub-space

Partition	Feature								
	2	3	4	5	6	7	8	9	10
2	4	8	16	32	64	128	256	512	1024
3	9	27	81	243	729	2187	6561	19683	59049
4	16	64	256	1024	4096	16384	65536	262144	1048576
5	25	125	625	3125	15625	78125	390625	1953125	9765625
6	36	216	1296	7776	46656	279936	1679616	10077696	60466176
7	49	343	2401	16807	117649	823543	5764801	40353607	2.82E+08
8	64	512	4096	32768	262144	2097152	16777216	1.34E+08	1.07E+09
9	81	729	6561	59049	531441	4782969	43046721	3.87E+08	3.49E+09
10	100	1000	10000	100000	1000000	10000000	1E+08	1E+09	1E+10

Growth sub-space with 10 features



Feature selection

Importance of feature selection

Ex. With 5 partitions and 10 features, rule size = 97,65,625

With 5 partitions and 9 features, rule size = 19,53,125

Reduction of a single feature here, reduces rule size by = 78,12,500

Need of Multiobjective Optimization

- Considering the complexity of the scenario, we expect a model with minimum number of rules with maximum classification accuracy.
- Further for comprehensibility, it is expected that the rule length (i.e. the number of antecedent conditions) should be minimum.
- This leads to multiple objectives optimization problem.
- Problem definition:

Maximize NCP(S1) and Minimize |S1| and Minimize |antecedent(S1)|

subject to *selected set of rules , S1 belongs to set of total rules, S*

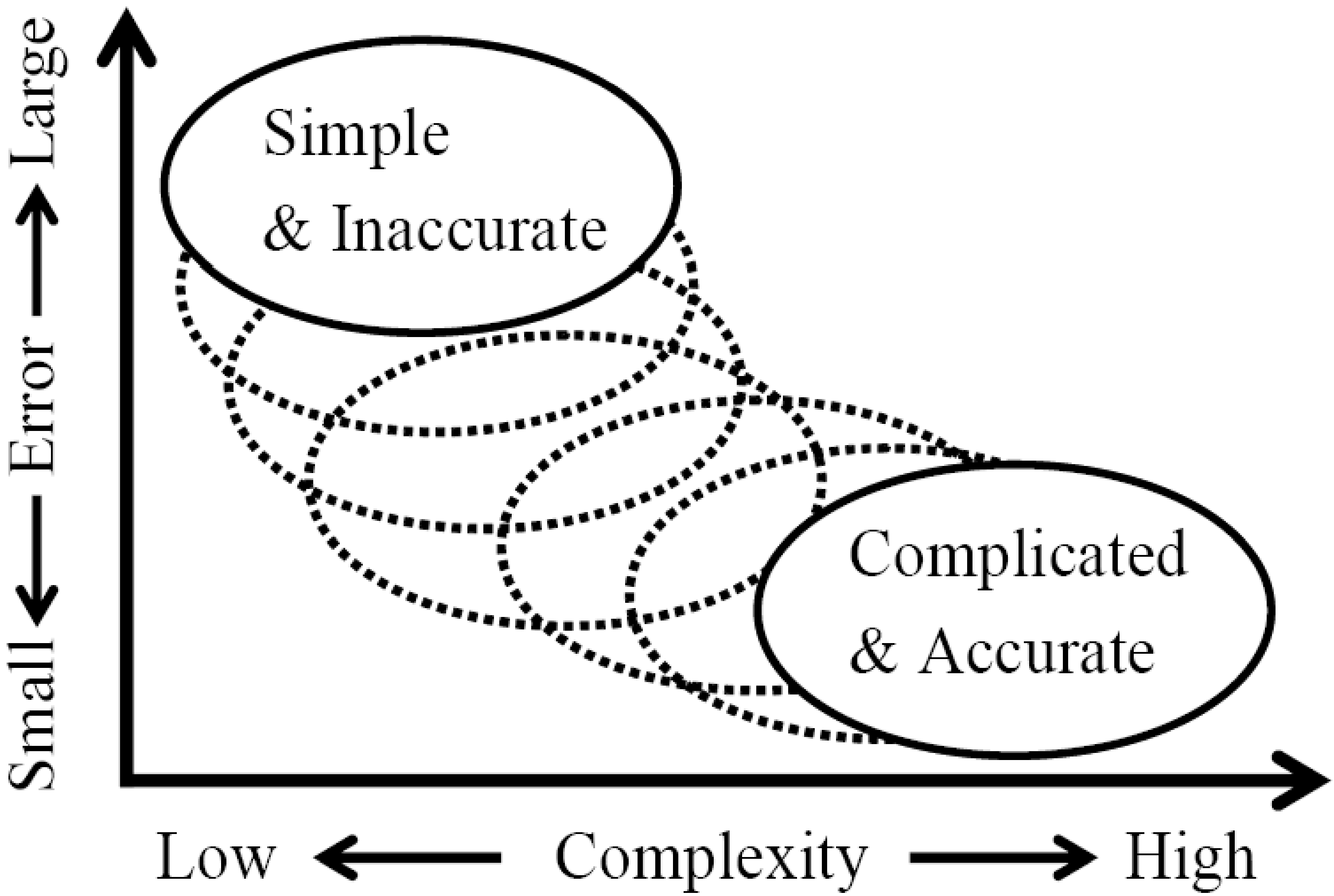
where NCP(S1) is the number of correctly classified patterns by *S1* *and |S1| is the cardinality of S1 (i.e., the number of fuzzy* if-then rules in S1) and |antecedent(S1)| *is the number of antecedent conditions in S1*.

