Fuzzy Logic & Data Processing

Lecture notes for Modern Method of Data Processing (CCOD)
in 2014

Akira Imada
Brest State Technical University, Belarus

(last modified on)

December 6, 2014

(still under construction)

PART I

Fuzzy Set Arithmetics

1 Fuzzy Set Theory

1.1 Fuzzy set vs. Crisp set

- Examples of crisp set
  - ⋆ $0 < x < 10$
  - ⋆ x=12
- Examples of fuzzy set
  - ⋆ $\{x$ is much smaller than $10\}$
  - ⋆ {x is close to 12}
  - ⋆ Beer is either of {very-cold, cold, not-so-cold, warm}

1.1.1 Membership function

How x is likely to be A is expressed by a function called *membership function.* Usually it is described as $\mu_A(x)$.

For example, a possible membership function for a fuzzy expression {x is close to 12} will be

$$\mu(x) = \frac{1}{1 + (x - 12)^2} \tag{1}$$

See Figure 1.

1.1.2 AND and OR in Fuzzy Logic

In the logic of crisp set *A and B* and *A or B* are defined as in Figure 3.
In Fuzzy Logic, on the other hand, the membership function of *A and B* and *A or B* are specified in various way, but most popular ones are:

$$\mu_{A\cap B}(x) = \min\{\mu_A(x), \mu_B(x)\} \tag{2}$$

and

$$\mu_{A\cup B}(x) = \max\{\mu_A(x), \mu_B(x)\}, \tag{3}$$

respectively.

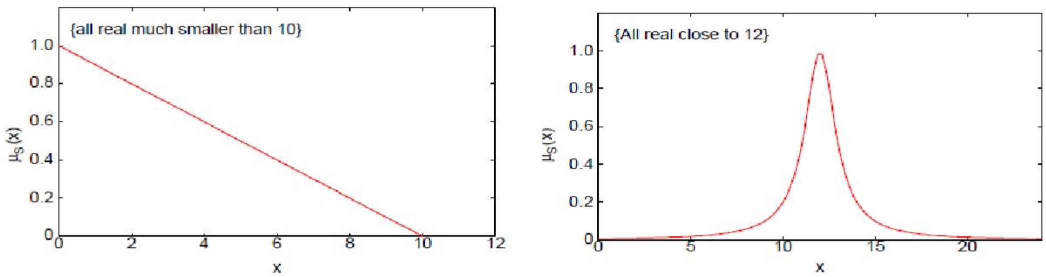


Figure 1: Examples of membership function $\{x$ is much smaller than $10\}$ (right) and {x is close to 12} (left).

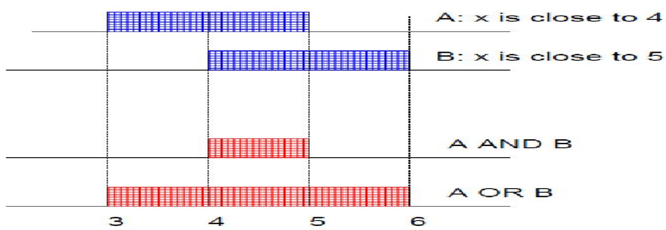


Figure 2: AND and OR in crisp set.

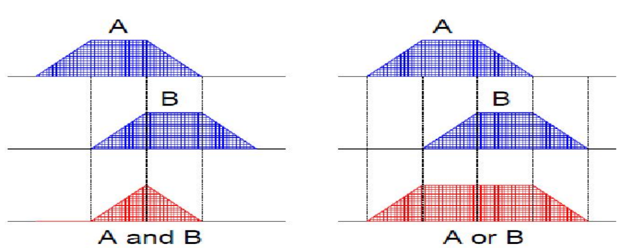


Figure 3: AND and OR in fuzzy set.

To be more concrete the membership function of *x is closer to 4 AND/OR x is closer to 5* is like a Figure 4.

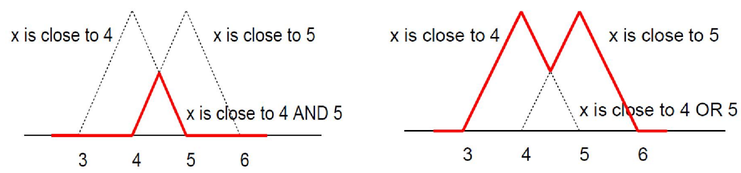
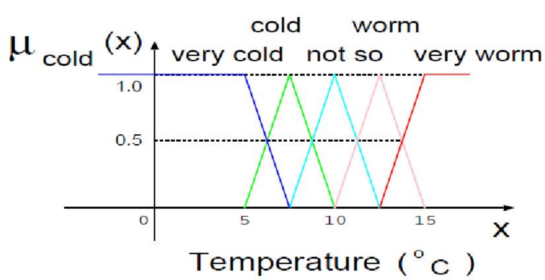


Figure 4: Membership function of *x is closer to 4 OR x is closer to 5*

- **Very cold or pretty cold beer.** ($\mu(x)$ is defined on temperature).

Assume we like very cold beer or pretty cold beer and now we have a beer the temperature of which is 3 degree. Then how is the beer likely to be our prefered one?



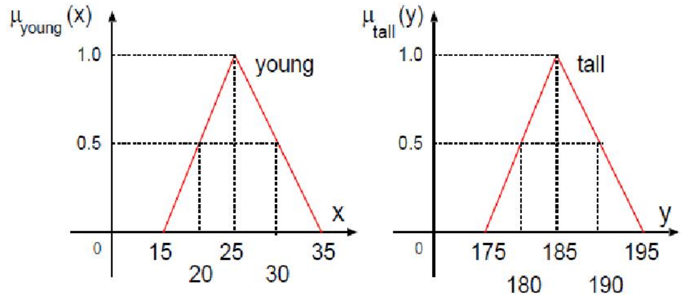
Note that this operation of *OR* was possible because both of the two membership function is defined on the same domain *temperature*. Then what if two membership functions are defined on different domains, such as age and height?

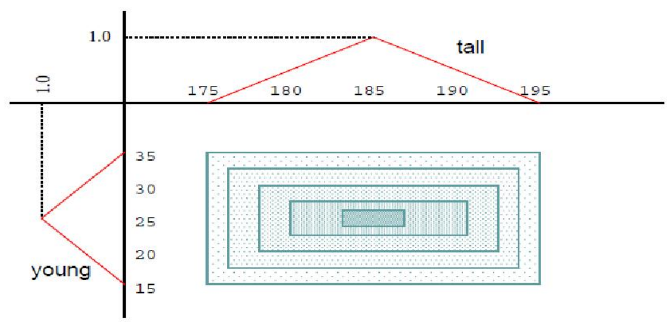
- **Young and tall.**

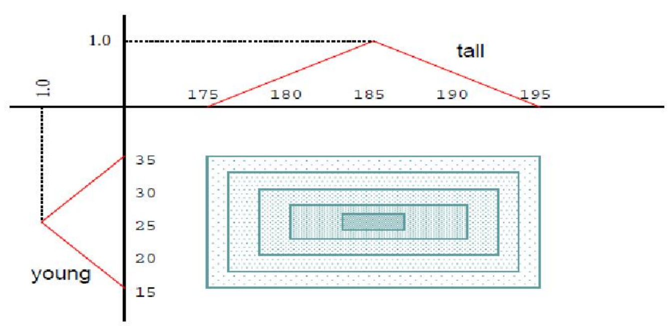
For example,
We cannot draw the membership function of *young and tall* on the 2-dimensional coordinate any more.

