School of Information Technology
Indian Institute of Technology Kharagpur
IT60108: Soft Computing and Application
Class Test - I

F.M. 20 Session 2014 – 2015 Time: 20 mins

Q. 1

(a) μ_A(x) and μ_B(x) are the membership functions of the fuzzy sets A and B, respectively.

$\mu_A(x) = e^{\frac{1}{1+x}}$

$\mu_B(x) = \frac{1}{1+(\frac{x-50}{10})^4}$

Decide whether A and B are closed or open.

Answer : A membership function $\mu(x)$ is said to be closed iff $\lim_{x\to-\infty} \mu_(x)$ =0= $lim_{x\to\alpha}\mu(x)$

Case 1 : $\mu_A(x) = e^{\frac{1}{1+n}}$: Here $\lim_{x\to-\infty} \mu_A(x) = \lim_{x\to+\infty}\mu_A(x) = 1$
Hence, it is neither closed nor open.

Case 2 : $\mu_B(x) = \frac{1}{1+(\frac{x-50}{10})^4}$: Here $\lim_{x\to-\infty} \mu_B(x) = \lim_{x\to+\infty} \mu_A(x) = 0$
Hence, it is closed.

[2+2]

(b) Given two fuzzy sets A and B defined over universe of discourses X and Y, respectively.
$A = \{(20, 0.2), (25, 0.4), (30, 0.6), (35, 0.6), (40, 0.7), (45, 0.8),$
$(50, 0.8)\}$
$B = \{(1, 0.8), (2, 0.8), (3, 0.6), (4, 0.4)\}$
$X = \{10, 15, 20, 25, 30, 35, 40, 45, 50, 55\}$
$Y = \{0, 1, 2, 3, 4, 5\}$

Draw the graphs for the following.

i. $A \times B$
ii. $A \Longrightarrow B$

(i) $A \times B$

$$\mu_{A\times B}(x, y) = min\{\mu_A(x), \mu_B(y)\} = \begin{array}{c} \\ 20 \\ 25 \\ 30 \\ 35 \\ 40 \\ 45 \\ 50 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 \\ 0.4 & 0.4 & 0.4 & 0.4 \\ 0.6 & 0.6 & 0.6 & 0.4 \\ 0.6 & 0.6 & 0.6 & 0.4 \\ 0.7 & 0.7 & 0.6 & 0.4 \\ 0.8 & 0.8 & 0.6 & 0.4 \\ 0.8 & 0.8 & 0.6 & 0.4 \end{bmatrix} \end{array}$$

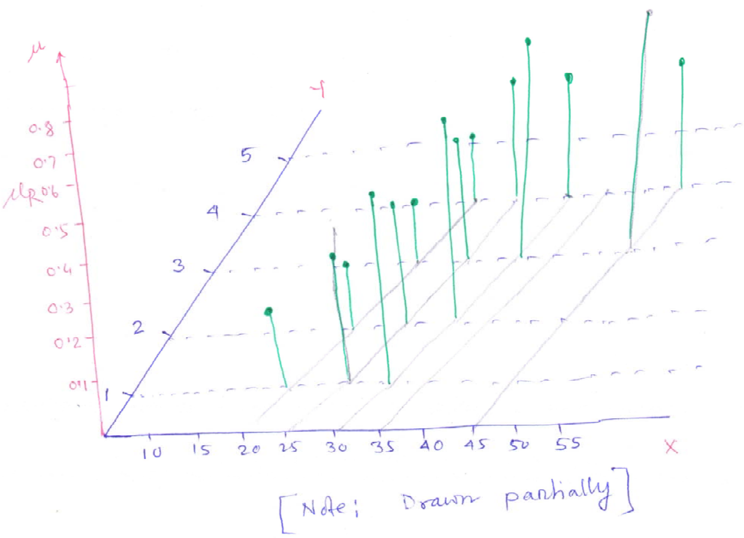
(ii) $A \Rightarrow B$

For this many interpretation are possible.

$A \Rightarrow B \equiv \bar{A} \cup B$ or $A \times B$ or $(A \times B) \cup (\bar{A} \times Y)$

Accordingly answer will be different.

[4+4]



Q. 2

(a) Suppose, a fuzzy relation is 'If x is A then y is B'. How to find the following:

i. x is C, given that y is D
ii. y is D, given that x is C

Answer :
GPM

if x is A then y is B
x is A'
— — — — — — — — — — — — — — — — — —
y is B'

GMT is

if x is A then y is B
y is B'
— — — — — — — — — — — — — — — — — —
x is A'

Since C and D are not mentioned as A' or B' and (vice-versa) none of the GMP and GMT are applicable in this case and hence we can not deduce anything.