Use of Evolutionary Algorithms

- Evolutionary algorithms is used to perform discrete optimization such as input selection, rule generation, rule selection and fuzzy partition.
- Learning tasks can be viewed as the following optimization problem:
  1. Maximize *Accuracy(S)*: where *S is a fuzzy system, and Accuracy(S) is an accuracy* measure (e.g., classification rate). or
  2. Optimize *f(S) = f(Accuracy(S), Interpretability(S)), or*
  3. Maximize Accuracy(S) and minimize $\text{Complexity}_1(S)$ and $\text{Complexity}_2(S)$.

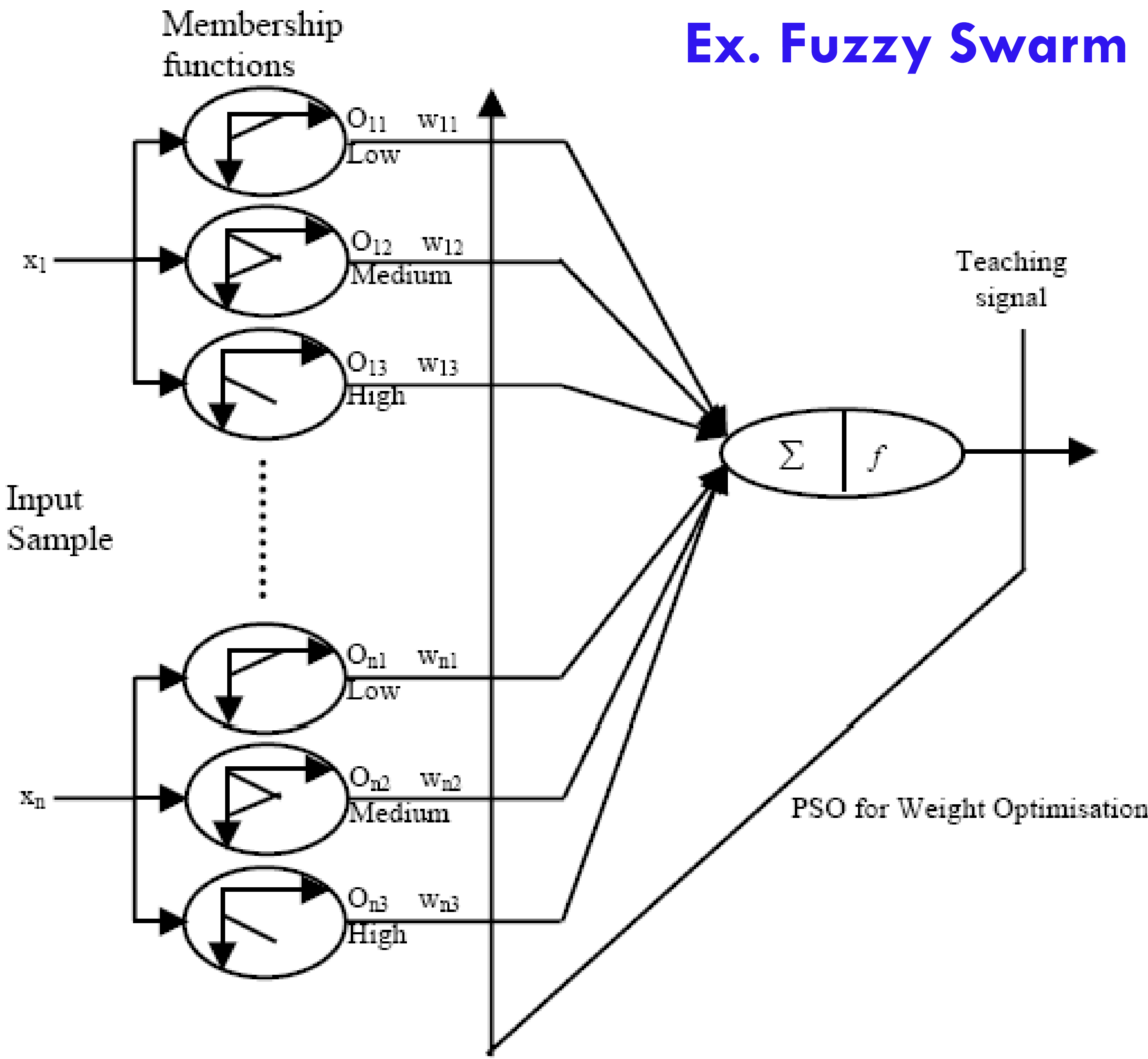
- Rule evaluation criteria : gain, variance, chi-squared value, entropy gain, gini, laplace, lift, and conviction.