(1) 3-D graphic (z = μ is defined on x = age and y = height)

(2) Matrix representation koko


μyoung (x)
1.0
0.5
0
young
x
15
25
35
20
30
μtall (y)
1.0
0.5
0
tall
y
175
185
195
180
190


1.0
tall
1.0
175
180
185
190
195
35
30
25
20
15
young


1.0
tall
1.0
175
180
185
190
195
35
30
25
20
15
young

height \ age	0	10	20	30	40	50	60	70	80	90	100
150											
160											
170											
180											
190											

1.1.3 IF-Then rule in Fuzzy Logic

In Fuzzy Logic, the membership of *IF A Then B* is specified also in many way. Here, let's take it as

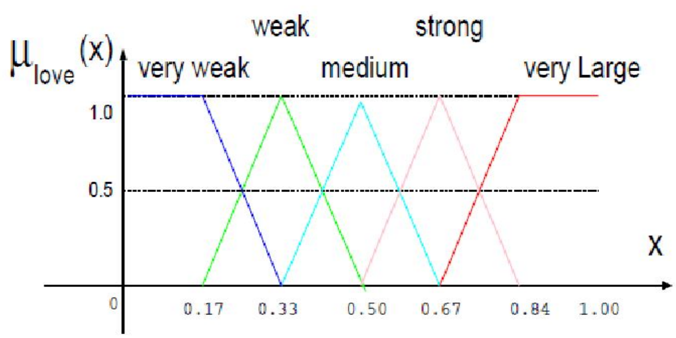
⋆ Mamdani's proposal

$$\mu_{A\to B}(x) = \min\{\mu_A(x), \mu_B(x)\} \tag{4}$$

⋆ Larsen's proposal

$$\mu_{A\to B}(x) = \mu_A(x) \times \mu_B(x) \tag{5}$$

- **If he is young then my love to him is strong.**



- **If he is young and tall then my love to him is very strong.**

1.2 How to express multidimensional membershipfunction

PART II

Fuzzy Controller

2 Fuzzy Controller

Let's construct a virtual metro and control trains by fuzzy controller.

A goal

We now assume x is speed of my car, y is distance to the car in front, and z is how strongly we push brake-pedal. Then let's controll my car with a set of rules, like

- IF x is *high* AND y is *short* THEN z should be *strong*
- IF x is *medium* AND y is *long* THEN z should be *medium*
- IF x is *medium low* or x is *medium* AND y is *long* THEN z should be *weak*
- IF x is *low* or x is *medium low* AND y is *short* or y is *medium hort* THEN z should be *medium weak*
- etc.

Then the results will be plotted like in the Figure below.

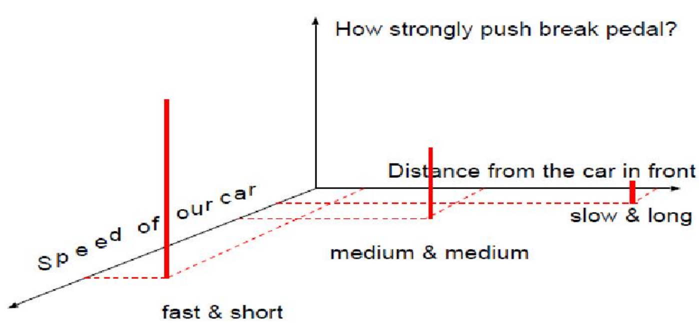


Figure 5: An example of the goal of Fuzzy controller

2.1 Virtual metro system with two trains in a loop line

We study Fuzzy Controllar via a simulation of virutal metro with one loop line on which two Train A and B run. To simplify we don't assume stations. That is, both trains always run. The speed of these trains are denoted as x_A and X_B. The distance from train A to train B is denoted as y_A and from train B to train A is y_B. Note that $x_A + y_B$ is constant (length of the loop line). Speed will be controlled by the distance to the train in front via its break. The shorter the distance, the storonger the break in order to avoid a collision.

3 Vertual metro

Let's create a virtual metro system with 2 cars on a loop line with 1000 pixels in which 4 stations 1, 2, 3 and 4 at pixel number 0, 250, 500 and 750, respectively.

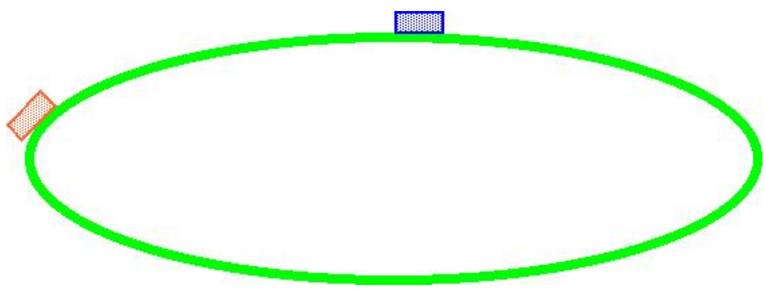


Figure 6: To de-fuzzify strength of break

Exercise 1 *Create your own simulation of metro with one loop on which two trains A and B run, using graphics. 6 parameters x_A, x_B, y_A, y_B, z_A, z_B, should also be desplayed on the screen. The simulation might be started with $x_A = x_B$, $y_A = y_B$, $z_A = z_B = 0$.*

3.1 Let's design a set of rules for driving a train

3.1.1 membership function

E.g., when we say "speed is 13," it is likely to be fast with 75% certain and medium with 25% .

Figure 7: An example of 5 membership function for speed.