[3+3]

(b) Two fuzzy sets P and Q are defined on $x \in$ X as follows.

	x_1	x_2	x_3	x_4	x_5
P	0.1	0.2	0.7	0.5	0.4
Q	0.9	0.6	0.3	0.2	0.8

Find (i.) $(P \cap \overline{Q})_{0.4}$ (ii.) $(P \times Q)_{0.4}$

Answer :
(i)
$P \cap \overline{Q} = (x_1, 0.1), (x_2, 0.2), (x_3, 0.7), (x_4, 0.5), (x_5, 0.2)$

$\therefore (P \cap \overline{Q})_{0.4} = \{x | \mu(x) \geq 0.4\} = \{x_3, x_4\}$

(ii)

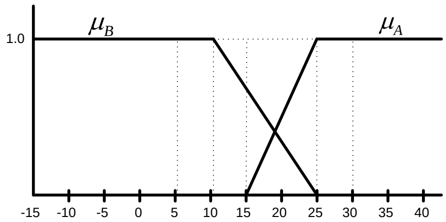
$$P \times Q = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \left[\begin{array}{ccccc} 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.7 & 0.6 & 0.3 & 0.2 & 0.7 \\ 0.5 & 0.5 & 0.3 & 0.2 & 0.4 \\ 0.4 & 0.4 & 0.3 & 0.2 & 0.4 \end{array} \right] \end{array}, \quad (P \times Q)_{0.4} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array} \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{array} \right] \end{array}$$

[3+3]

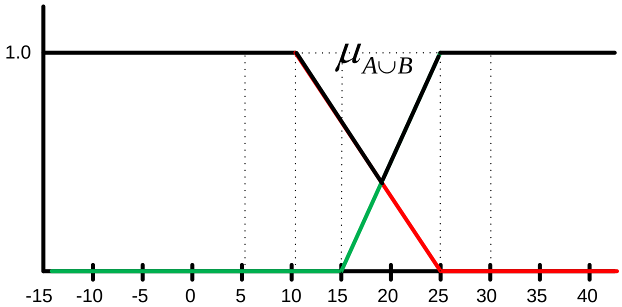
Q. 3

(a) The membership functions of two fuzzy sets A and B are shown in the following graph.
A: climate is Hot.

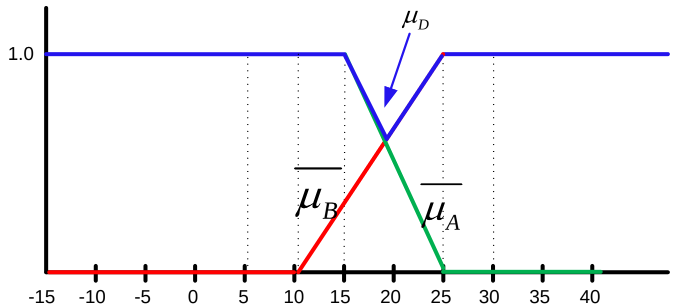
B: climate is Cold.



i. Draw the graph of the membership function, which represents the fuzzy set C: *climate is Extreme.*
Answer : Climate is Extreme $\approx$ Climate is Hot OR Climate is Cold.



ii. What would be the graph of the membership function μ_D of the fuzzy set $D = \overline{(A \cap C)}$? State D in terms of fuzzy linguistic.
Answer :



Lingustic interpretation : $A \cap B$ = Climate is Pleasant, $D = A \bar{\cap} B$ = Climate is not Pleasant

[3+3]

(b) Two fuzzy relations 'likes' and 'earns' are defined below.

$$\text{likes} = \begin{array}{c} \\ Dhoni \\ Virat \\ Rohit \\ Sekhar \end{array} \begin{array}{c} \begin{array}{ccc} Football & Hockey & Cricket \end{array} \\ \begin{bmatrix} 0.1 & 0.3 & 0.8 \\ 0.2 & 0.7 & 0.5 \\ 0.5 & 0.4 & 0.2 \\ 0.4 & 0.5 & 0.6 \end{bmatrix} \end{array}$$

For example, x likes Game.

$$\text{earns} = \begin{array}{c} \\ Dhoni \\ Virat \\ Rohit \\ Sekhar \end{array} \begin{array}{c} \begin{array}{ccc} 10L & 50L & 100L \end{array} \\ \begin{bmatrix} 0.6 & 0.3 & 0.2 \\ 0.4 & 0.7 & 0.8 \\ 0.1 & 0.3 & 0.2 \\ 0.5 & 0.2 & 0.6 \end{bmatrix} \end{array}$$

For example, x earns Money.