Non-dominated fuzzy systems along the accuracy-complexity tradeoff curve



A different approach to use fuzzy systems for classification

Ex. Fuzzy Swarm Net Classifier



Description of the features of the databases employed

	Number of Patterns	Number of Attributes	Number of Classes	Number of Patterns in Class1	Number of Patterns in Class2
Pima Indian Diabetes	768	8	2	500	268
Bupa Liver Disorders	345	6	2	145	200
WBC	699	10	2	458	241

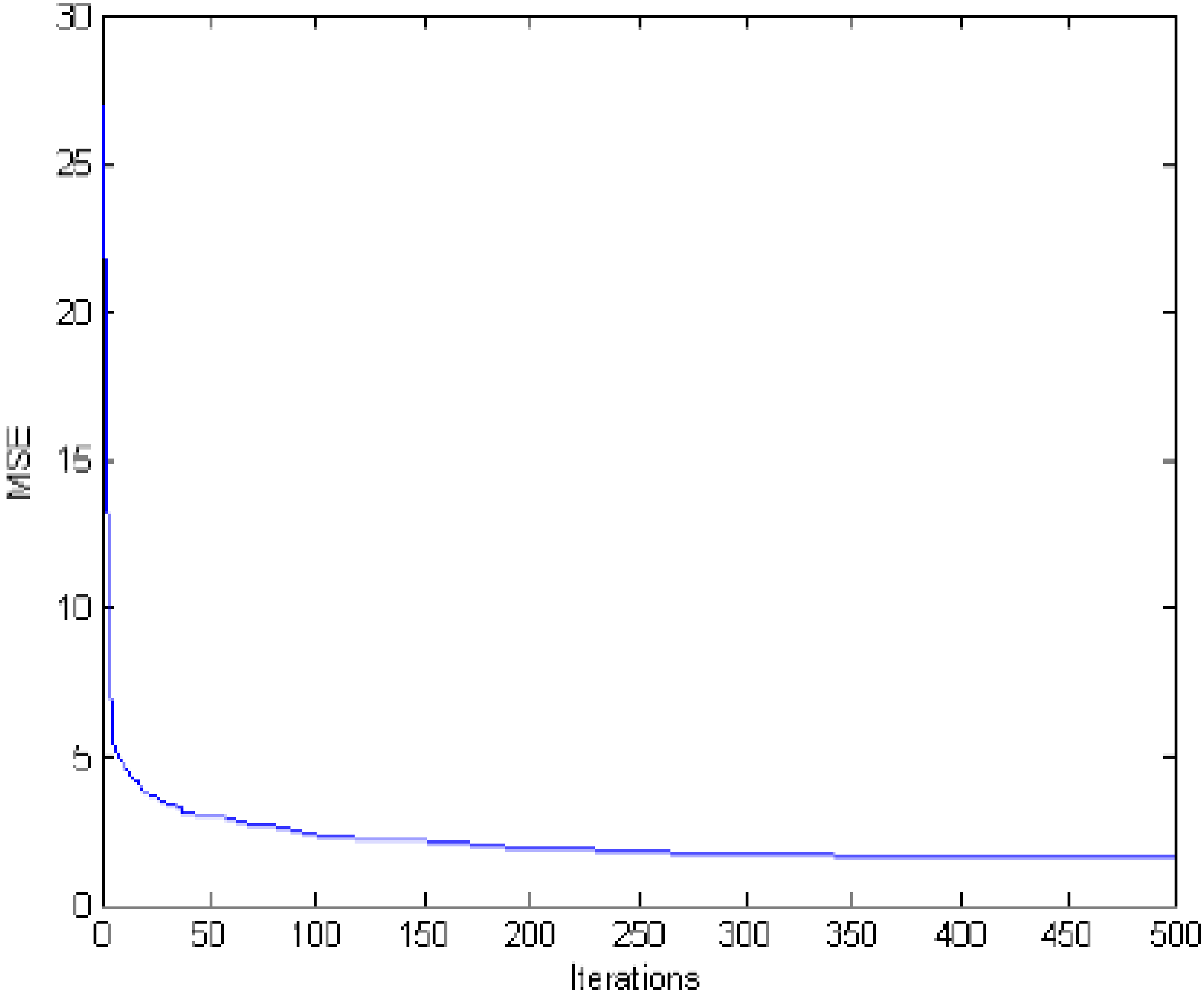
Results obtained with the FSN model for classification

Data set used for testing	Hit Percentage in the training set	Hit percentage in the test set
WBC1.dat *WBC2.dat* *Average WBC*	97.0858 97.9656 97.5257	95.5714 97.1346 **96.353**
pima1.dat *pima2.dat* *Average PIMA*	81.0156 79.4532 80.2344	75.9376 75.1302 **75.5309**
liver1.dat *liver2.dat* *Average LIVER*	75.3488 76.9368 76.1428	70.1745 68.1502 **69.1476**