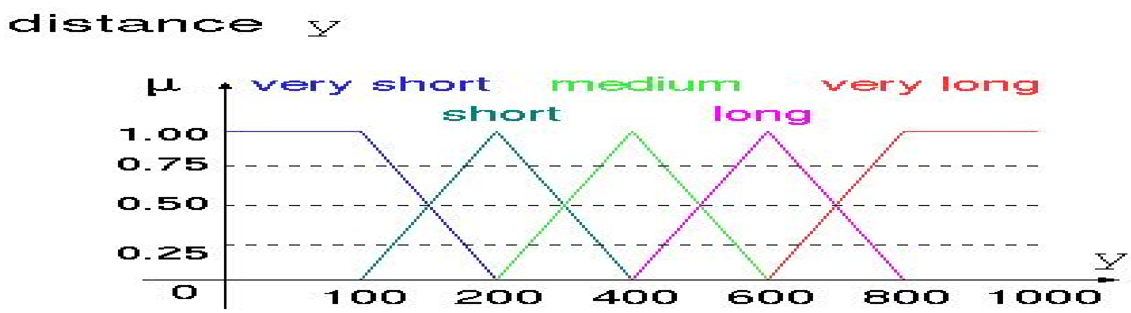


Figure 8: An example of 5 membership function for speed.

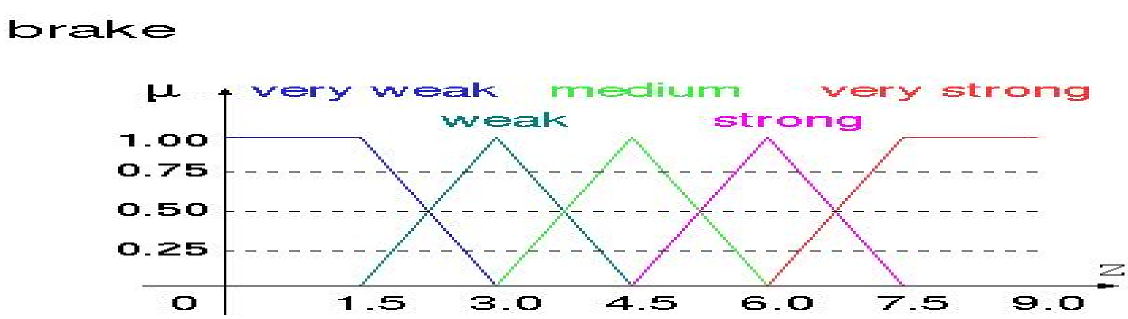


Figure 9: An example of 5 membership function for speed.

3.1.2 One point of break

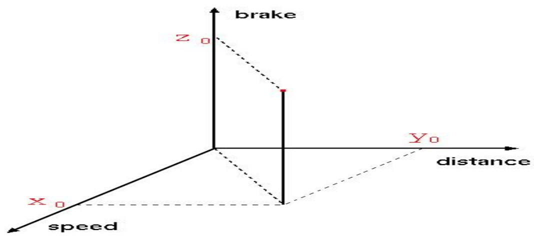


Figure 10: An example of 5 membership function for speed.

3.1.3 One plaine of break

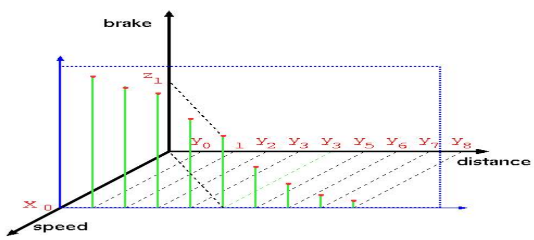


Figure 11: An example of 5 membership function for speed.

3.1.4 3-D control of break

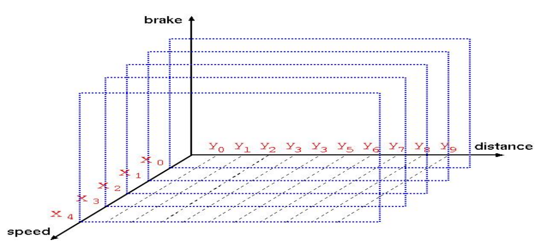


Figure 12: An example of 5 membership function for speed.

3.1.5 To calculate one point of break value for fixed speed and distance under one rule

We now assume our rule is

IF speed is fast AND distance is short THEN break is strong

Then

speed	μ	distance	μ	break	μ	total μ
13	0.75	125	0.25	0	0.00	0.00
13	0.75	125	0.25	1	0.00	0.00
13	0.75	125	0.25	2	0.00	0.00
13	0.75	125	0.25	3	0.00	0.00
13	0.75	125	0.25	4	0.00	0.00
13	0.75	125	0.25	5	0.33	0.08
13	0.75	125	0.25	6	1.00	0.25
13	0.75	125	0.25	7	0.33	0.08
13	0.75	125	0.25	8	0.00	0.00
13	0.75	125	0.25	9	0.00	0.00

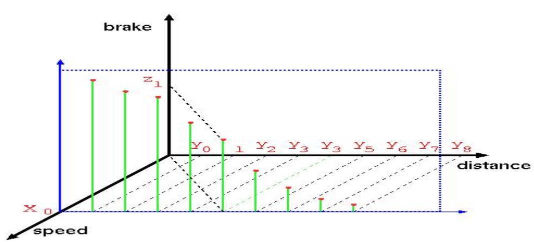


Figure 13: An example of 5 membership function for speed.

We now know the center of gravity of break is at at 6. So the point is (13, 125, 6).

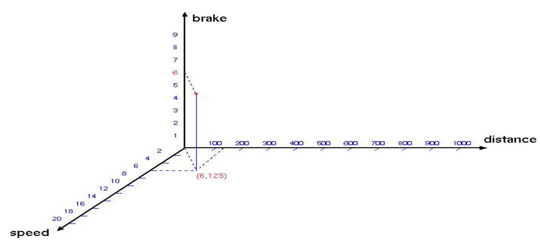


Figure 14: An example of 5 membership function for speed.

3.1.6 To calculate one point of break value for fixed speed and distance under two rules

Rule 1							Rule 2							Total
x	μ	y	μ	z	μ	$\mu 1$	x	μ	y	μ	z	μ	$\mu 2$	μ-final
15	0.25	80	0.25	0	0.25	0.25	15	0.25	80	0.25	0	0.25	0.25	0.25
15	0.25	80	0.25	1	0.25	0.25	15	0.25	80	0.25	1	0.25	0.25	0.25
15	0.25	80	0.25	2	0.25	0.25	15	0.25	80	0.25	2	0.25	0.25	0.25
15	0.25	80	0.25	3	0.25	0.25	15	0.25	80	0.25	3	0.25	0.25	0.25
15	0.25	80	0.25	4	0.25	0.25	15	0.25	80	0.25	4	0.25	0.25	0.25
15	0.25	80	0.25	5	0.25	0.25	15	0.25	80	0.25	5	0.25	0.25	0.25
15	0.25	80	0.25	6	0.25	0.25	15	0.25	80	0.25	6	0.25	0.25	0.25
15	0.25	80	0.25	7	0.25	0.25	15	0.25	80	0.25	7	0.25	0.25	0.25
15	0.25	80	0.25	8	0.25	0.25	15	0.25	80	0.25	8	0.25	0.25	0.25
15	0.25	80	0.25	9	0.25	0.25	15	0.25	80	0.25	9	0.25	0.25	0.25