Obtain the relation between a game to a money? [**6**]

Answer :

This relation can be obtained as $likes^T \circ earns$. That is,

$$\begin{array}{c} \\ Footboll \\ Hockey \\ Cricket \end{array}\begin{array}{c} \begin{array}{cccc} Dhoni & Virat & Rohit & Sekhan \end{array} \\ \left[\begin{array}{cccc} 0.1 & 0.2 & 0.5 & 0.4 \\ 0.3 & 0.7 & 0.4 & 0.5 \\ 0.8 & 0.5 & 0.2 & 0.6 \end{array}\right] \end{array} \circ \begin{array}{c} \\ Dhoni \\ Virat \\ Rohit \\ Sekhan \end{array}\begin{array}{c} \begin{array}{ccc} 10L & 50L & 100L \end{array} \\ \left[\begin{array}{ccc} 0.6 & 0.3 & 0.2 \\ 0.4 & 0.7 & 0.8 \\ 0.1 & 0.3 & 0.2 \\ 0.5 & 0.2 & 0.6 \end{array}\right] \end{array} =$$

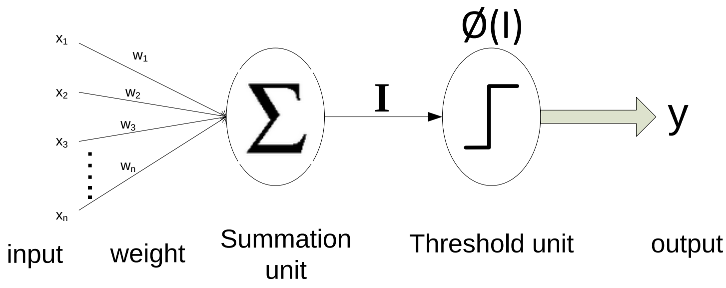
$$\begin{array}{c} \\ Football \\ Hockey \\ Circket \end{array}\begin{array}{c} \begin{array}{ccc} 10L & 50L & 100L \end{array} \\ \left[\begin{array}{ccc} 0.4 & 0.3 & 0.4 \\ 0.5 & 0.7 & 0.7 \\ 0.6 & 0.5 & 0.6 \end{array}\right] \end{array}$$

Q. 4

(a) What are the components you should consider in order to mathematically model an artificial neuron ?

[4]

Answer : Two components namely summation until and threshold unit are required to mathematically model as artificial neuron. Two components can be defined as follows.



Summation unit $I = \sum_{i=1}^{n} x_i \cdot w_i$. Threshold unit $\bigcirc = \phi(I)$, where ϕ is some transfer function.

(b) If $\phi(I) = \frac{1}{1+e^{-\alpha I}}$ is a transfer function in a perceptron, then show that

Answer :

$\phi(I) = \frac{1}{1+e^{-\alpha I}}$; Let $z = 1 + e^{-\alpha I}$. $\therefore \frac{\partial z}{\partial I} = -\alpha \cdot e^{-\alpha I} = -\alpha \cdot \frac{1-\phi(I)}{\phi(I)}$

$\frac{\partial \phi(I)}{\partial I} = \frac{\partial \phi(I)}{\partial z} \cdot \frac{\partial z}{\partial I}$

$= -\frac{1}{z^2} \cdot -\alpha \cdot \frac{1-\phi(I)}{\phi(I)}$

$= \phi(I)^2 \cdot \alpha \cdot \frac{1-\phi(I)}{\phi(I)}$, $\because \frac{1}{z} = \phi(I)$

