II. FUZZY SET METHODS – CLUSTER ANALYSIS

OBJECTIVES

To study fuzzy cluster analysis and how to solve basic problems using fuzzy cluster analysis

Fuzzy Clustering - Theory

CLUSTER ANALYSIS

- way to search for structure in a dataset X
- a component of pater recognition
- clusters form a partition

Examples:

- partition all credit card users into two groups, those that are legally using their credit cards and those who are illegally using stolen credit cards
- partition UCD students into two classes, those who will go sking over winter vacation and those who will go to the beach

Fuzzy Clustering - Theory

REMARKS: (1) The dataset, in the case of students would include such things as **agschool, income of parents, number of years as student, marital status**

(2) Classical cluster analysis would partition the set of student (with respect to their characteristics; that is, the items in the dataset) into disjoint sets P_i so that we would have:

$$\bigcup_{i=1}^{c} P_i \text{ and } P_i \cap P_j = \{\Phi\} \text{ for } i \neq j.$$

Fuzzy Clustering - Theory

Let's suppose that our dataset has:

Age = {17,18,...,35}

School = {Arts, Drama, ..., Civil Engineering, Natural Science, Mathematics, Computer Science}

Income = {$0 ... $500,000}

Note: It is (or should be) intuitively clear that for this problem the partitions are intersecting since for many students there is an equal preference between going to the beach and going to ski for vacation and the preferences are not zero/one for most students.

Fuzzy Clustering - Theory

The idea of cluster analysis is to obtain centers (i=1,…,c where c=2 for the example of sking and going to the beach) $v_1,\ldots,v_c$ that are exemplars and radii that will define the partition. Now, the centers serve as exemplars and an advertising company could sent sking brochures to the group that is defined by the first center and another brochure for beach trips for students. The idea of **fuzzy clustering** (**fuzzy c-means** clustering where c is an *a-priori* chosen number of clusters) is to allow overlapping clusters with partial membership of individuals in clusters.

Fuzzy Clustering - Theory

Given: dataset $X = \{x_1, \ldots, x_N\}$

fuzzy partition $P = \{A_1, \ldots, A_c\}$

such that

$\sum_{i=1}^{c} \mu_{A_i}(x_j) = 1$ - says that each data point must be completely distributed

$0 < \sum_{j=1}^{N} \mu_{A_i}(x_j) < 1$ - there are no empty partitions, a partition must contain some element(s) to some degree

Fuzzy Clustering – Example (from Klir&Yuan)

$A_1 = \{0.6/x_1, 1/x_2, 0.1/x_3\}$

$A_2 = \{0.4/x_1, 0/x_2, 0.9/x_3\}$

Property 1:

$$\mu_{A_1}(x_1) + \mu_{A_2}(x_1) = 1.0$$

$$\mu_{A_1}(x_2) + \mu_{A_2}(x_2) = 1.0$$

$$\mu_{A_1}(x_3) + \mu_{A_2}(x_3) = 1.0$$

Property 2:

$$0 < \mu_{A_1}(x_1) + \mu_{A_1}(x_2) + \mu_{A_1}(x_3) = 1.7 < 3$$

$$0 < \mu_{A_2}(x_1) + \mu_{A_2}(x_2) + \mu_{A_2}(x_3) = 1.3 < 3$$

Fuzzy Clustering

In general:

$$\vec{x}_j = \begin{pmatrix} x_{j_1} \\ \vdots \\ x_{j_p} \end{pmatrix} \in \Re^p \text{ a p-dimensional vector. For our example,}$$

$$\text{particular student \#17, } x_{17} = \begin{pmatrix} 20 \\ \textit{mathematics} \\ \$60{,}000 \end{pmatrix}$$

Fuzzy Clustering

Suppose all components to the vectors in the dataset are numeric, then:

$$v_i = \frac{\sum_{j=1}^{N} [\mu_{A_i}(x_j)]^m x_j}{\sum_{j=1}^{N} [\mu_{A_i}(x_j)]^m} \qquad (4.1)$$

m>1 governs the effect of the membership grade.