3.1.7 To calculate one point of break value for fixed speed and distance under ten rules

Rule 1				Rule 2				⋯	Rule 10				Total
x	y	z	μ	x	y	z	μ	⋯	x	y	z	μ	μ
15	250	0	0.84	15	250	0	0.52	⋯	15	250	0	0.7	0.7
15	250	1	0.84	15	250	1	0.52	⋯	15	250	1	0.7	0.7
15	250	2	0.84	15	250	2	0.52	⋯	15	250	2	0.7	0.7
15	250	3	0.84	15	250	3	0.52	⋯	15	250	3	0.7	0.7
15	250	4	0.84	15	250	4	0.52	⋯	15	250	4	0.7	0.7
15	250	5	0.84	15	250	5	0.52	⋯	15	250	5	0.7	0.7
15	250	6	0.84	15	250	6	0.52	⋯	15	250	6	0.7	0.7
15	250	7	0.84	15	250	7	0.52	⋯	15	250	7	0.7	0.7
15	250	8	0.84	15	250	8	0.52	⋯	15	250	8	0.7	0.7
15	250	9	0.84	15	250	9	0.52	⋯	15	250	9	0.7	0.7

Rule 1				Rule 2				Rule 3				Rule 4				Rule 5			
S	D	B	μ1	S	D	B	μ2	S	D	B	μ3	S	D	B	μ4	S	D	B	μ5
15	200	0	0	15	200	0	0.1	15	200	0	0	15	200	0	0	15	200	0	0
15	200	1	0	15	200	1	0.03	15	200	1	0.01	15	200	1	0	15	200	1	0
15	200	2	0	15	200	2	0.05	15	200	2	0.02	15	200	2	0	15	200	2	0
15	200	3	0	15	200	3	0.03	15	200	3	0.03	15	200	3	0	15	200	3	0.15
15	200	4	0	15	200	4	0.01	15	200	4	002	15	200	4	0	15	200	4	0.25
15	200	5	0	15	200	5	0	15	200	5	0.01	15	200	5	0.02	15	200	5	0.35
15	200	6	0	15	200	6	0	15	200	6	0	15	200	6	0.04	15	200	6	0.25
15	200	7	0	15	200	7	0	15	200	7	0	15	200	7	0.06	15	200	7	0.15
15	200	8	0	15	200	8	0	15	200	8	0	15	200	8	0.04	15	200	8	0
15	200	9	0	15	200	9	0	15	200	9	0	15	200	9	0.02	15	200	9	0

Rule 6				Rule 7				Rule 8				Rule 9				Rule 10				
S	D	B	μ6	S	D	B	μ7	S	D	B	μ8	S	D	B	μ9	S	D	B	μ10	μ
15	200	0	0.05	15	200	0	0	15	200	0	0	15	200	0	0	15	200	0	0	0.1
15	200	1	0.025	15	200	1	0	15	200	1	0	15	200	1	0	15	200	1	0.01	0.03
15	200	2	0	15	200	2	0	15	200	2	0.1	15	200	2	0	15	200	2	0.02	0.1
15	200	3	0	15	200	3	0	15	200	3	0.5	15	200	3	0	15	200	3	0.03	0.5
15	200	4	0	15	200	4	0	15	200	4	0.8	15	200	4	0.1	15	200	4	0.02	0.8
15	200	5	0	15	200	5	0.01	15	200	5	0.5	15	200	5	0.25	15	200	5	0.01	0.5
15	200	6	0	15	200	6	0.02	15	200	6	0.1	15	200	6	0.5	15	200	6	0	0.5
15	200	7	0	15	200	7	0.03	15	200	7	0	15	200	7	0.25	15	200	7	0	0.25
15	200	8	0	15	200	8	0.02	15	200	8	0	15	200	8	0.1	15	200	8	0	0.1
15	200	9	0	15	200	9	0.01	15	200	9	0	15	200	9	0	15	200	9	0	0.02

Figure 15: An example of 5 membership function for speed.

PART III

Fuzzy Classification

4 Classify data by a rule set

Assume we classiy M data to be classified by using N features.

$$x_1, x_2, x_3, \cdots, x_N.$$

A rule such as

$$\text{If } x_1 = A_1 \text{ and } x_2 = A_2, \text{ and } \cdots, \text{ and } x_N = A_N \text{ then } class = \omega_p.$$

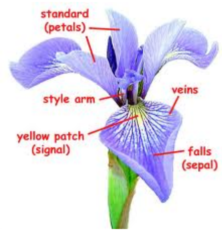
classifies the data to one class ω_p.

A_i is called attribute. For instance, (i) IF $x_i = 30$, (ii) IF $15 < x_i < 20$, (iii) IF x_i is Large, or (iv) IF x_i is Female, etc. The first two are called *crisp*, second is *fuzzy*, and fourth is called *categorical*. Let's take an example.

IF$x_1 = 20g$ AND $10cm < x_2 < 20cm$ AND $x_3 = Green,$ AND $x_4 = Fruits$ THEN this is *apple*.

5 A benchmark – Iris database

As an example target here, we classify Iris flowers. Iris flower dataset[1] is made up of 150 samples consists of three species of iris flower, that is, *setosa, versicolor* and *virginica.* Each of these three families includes 50 samples. Each sample is a four-dimensional vector representing four attributes of the iris flower, that is, *sepal-length, sepal-width, petal-length*, and *petal-width.*



All data are given as crisp as below.

x_1	x_2	x_3	x_4	class
5.1	3.5	1.4	0.2	1 (Setosa)
4.9	3.0	1.4	0.2	1 (Setosa)
4.7	3.2	1.3	0.2	1 (Setosa)
...	...	...	...	...
7.0	3.2	4.1	1.4	2 (Versicolor)
6.4	3.2	4.5	1.5	2 (Versicolor)
6.9	3.1	4.9	1.5	2 (Versicolor)
...	...	...	...	...
5.8	2.7	5.1	1.9	3 (Virginica)
7.1	3.0	5.9	2.1	3 (Virginica)
6.3	2.9	5.6	1.8	3 (Virginica)
...	...	...	...	...

[1] University of California Urvine Machine Learning Repository.
ics.uci.edu: pub/machine-learning-databases.

5.1 Let's visualize data

Each data is a point in 4-dimensional space, i.e., x_1, x_2, x_3 and x_4.

5.1.1 distribution of each of 4 attributes

First, let's see how each x_i $(i = 1, 2, 3, 4)$ of these dataset distributed.