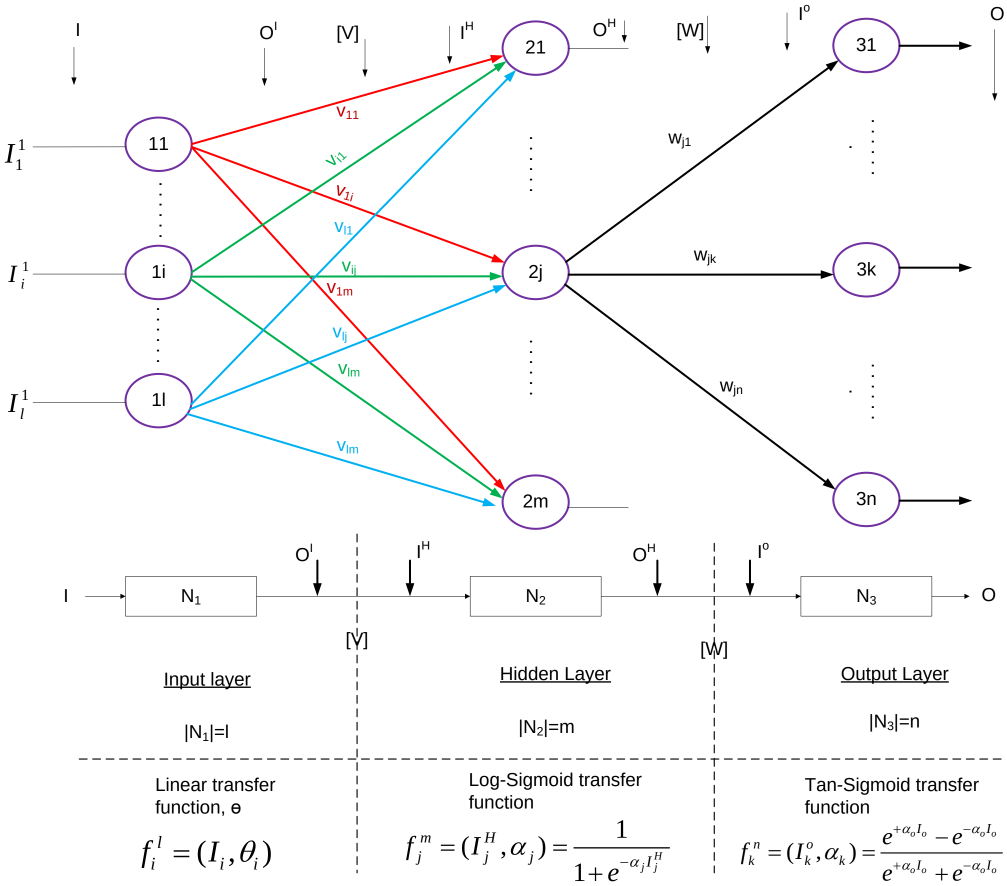
$= \alpha(1 - \phi(I)) \cdot \phi(I)$ **[3]**

(c) Draw a schematic diagram of a multi-layer feed-forward artificial neural network architecture and clearly label the different elements in it.

Give one application, where you should apply such an ANN architecture.

Answer :

Application : Such an ANN would be applied to problems whose output are non-separable with resepct to input. **[4+1]**



Q. 5

(a) Show how the computations in input, hidden and output layers of an ANN can be accomplished in terms of matrix algebra.

[2+2+2]

Answer :

Whole learning method consists of the following three computations:

(a) **Input layer computation**

(b) Hidden layer computation

(c) Output layer computation

In our computation, we assume that $< T_0, T_I >$ be the training set of size $|T|$.

- Let us consider an input training data at any instant be $I^I = [I_1^1, I_2^1, \cdots, I_i^1, I_l^1]$ where $I^I \in T_I$
- Consider the outputs of the neurons lying on input layer are the same with the corresponding inputs to neurons in hidden layer. That is,

  $O^I = I^I$

  $[l \times 1] = [l \times 1]$ [Output of the input layer]
- The input of the j-th neuron in the hidden layer can be calculated as follows.

$$I_j^H = v_{1j}o_1^I + v_{2j}o_2^I +, \cdots, +v_{ij}o_j^I + \cdots + v_{ij}o_l^I$$

where $j = 1, 2, \cdots m$.

[Calculation of input of each node in the hidden layer]

- In the matrix representation form, we can write

$$I^H = V^T \cdot O^I$$
$$[m \times 1] = [m \times l]\ [l \times 1]$$

- Let us consider any j-th neuron in the hidden layer.
- Since the output of the input layer's neurons are the input to the j-th neuron and the j-th neurons follows the log-sigmoid transfer function, we have

$$O_j^H = \frac{1}{1+e^{-\alpha_H \cdot I_j^H}}$$

where $j = 1, 2, \cdots, m$ and α_H is the constant co-efficient of the transfer function.