Fuzzy Clustering

Given a way to compute the center v_i we need a way to measure how good these centers are (one by one). This is done by a **performance measure** or **objective function** as follows:

$$J_m(P) = \sum_{j=1}^{N} \sum_{i=1}^{c} [\mu_{A_i}(x_j)]^m \left\| x_j - v_i \right\|^2 \qquad (4.2)$$

Fuzzy Clustering: Fuzzy c-means algorithm

Step 1: Set k=0, select an initial partition $P^{(0)}$

Step 2: Calculate centers $v_i^{(k)}$ according to equation (4.1)

Step 3: Update the partition to $P^{(k+1)}$ according to:

Fuzzy Clustering: Fuzzy c-means algorithm (step 3 continued)

$$\forall x_j \in X \text{(dataset)}, \text{ if } \left\| x_j - v_i^{(k)} \right\|^2 > 0, \text{ then}$$

$$\mu_{A_i}^{(k+1)}(x_j) = \left[\sum_{l}^{c} \left(\frac{\left\| x_j - v_i^{(k)} \right\|^2}{\left\| x_j - v_l^{(k)} \right\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}.$$

$$\text{If } \left\| x_j - v_i^{(k)} \right\|^2 = 0 \text{ for some } i \in I \subseteq \{1,\ldots,c\}, \text{ then assign (arbitrarily) values so that}$$

$$\sum_{i \in I} \mu_{A_i}^{(k+1)}(x_j) = 1 \text{ and}$$

$$\mu_{A_i}^{(k+1)}(x_j) = 0 \text{ for } i \notin I.$$

Fuzzy Clustering: Fuzzy c-means algorithm

Step 4: Compare $P^{(k)}$ to $P^{(k+1)}$. If $|| P^{(k)} - P^{(k+1)} || < \varepsilon$ then stop. Otherwise set k:=k+1 and go to step 2.

Remark: the computation of the updated membership function is the condition for the minimization of the objective function given by equation (4.2).

The example that follows uses c=2, $\varepsilon=0.01$, the Euclidean norm and $A_1 = \{0.854/x_1 ,\ldots, 0.854/x_{15}\}$ and

$$A_2 = \{0.146/x_1 ,\ldots, 0.146/x_{15}\}.$$

For k=6, A_1 and A_2 are given in the following slide where $v_1^{(6)}=(0.88,2)^T$ and $v_2^{(6)}=(5.14,2)^T$

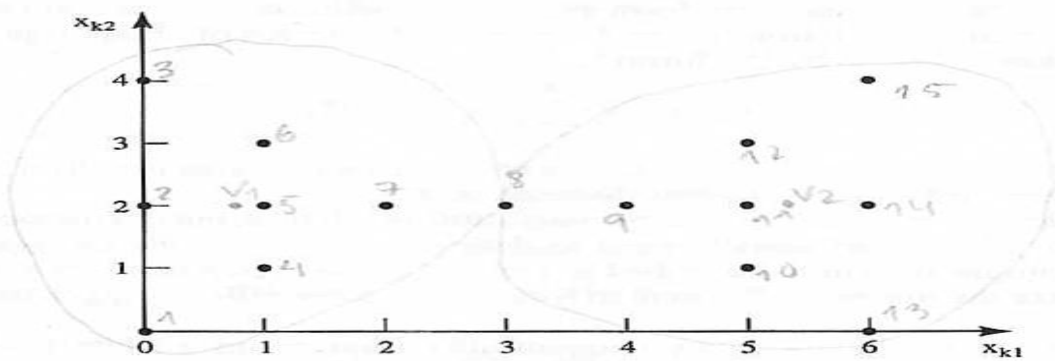
Fuzzy Clustering - Example

To illustrate the fuzzy c-means algorithm, let us consider a data set X that consists of the following 15 points in $\mathbb{R}^2$:

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_{k1}	0	0	0	1	1	1	2	3	4	5	5	5	6	6	6
x_{k2}	0	2	4	1	2	3	2	2	2	1	2	3	0	2	4