5.1.2 What if we had only two attributes?

5.1.3 Demension reduction from 4 to 2 by Samon mapping

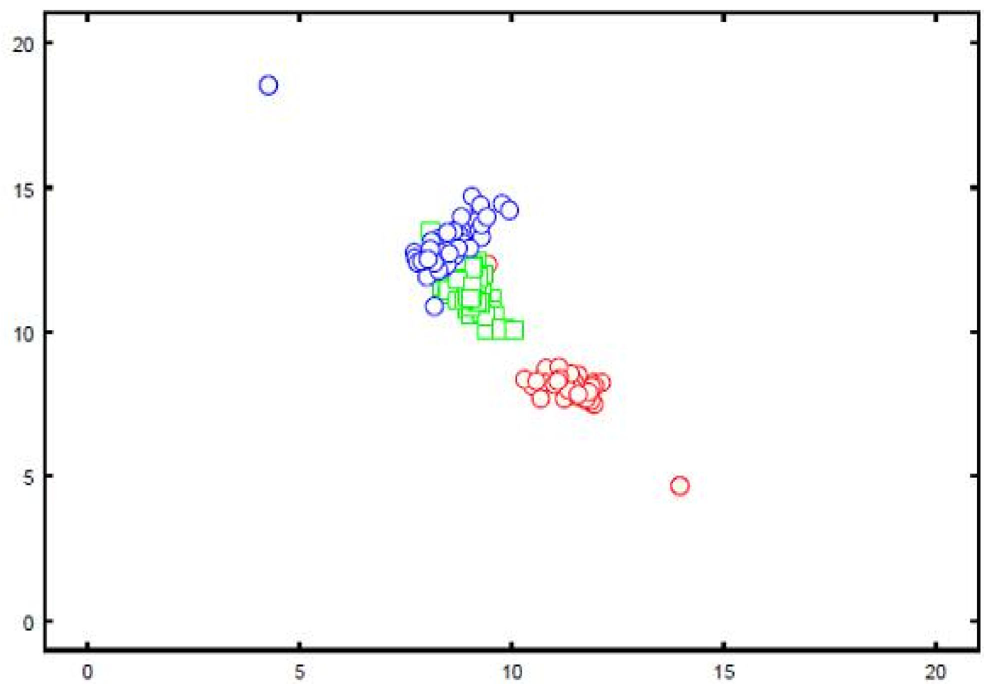


Figure 16: Sammon mapping of iris data.

5.2 Non-fuzzy approach

If

$a_1^1 \leq x_1 < b_1^1$ and $a_2^1 \leq x_2 < b_2^1$ and $a_3^1 \leq x_3 < b_3^1$ and $a_4^1 \leq x_4 < b_4^1$ then class $= 1$.

Else if

$a_1^2 \leq x_1 < b_1^2$ and $a_2^2 \leq x_2 < b_2^2$ and $a_3^2 \leq x_3 < b_3^2$ and $a_4^2 \leq x_4 < b_4^2$ then class $= 2$.

Else

clase $= 3$.

6 TS-Fuzzy Formula

6.1 Singleton Consequence

R_i: If x_1 is A_1^i and x_2 is A_2^i and $\cdots$ and x_n is A_n^i then y is g^i.

6.2 Linear Regression Consequence

R_i: If x_1 is A_1^i and x_2 is A_2^i and $\cdots$ and x_n is A_n^i then $y = a_1^1 x_1 + a_2^1 x_1 + \cdots + a_n^i x_n + b^i$.

6.3 Triangle vs. Gaussian membership function

$$\mu(x) = \exp\{-\frac{(x-c)^2}{w^2}\}$$

6.4 Example - Iris data classification

6.4.1 Singleton Consequence with triangle membership

Apply the TS-fuzzy formula above to the iris flower database, assumming the folloing $p = 3$ rules and membership functions[2].

R_1: IF x_1 is short AND x_2 is long AND x_3 is short AND x_4 is short THEN y = 1.00.
R_2: IF x_1 is medium AND x_2 is small AND x_3 is medium AND x_4 is medium THEN y = 2.10.
R_3: IF x_1 is long AND x_2 is medium AND x_3 is long AND x_4 is long THEN y=2.95.

where each of the membership functions are adjusted as Figure 1 below.

[2]Taken frome the draft of H. Roubos et al. (2001) IEEE Transactions on Fuzzy Systems, Vol. 9, No. 4, pp. 516-524.

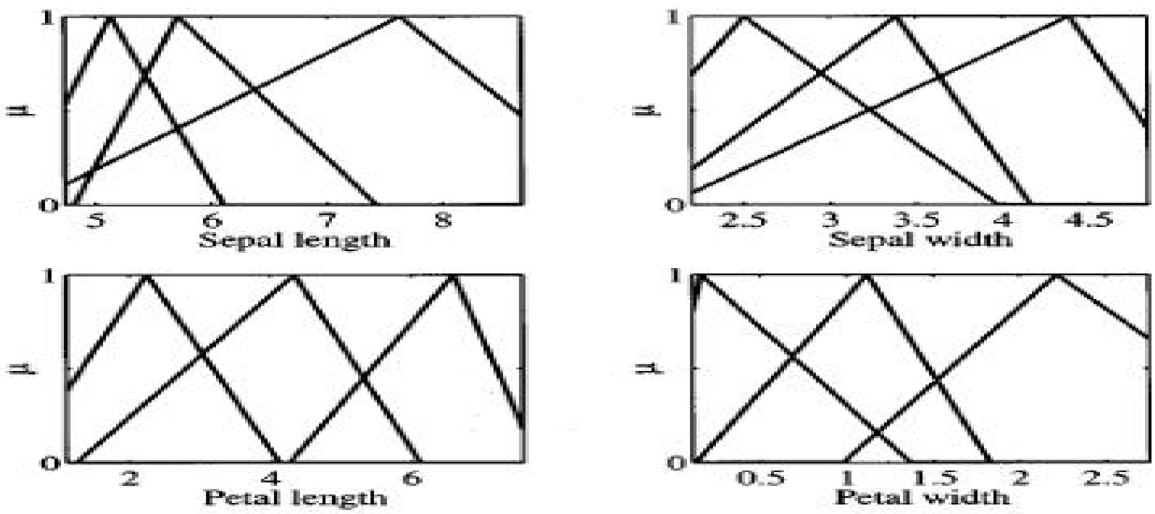


Figure 17: Triangle membership functions representing small, medium and large for x_1 (up left), x_2 (Up right), x_3 (bottom left) and x_4 (bottom left).

x_1	S=(4.2046, 5.3171, 5.9475)	M=(3.7522, 5.9779, 8.2048)	L=(4.3795, 6.5593, 8.2816)
x_2	N=(1.2431, 2.5759, 4.4562)	M=(1.7047, 3.5304, 3.8480)	W=(1.8674, 4.143, 4.935)
x_3	S=(0.3542, 1.061, 4.1049)	M=(0.5027, 3.9518, 5.4715)	L=(4.1304, 6.8883, 7.808)
x_4	N=(−0.3903, −0.0909, 1.097)	M=(0.4008, 1.130, 2.3095)	W=(1.1088, 2.1603, 3.0233)

Figure 18: Data for 12 triangle membership function above, indicating ($start-peak-end$) of each triangle for small, medium and large for each of x_1, x_2, x_3 and x_4.