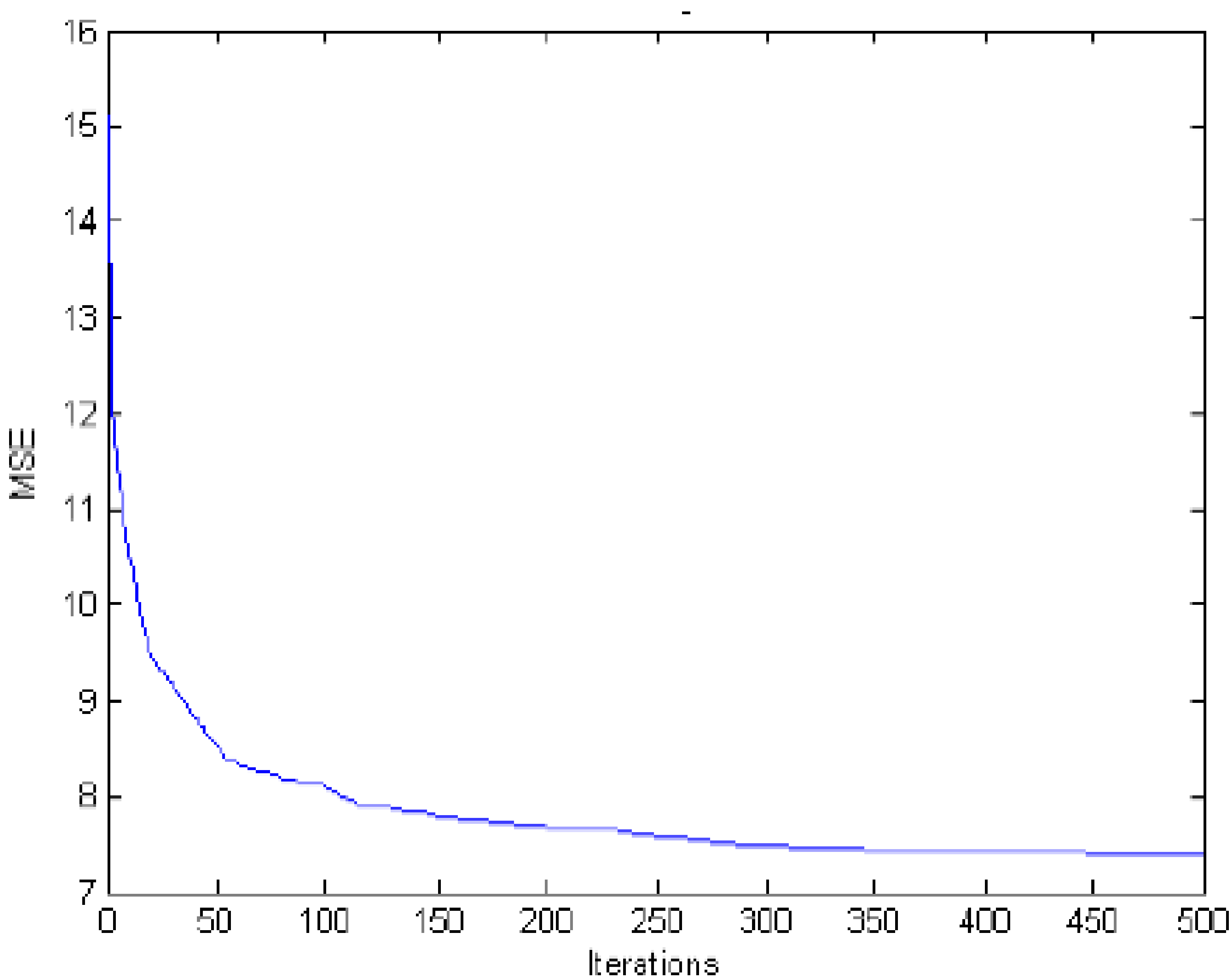
Standard Deviation of 50 simulations

Data set used for testing	Standard Deviation in the training set	Standard Deviation in the test set
WBC1.dat *WBC2.dat* *Average WBC*	0.2256 0.3432 0.2844	0.4312 0.3307 **0.381**
pima1.dat *pima2.dat* *Average PIMA*	0.4331 0.7613 0.5972	1.2351 0.7287 **0.9819**
liver1.dat *liver2.dat* *Average LIVER*	1.1693 0.4265 0.7979	0.9901 1.3746 **1.1824**