The data are also shown in Fig. 13.1a. Assume that we want to determine a fuzzy pseudopartition with two clusters (i.e., $c = 2$). Assume further that we choose $m = 1.25$, $\varepsilon = 0.01$; $\|\cdot\|$ is the Euclidean distance, and the initial fuzzy pseudopartition is $\mathcal{P}^{(0)} = \{A_1, A_2\}$ with

$$A_1 = .854/x_1 + .854/x_2 + \ldots + .854/x_{15},$$
$$A_2 = .146/x_1 + .146/x_2 + \ldots + .146/x_{15}.$$



(a)

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$A_1(\mathbf{x}_k)$	.99	1	.99	1	1	1	.99	.47	.01	0	0	0	.01	0	.01
$A_2(\mathbf{x}_k)$	.01	0	.01	0	0	0	.01	.53	.99	1	1	1	.99	1	.99

(b)

(a) data; (b) fuzzy pseudopartition $\mathcal{P}^{(6)} = \{A_1, A_2\}$.

[Bezdek, 1981]

Neural Networks

A neural network works as follows:

INPUT ----→ PROCESS ----→ OUTPUT

^

neural network

Thus, a neural network is a (mathematical) function.

We obtain values by **training** the neural network

Neural Networks

ISSUE:

Structure of the neural network

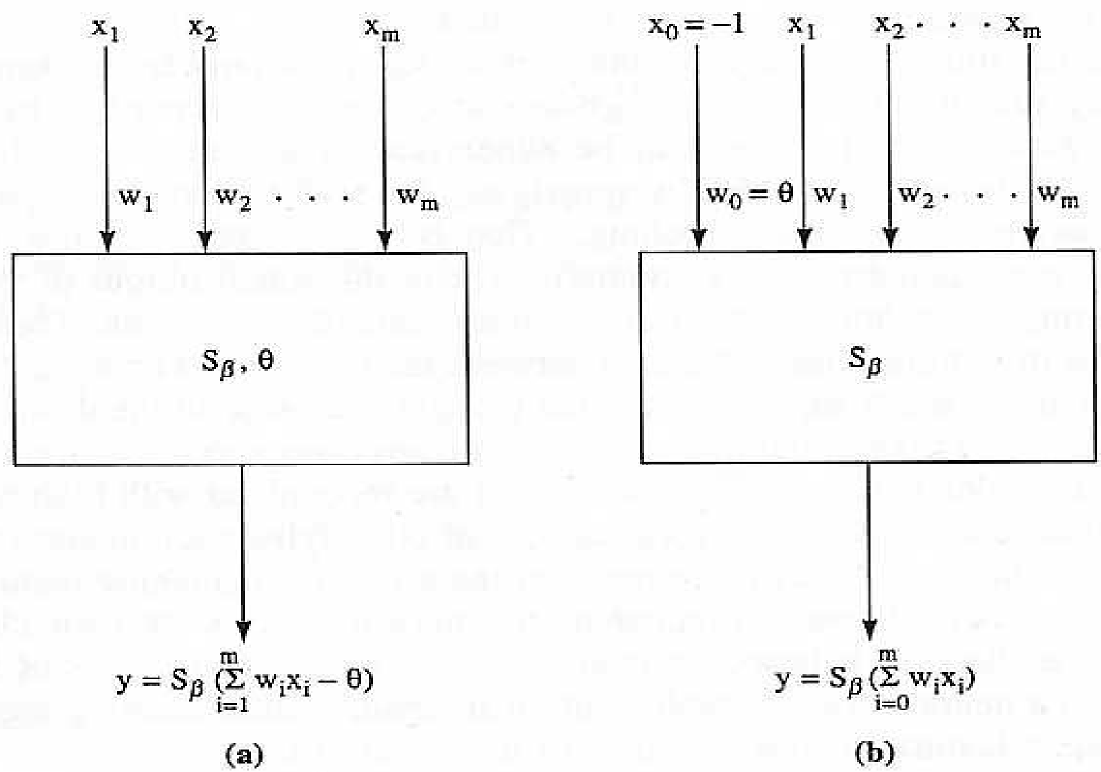
- how to connect nodes (output bins)
- weights on the connections
- number of layers
- activation function
- bias