Apply TS-formula above and then estimated class is:

$$y = \begin{cases} 1 & \text{... if } \hat{y} < 1.5 \\ 2 & \text{... if } 1.5 \leq \hat{y} < 3.0 \\ 3 & \text{... if } 3.0 \leq \hat{y} \end{cases}$$

Exercise 2 *How many y out of 150 data are collect?*

6.4.2 Lenear Regression Consequence with Gaussian membership

Apply the TS-fuzzy formula above to the iris flawer database, assumming the folloing $p = 3$ rules and membership functions[3].

$$\begin{bmatrix} a_1^1 & a_2^1 & a_3^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 & a_3^1 \\ a_1^3 & a_2^3 & a_3^3 & a_3^1 \end{bmatrix} = \begin{bmatrix} -0.0000 & 0.0000 & -0.0001 & 0.0005 \\ -0.1121 & -0.2234 & 0.0029 & 0.0005 \\ -0.1020 & -0.0624 & 0.1276 & 0.0005 \end{bmatrix}$$

and

$$\begin{bmatrix} b^1 \\ b^2 \\ b^3 \end{bmatrix} = \begin{bmatrix} 0.6667 \\ 1.7547 \\ 1.8412 \end{bmatrix}.$$

Gaussian membership function is defiend here as:

$$\mu(x) = \exp\{-\frac{(x-c)^2}{w^2}\}$$

where c and w represent center and width of distribution, respectively.

6.5 Neuro Fuzzy approach

6.5.1 Fuzzy Neural Network Implementation.

The procedure described in the previous sub-subsection can be realized when we assume a neural network architecture such as depicted in Fig. 2. The 1st layer is made up of n input neurons. The 2nd layer is made up of H groups of a neuronal structure each contains n neurons where the i-th neuron of the k-th group has a connection to the i-th neuron in the 1st layer with a synaptic connection which has a pair of weights (w_{ik}, σ_{ik}).

[3]Taken frome M.H. Kim et al. (2004) A novel appreoach to design of Takagi-Sugeno fuzzy classifier. Joint International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems.

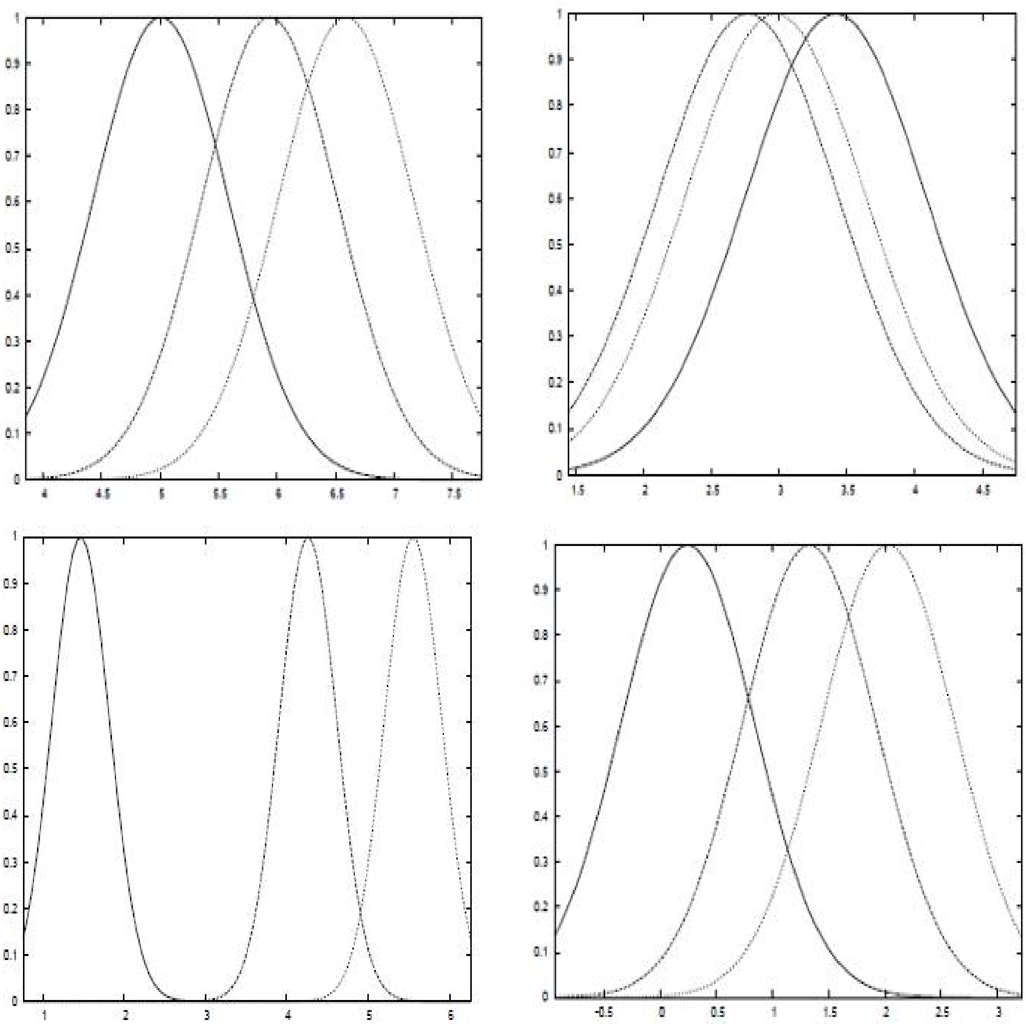


Figure 19: Gaussian membership functions representing small, medium and large for x_1 (up left), x_2 (Up right), x_3 (bottom left) and x_4 (bottom left).

Then k-th group in the second layer calculates the value $\mu_k(\mathbf{x})$ from the values which are received from each of the n neurons in the first layer. The 3rd layer is made up of m neurons each of which collects the H values from the output of the second layer, that is j-th neuron of the 3rd layer receives the value from k-th output in the second layer with the synapse which has the weight ν_{kj}

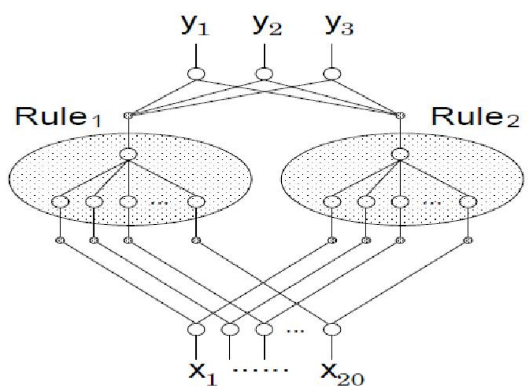


Figure 20: Architecture of the proposed fuzzy neural network which infers how an input $\mathbf{x} = (x_1, \cdots x_n)$ is likely to belong to the j-th class by generating outputs y_j each of which reflect the degree of the likeliness. In this example, a 20-dimension data input will be inferred to which of the 3 classes the input belongs by using 2 rules.