Note that all output of the nodes in the hidden layer can be expressed as a one-dimensional column matrix.

$$O^H = \begin{bmatrix} \cdots \\ \cdots \\ \vdots \\ \frac{1}{1+e^{-\alpha_H \cdot I_j^H}} \\ \vdots \\ \cdots \\ \cdots \end{bmatrix}_{m \times 1}$$

Let us calculate the input to any k-th node in the output layer. Since, output of all nodes in the hidden layer go to the k-th layer with weights $w_{1k}, w_{2k}, \cdots, w_{mk}$, we have

$$I_k^O = w_{1k} \cdot o_1^H + w_{2k} \cdot o_2^H + \cdots + w_{mk} \cdot o_m^H$$

where $k = 1, 2, \cdots, n$

In the matrix representation, we have

$$I^O = W^T \cdot O^H$$
$$[n \times 1] = [n \times m]\ [m \times 1]$$

Now, we estimate the output of the k-th neuron in the output layer. We consider the tan-sigmoid transfer function.

$$O_k = \frac{e^{\alpha_o \cdot I_k^o} - e^{-\alpha_o \cdot I_k^o}}{e^{\alpha_o \cdot I_k^o} + e^{-\alpha_o \cdot I_k^o}}$$

for $k = 1, 2, \cdots, n$

Hence, the output of output layer's neurons can be represented as

$$O = \begin{bmatrix} \cdots \\ \cdots \\ \vdots \\ \frac{e^{\alpha_o \cdot I_k^o} - e^{-\alpha_o \cdot I_k^o}}{e^{\alpha_o \cdot I_k^o} + e^{-\alpha_o \cdot I_k^o}} \\ \vdots \\ \cdots \\ \cdots \end{bmatrix}_{n \times 1}$$

(b) Explain the basic principle of calculating error in supervised learning.

[2]

Answer :

- Let us consider any k-th neuron at the output layer. For an input pattern $I_i \in T_I$ (input in training) the target output T_{Ok} of the k-th neuron be T_{Ok}.
- Then, the error e_k of the k-th neuron is defined corresponding to the input I_i as

$$e_k = \frac{1}{2}\left(T_{Ok} - O_{Ok}\right)^2$$

where O_{Ok} denotes the observed output of the k-th neuron.

- For a training session with $I_i \in T_I$, the error in prediction considering all output neurons can be given as

$$e = \sum_{k=1}^{n} e_k = \frac{1}{2}\sum_{k=1}^{n}\left(T_{Ok} - O_{Ok}\right)$$

where n denotes the number of neurons at the output layer.

- The total error in prediction for all output neurons can be determined considering all training session $< T_I, T_O >$ as

$$E = \sum_{\forall I_i \in T_I} e = \frac{1}{2}\sum_{\forall t \in <T_I, T_O>} \sum_{k=1}^{n}\left(T_{Ok} - O_{Ok}\right)^2$$

(c) Derive the 'delta rule' according to the method of *Steepest descent*.

- For simplicity, let us consider the connecting weights are the only design parameter.
- Suppose, V and W are the wights parameters to hidden and output layers, respectively.
- Thus, given a training set of size N, the error surface, E can be represented as

$$E = \sum_{i=1}^{N} e^i\left(V, W, I_i\right)$$

where I_i is the i-th input pattern in the training set and $e^i(...)$ denotes the error computation of the i-th input.

- Now, we will discuss the steepest descent method of computing error, given a changes in V and W matrices.
- Suppose, A and B are two points on the error surface (see figure in Slide 30). The vector $\vec{AB}$ can be written as

$$\vec{AB} = (V_{i+1} - V_i)\cdot \bar{x} + (W_{i+1} - W_i)\cdot \bar{y} = \Delta V \cdot \bar{x} + \Delta W \cdot \bar{y}$$

The gradient of $\vec{AB}$ can be obtained as

$$e_{\vec{AB}} = \frac{\partial E}{\partial V}\cdot \bar{x} + \frac{\partial E}{\partial W}\cdot \bar{y}$$

Hence, the unit vector in the direction of gradient is

$$\bar{e}_{\vec{AB}} = \frac{1}{|e_{\vec{AB}}|}\left[\frac{\partial E}{\partial V}\cdot \bar{x} + \frac{\partial E}{\partial W}\cdot \bar{y}\right]$$

- With this, we can alternatively represent the distance vector AB as

$$\vec{AB} = \eta\left[\frac{\partial E}{\partial V}\cdot \bar{x} + \frac{\partial E}{\partial W}\cdot \bar{y}\right]$$

where $\eta = \frac{k}{|e_{\vec{AB}}|}$ and k is a constant

- So, comparing both, we have

$$\Delta V = \eta \frac{\partial E}{\partial V}$$
$$\Delta W = \eta \frac{\partial E}{\partial W}$$

This is also called as **delta rule** and η is called **learning rate**.

[2+2]