Training

- how to train when you don't know what output you want, this is usually *unsupervised*
- how to train when you know the output you want, this is *supervised*

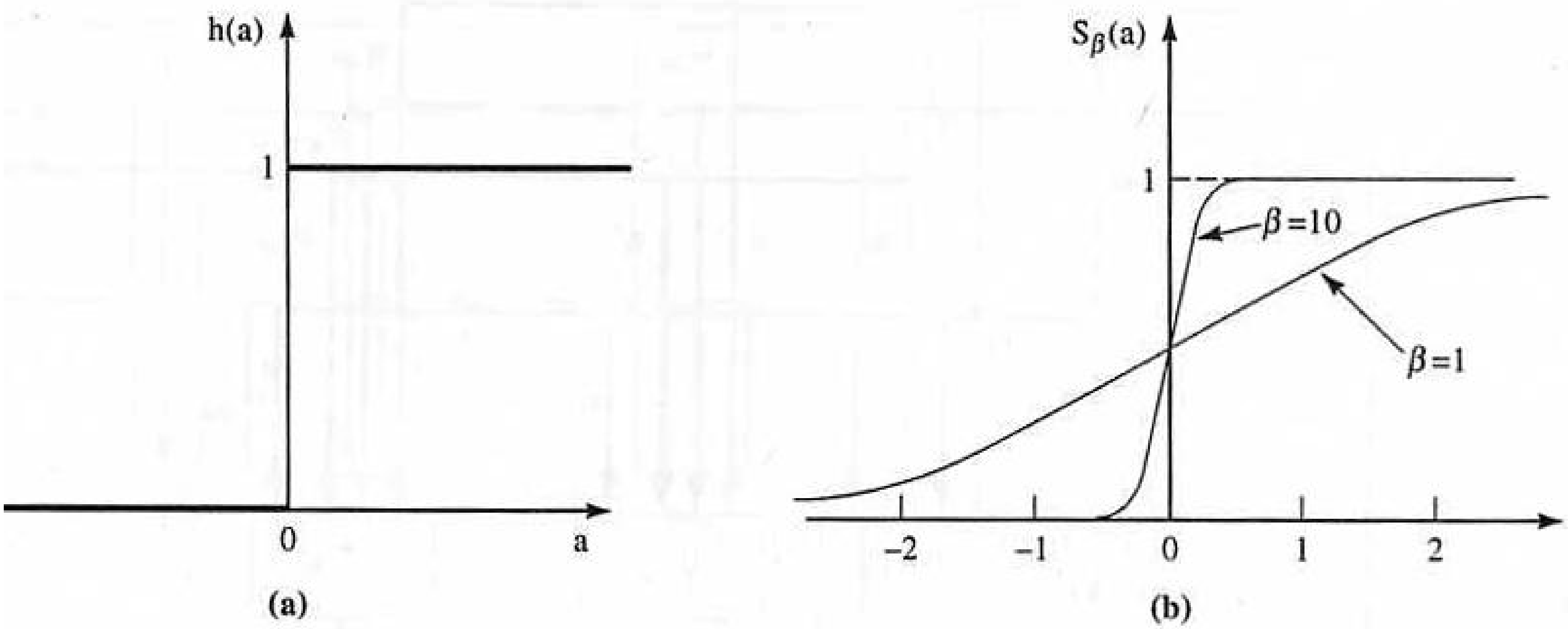
Neural Networks – Crisp Neural Networks

$$y = s_\beta\left(\sum_{i=1}^{n} w_i x_i - \theta\right)$$



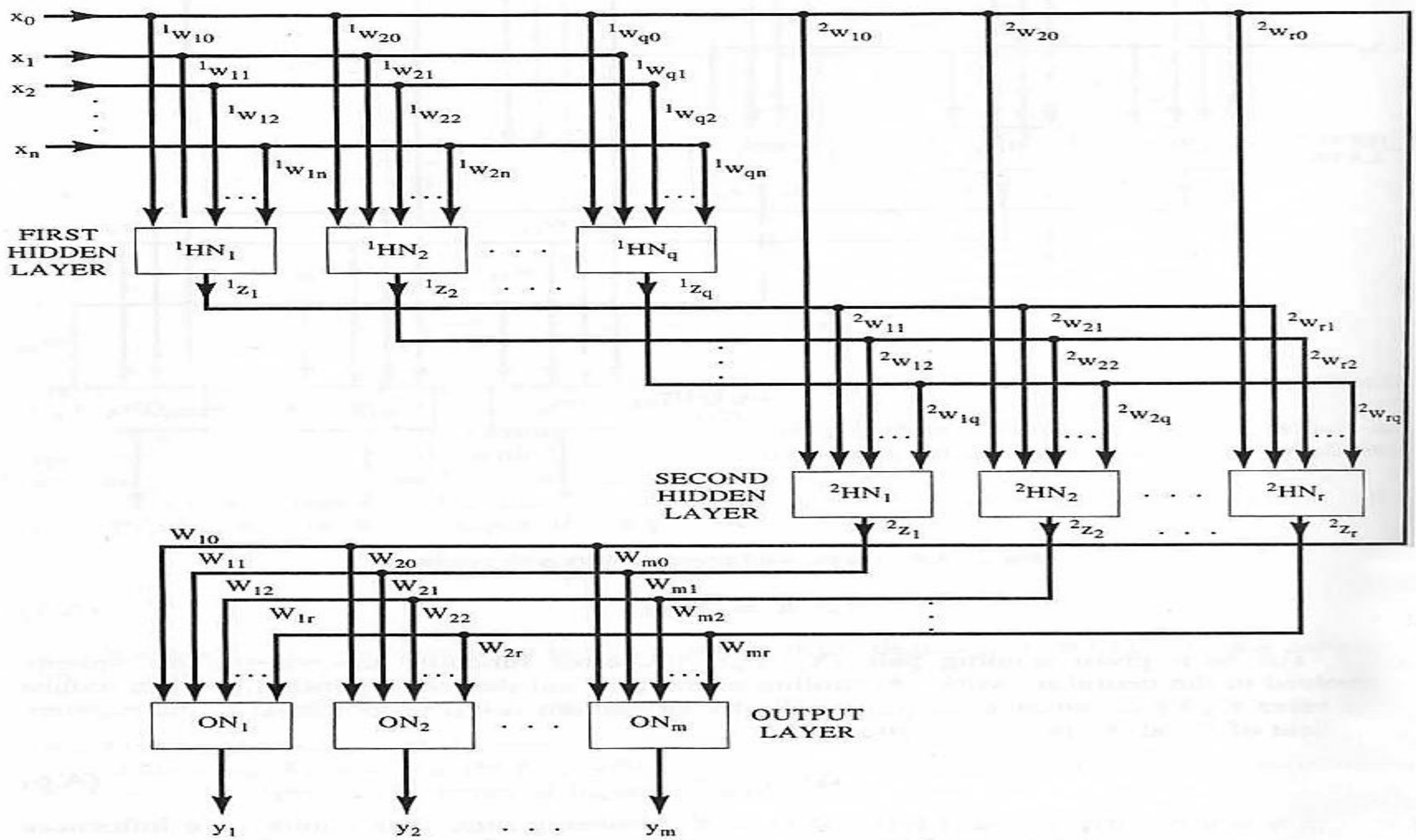
Two equivalent representation of a neuron activated by a sigmoid function s_β with bias θ.

Crisp Neural Networks



Examples of activation functions: (a) Heaviside function h; (b) sigmoid functions s_β.

Crisp Neural Networks



Feedforward neural network with three layers (two hidden layers and one output layer).

Neural Networks and Fuzzy Set - Applications

In determining which fuzzy rules are useful in a fuzzy logic systems. For example, suppose we are trying to control an electrical power system and we have demands for power by hour, by day of the year and each demand has five states (very low, low, medium, high, very high) and output levels of the generator are say ten each of which has five states. The on the input side, one has 24x365x5=43,800 different possibilities and on the output side, one has 10x5=50. One would have 43,500x50 possible rules = 2,190,000. Say that one concentrated rules for a single day then one is looking at 24x5x10x5 or 6,000 rules which is still many rules. Neural networks can be used to determine which rules are not needed.