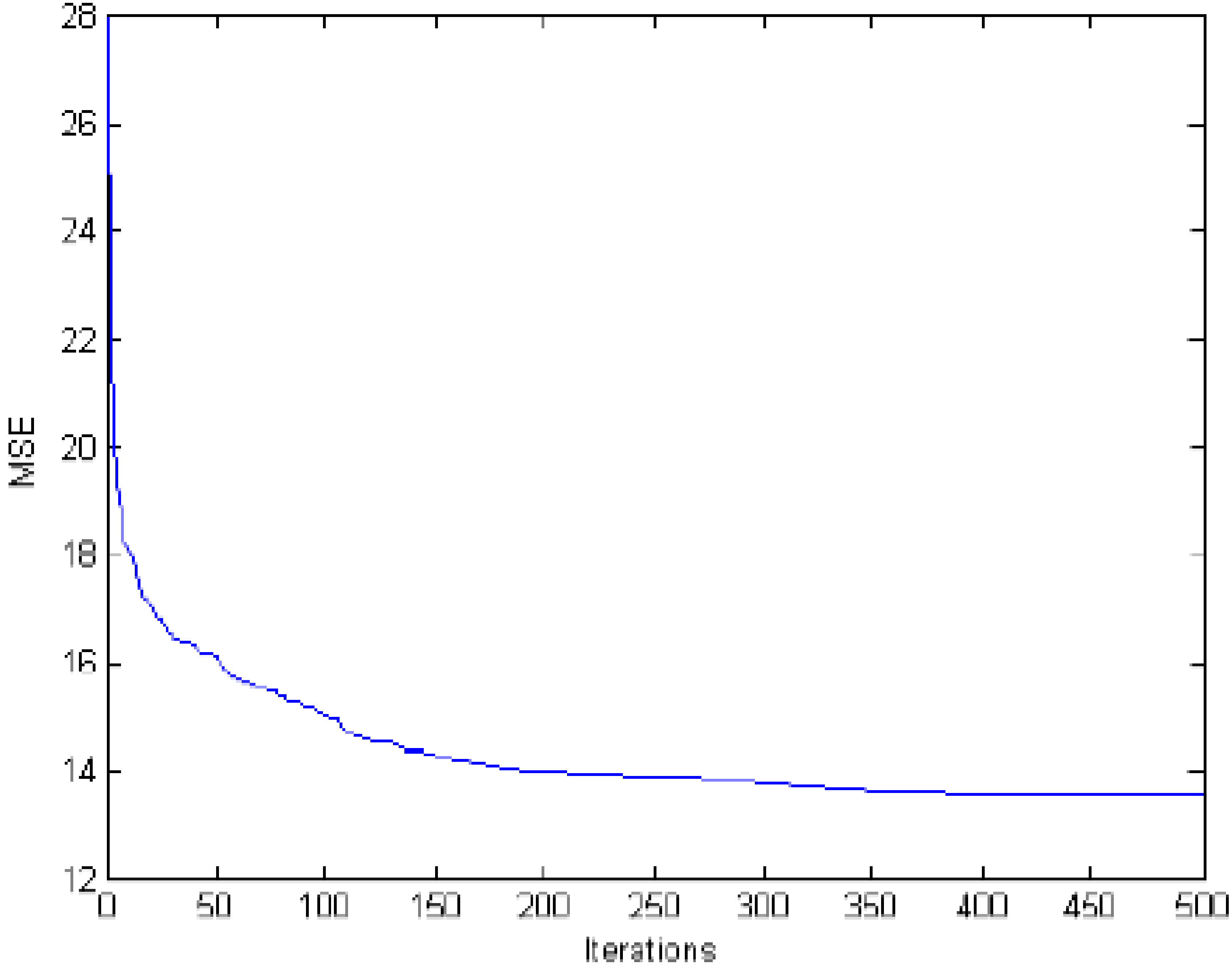
Error curve for training WBC database



Error curve for training BUPA database



Error curve for training PIMA database



References

- Hisao Ishibuchi, Takashi Yamamoto, “Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining,” Fuzzy Sets and Systems 141 (2004) 59–88.
- J, Casillas, O. Cordon, M. J. del Jesus, and F. Herrera, “Genetic Feature Selection in a Fuzzy Rule-Based Classification System Learning Process for High Dimensional Problem,” Technical Report #DECSAI-000122, November-2000.
- H. Ishibuchi, “Multiobjective genetic fuzzy systems: Review and future research directions,” Proc. of 2007 IEEE International Conference on Fuzzy Systems, pp. 913-918, London, UK, July 23-26, 2007.
- H. Ishibuchi and T. Yamamoto, “Rule weight specification in fuzzy rule-based classification systems,” IEEE Trans. on Fuzzy Systems, vol. 13, no. 4, pp. 428-435, August 2005.
- R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification, John Wiley & Sons, Inc., 2001.
- J. A. Roubos, M. Setnes, and J. Abonyi “Learning fuzzy classification rules from labeled data,” Information Sciences 150 (2003) 77–93.
- B. B. Misra, S. Dehuir, G. Panda, P.K. Dash, "Fuzzy Swarm Net (FSN) for Classification in Data Mining," the CSI Journal on Computer Science and Engineering, Vol. 5, No. 2&4(b), pp. 1-8, 2007.

?

Thank

You