6.5.2 How it learns?

Castellano et al. [**?**] used (i) a competitive learning to determine how many rules are needed under initial weights created at random. Then, in order to optimize the initial random weight configuration, they use (ii) a gradient method performing the steepest descent on a surface in the weight space employing the same training data, that is, supervised learning.

Here, on the other hand, we use a simple genetic algorithm, since our target space is specific enough to know the network structure in advance, i.e., only unique rule is necessary. Our concern, therefore, is just obtaining the solution of weight configuration of the network. That is to say, all we want to know is a set of parameters w_{ik}, σ_{ik} and ν_{kj} $(i = 1, \cdots n)$, $(k = 1, \cdots H)$, $(j = 1, \cdots m)$ where n is the dimension of data, H is the number of rules, and m is the number of outputs. Hence our chromosome has those $n \times H \times m$ genes. Starting with a population of chromosomes whose genes are randomly created, they evolve under *simple truncate selection* where higher fitness chromosome are chosen, with *uniform crossover* and occasional *mutation* by replacing some of a few genes with randomly created other parameters, expecting higher fitness chromosomes will be emerged. These settings are determined by trials and errors experimentally.

6.6 Multi input multi output

PART IV

Time series data forecasting with TS-Fuzzy formula

7 Two different formulae

7.1 Forcasting a value from its history

Assume $y(t)$ is a value of a variable y at time t such as maximum price of a stock during a day. Then T-S formula for singleton consequeance is as follows[4].

R_i: If $y(t-1)$ is A_1^i and $y(t-2)$ is A_2^i and $\cdots$ and $y(t-n+1)$ is A_n^i then $y(t)$ is g^i.

7.2 Forcasting a value from its history

R_i: If $x_1(t)$ is A_1^i and $x_2(t)$ is A_2^i and $\cdots$ and $x_n(t)$ is A_n^i then $y(t)$ is g^i.

[4] Taken from Sheta, A. F. () Forecasting the Nile river flow using fuzzy logic model.

PART V

Fuzzy Relation

8 Relation between two sets

In this section we study fuzzy expressions such as "at least middle-aged," brighter than average," more or less expensive" and "younger than about 20."
First, let's recall Cartesian product $X \times Y$ in which both X and Y is a set. Let me take an example. Assume now $X = \{1, 2\}$ and $Y = \{a, b, c\}$ then $X \times Y = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}$. Then relation is defiened over Cartesian product $X \times Y$, that is, a subset of $X \times Y$. In other words relation is a set of ordered pair in which order is important.

Generally it is defined over multipel set, like $X_1 \times X_2 \times \cdots \times X_n$, but here we think of only product of two set, and call it *binary relation.*

To visualize we can plot $\mu_R(X, Y)$ 3-D Cartesian space.

⋆ Example 1 ... $X = \{1, 2\}$, $Y = \{2, 3, 4\}$, $R : X < Y$

Let's think of it as a *crisp* logic, that is, the value is 1 (yes) or 0 (no). Then membership function of this relation will be:

$X \setminus Y$	2	3	4
1	1	0	0
2	1	1	0
3	0	0	1

Then what about the relation $R : x \approx y$. Let's think of this example with *fuzzy* logic.

⋆ Example 2 ... $X = \{1, 2\}$, $Y = \{2, 3, 4\}$, $R : X \approx Y$

$X \setminus Y$	2	3	4
1	2/3	1/3	0
2	1	2/3	1/3
3	2/3	1	2/3

The values are just examples. Further more we think of X and Y as a continuous values instead of integer. Then membership function is a surface instead of just 9 points, over $X - Y$ coordinate.

We now proceed to examples where we use fuzzy linguistic expression instead of numbers.

First, these matrices are not necessarily rectangular. For example:

⋆ Example 3 ... X = {Brest, London, BuenosAires} Y=Tokyo, NewYork, Minsk, Johanesburg R: very far.

$X \setminus Y$	Tokio	New York	Minsk	Johanesburg
Brest				
London				
Buenos Aires				

Try to fill those blancs by yourself.
⋆ Example 4 ... X = {green, yellow, red}, Y = {unripe, semiripe, ripe}.

Imagine an apple. First, with a *crisp* logic. A red apple is usually ripe but a green apple is unripe. Thus:

$X \setminus Y$	unripe	semiripe	ripe
green	1	0	0
yellow	0	1	0
red	0	0	1

Now, secondly, with a *fuzzy* logic. A red apple is *provably* ripe, but a green apple is *most likely*, and so on. Thus, for example:

$X \setminus Y$	unripe	semiripe	ripe
green	1	0.5	0
yellow	0.3	1	0.4
red	0	0.2	1

8.1 Combine two fuzzy relations

We now return to the previous example of tomato.

$X \setminus Y$	unripe	semiripe	ripe
green	1	0.5	0
yellow	0.3	1	0.4
red	0	0.2	1

This is the relation of two sets:

$$X = \{green, yellow, red\}$$

and

$$Y = \{unripe, semiripe, ripe\}$$

Let's call this relation R_1. Then we think a similar but new Relation.

$$Y = \{unripe, semiripe, ripe\}$$

and

$$Z = \{sour, sour - sweet, sweet\}$$

Let's call this relation R_2.

$X \setminus Y$	sour	sour-sweet	sweet
unripen	0.8	0.5	0.1
semiripe	0.1	0.7	0.5
ripe	0.2	0.3	0.9

If we combine these two relations R_1 and R_2 by the formula

$$\mu_R(x, z) \geq \max_{y \in X}\{\min\{\mu_R(x, y), \mu_R(y, z)\}\},$$

the result is:
This relation could be expressed by our daily language like

> "If tomato is red then it's most likely sweet , possibly sour-sweet, and unlikely sour."

$X \setminus Y$	sour	sour-sweet	sweet
red	0.8	0.5	0.5
yellow	0.3	0.7	0.5
green	0.2	0.3	0.9

"If tomato is yellow then probably it's sour-sweet , possibly sour, maybe sweet."

"If tomato is green then almost always sour, less likely sour-sweet, unlikely sweet."

Or, we could say:

"Now tomato is more or less red, then what is taste like?"

9 Clustering by fuzzy relation

Algorithm 1 *1. Calculate a max-min similarity-relation $R = [a_{ij}]$*

2. Set $a_{ij} = 0$ for all $a_{ij} < \alpha$ and $i = j$

3. Select s and t so that $a_{ij} = \max\{a_{ij} | i < j \text{ and } i, j \in I\}$. When tie, select one of these pairs at random

WHILE $a_{st} \neq 0$ DO put s and t into the same cluster $C = \{s, t\}$ ELSE [4.]
ELSE all indices $\in I$ into separated clusters and STOP

4. Choose $u \in I \setminus C$ so that

$$\sum_{i \in C} a_{iu} = \max_{j \in I \setminus C} \{\sum_{i \in C} a_{ij} | a_{ij} \neq 0\}$$

When a tie, select one such u at random.

WHILE such a u exists, put u into $C = \{s, t, u\}$ and REPEAT [4.]

5. Let $I = I \setminus C$ and GOTO [3.]

Example

Exercise 3 *Starting from the following* 10×10 *proximity-relation* $R^(0)$*, apply the algorithm above. Assume now* $\alpha = 0.55$.

$$R^{(0)} = \begin{bmatrix} 1 & .7 & .5 & .8 & .6 & .6 & .5 & .9 & .4 & .5 \\ .7 & 1 & .3 & .6 & .7 & .9 & .4 & .8 & .6 & .6 \\ .5 & .3 & 1 & .5 & .5 & .4 & .1 & .4 & .7 & .6 \\ .8 & .6 & .5 & 1 & .7 & .5 & .5 & .7 & .5 & .6 \\ .6 & .7 & .5 & .7 & 1 & .6 & .4 & .5 & .8 & .9 \\ .6 & .9 & .4 & .5 & .6 & 1 & .3 & .7 & .7 & .5 \\ .5 & .4 & .1 & .5 & .4 & .3 & 1 & .6 & .2 & .3 \\ .9 & .8 & .4 & .7 & .5 & .7 & .6 & 1 & .4 & .4 \\ .4 & .6 & .7 & .5 & .8 & .7 & .2 & .4 & 1 & .7 \\ .5 & .6 & .6 & .6 & .9 & .5 & .3 & .4 & .7 & 1 \end{bmatrix}$$

An example solution

By repeating $R^{(n+1)} = R^{(n)} \circ R^{(n)}$ till $R^{(n)} = R^{(n+1)}$. In this way, similarity-relation $R^{(n)}$ will be calculated as:

$$R^{(n)} = \begin{bmatrix} 1 & .2 & .5 & .8 & .6 & .2 & .3 & .9 & .4 & .3 \\ .2 & 1 & .3 & .6 & .7 & .9 & .2 & .8 & .3 & .2 \\ .5 & .3 & 1 & .5 & .3 & .4 & .1 & .3 & .7 & .6 \\ .8 & .6 & .5 & 1 & .7 & .3 & .5 & .4 & .1 & .3 \\ .6 & .7 & .3 & .7 & 1 & .2 & .4 & .5 & .8 & .9 \\ .2 & .9 & .4 & .3 & .2 & .4 & .1 & .3 & .7 & .2 \\ .3 & .2 & .1 & .5 & .4 & .1 & 1 & .6 & .1 & .3 \\ .9 & .8 & .3 & .4 & .5 & .3 & .6 & 1 & 0 & .2 \\ .4 & .3 & .7 & .1 & .8 & .7 & .1 & 0 & 1 & .1 \\ .3 & .2 & .6 & .3 & .9 & .2 & .3 & .2 & .1 & 1 \end{bmatrix}$$

Now apply [1.] and [2.]

$$\begin{bmatrix} 0 & .7 & 0 & .8 & .6 & .6 & 0 & .9 & 0 & 0 \\ .7 & 0 & 0 & .6 & .7 & .9 & 0 & .8 & .6 & .6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 & .6 \\ .8 & .6 & 0 & 0 & .7 & 0 & 0 & .7 & 0 & .6 \\ .6 & .7 & 0 & .7 & 0 & .6 & 0 & 0 & .8 & .9 \\ .6 & .9 & 0 & 0 & .6 & 0 & 0 & .7 & .7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .6 & 0 & 0 \\ .9 & .8 & 0 & .7 & 0 & .7 & .6 & 0 & 0 & 0 \\ 0 & .6 & .7 & 0 & .8 & .7 & 0 & 0 & 0 & .7 \\ 0 & .6 & .6 & .6 & .9 & 0 & 0 & 0 & .7 & 0 \end{bmatrix}$$

First, set $I = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and $C = \{\ \}$.

Then

3. Now $a_{18} = a_{26} = a_{5\ 10} = 0.9$ are maximum and a_{18} is randomly selected. Then $C = \{1, 8\}$.

4. $a_{12} + a_{82} = a_{14} + a_{84} = 1.5$ are maximum and $j = 4$ is randomly selected. Then $C = \{1, 8, 4\}$.

4. $a_{12} + a_{42} + a_{82} = 2.1$ is maximum, then $C = \{1, 8, 4, 2\}$.

4. There are no j such that $a_{1j} + a_{2j} + a_{4j} + a_{8j}$ is maximum. Then final $C = \{1, 8, 4, 2\}$.

   ⋆ $a_{16} + a_{26} + a_{46} + a_{86} = 0.6 + 0.9 + 0 + 0.7 = 2.2$ seems maximum but actually not because $a_{46} = 0$

   Note that $\sum_{i \in C} a_{iu} = \max_{j \in I \setminus C} \{\sum_{i \in C} a_{ij} | a_{ij} \neq 0\}$

5. Let $I = \{3, 5, 6, 7, 9, 10\}$

3. $a_{5\ 10} = 0.9$ is maximum. Then renew C as $\{5, 10\}$.

4. $a_{59} + a_{10\ 9} = 1.5$ is maximum. Then $C = \{5, 10, 9\}$.

4. There are no j in $\{3, 6, 9\}$ such that $a_{5j} + a_{9j} + a_{10j}$ is maximum. Then final $C = \{5, 10, 9\}$.

5. Let $I = \{3, 6, 7\}$.

3. Now $a_{36} = a_{37} = a_{67} = 0$. Then $\{3\}$, $\{6\}$and $\{7\}$are three separated clusters. In fact,

$$\begin{bmatrix} a_{33} & a_{36} & a_{37} \\ a_{63} & a_{66} & a_{67} \\ a_{73} & a_{76} & a_{77} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

   So $\sum_{i \in \{3,6,7\}} a_{iu} = \max_{j \in \{3,6,7\}} \{\sum_{i \in C} a_{ij} | a_{ij} \neq 0\}$ does not exit any more.

In this way, when $\alpha = 0.55$, we have 5 clasters $\{1, 8, 4, 2\}$,$\{5, 10, 9\}$, $\{3\}$, $\{6\}$ and $\{7\}$ are obtained.