. E.

Seventh Semester Examination, Dec-2007

NEURAL NETWORKS

Note : Attempt any five questions.

Q. 1. With the help of suitable diagram explain the structure of biological neurons? Also show how information flows in the neutral system?

Ans. It is claimed that the human central nervous system is comprised of about 1, 3 × 1010 neurons and that about 1 × 1010 of them takes place in the brain. At any time, some of these neurons are firing and the power dissipation due this electrical activity is estimated to be in the order of 10 watts. Monitoring the activity in the brain has shown that, even when asleep, 5 × 107 nerve impulses per second are being relayed back and forth between the brain and other parts of the body. This rate is increased significantly when awake A neuron has a roughly spherical cell body called soma (Figure 1.). The signals generated in soma are transmitted to other neurons through an extension on the cell body called axon or nerve fibres. Another kind of extensions around the cell body like bushy tree is the dendrites, which the responsible from receiving the incoming signals generated by other neurons.

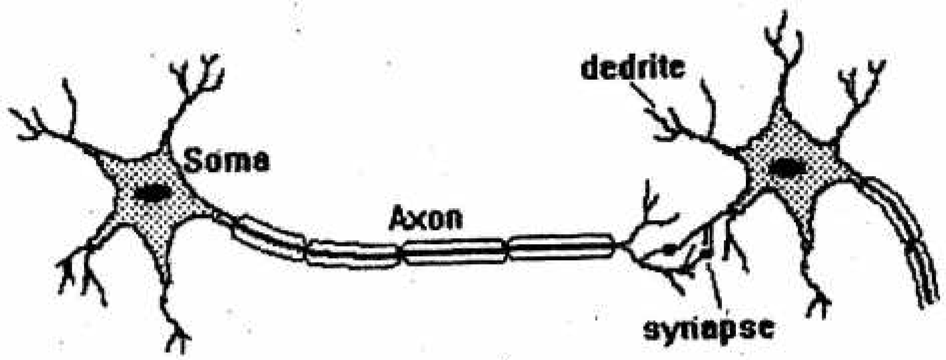


Fig. 1

An axon (figure 2), having a length varying from a fraction of a millimeter to a meter in human body, prolongs from the cell body at the point called axon hillock. At the other end, the axon is separated into several branches, at the very end of which the axon enlarges and forms terminal buttons. Terminal buttons are placed in special structures called the synapses which are the junctions transmitting signals from one neuron to another (figure 3). A neuron typically drive 103 to 104 synaptic junctions.

Fig. The synaptic vesicles holding several thousands of molecules of chemical transmitters, take place in terminal buttons. When a nerve impulse arrives at the synapse, some of these chemical transmitters are discharged into synaptic cleft, which is the narrow gap between the terminal button of the neuron transmitting the signal and the membrane of the neuron receiving it. In general the synapses take place between an axon branch of a neuron and the dendrite of another one. Although it is not very common, synapses may also take place between two exons or two dendrites of different cells or between an axon and a cell body.

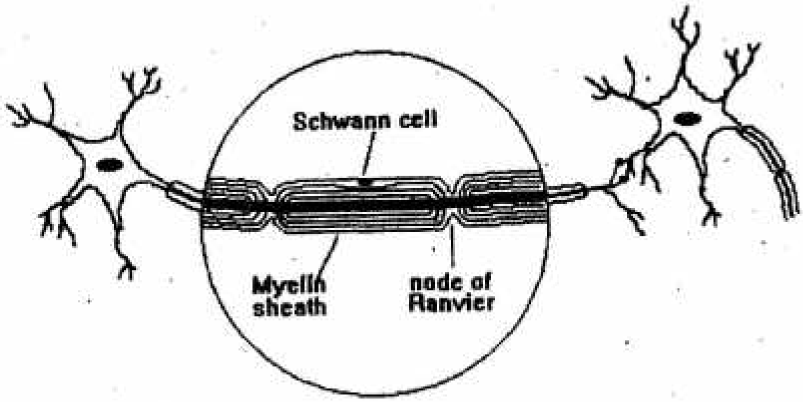


Fig. 2

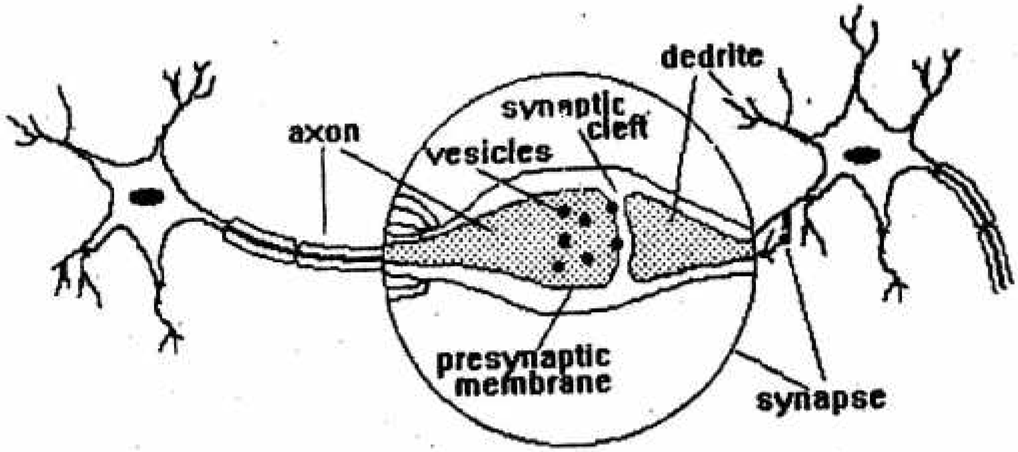


Fig. 3

Neurons are covered with a semi-permeable membrane, with only 5 nanometer thickness. The membrane is able to selectively absorb and reject ions in the intracellular fluid. The membrane basically acts as an ion pump to maintain a different ion concentration between the intracellular fluid and extracellular fluid. While the sodium ions are continually removed from the intracellular fluid to extracellular fluid, the potassium ions are absorbed from the extracellular fluid in order to maintain an equilibrium condition. Due to the difference in the ion concentrations inside and outside, the cell membrane become polarized. In equilibrium the interior of the cell is observed to be 70 millivolts negative with respect to the outside of the cell. The mentioned potential is called the resting potential. A neuron receives inputs from a large number of neurons via its synaptic connections. nerve signals arriving at the presynaptic cell membrane cause chemical transmitters to be released in to the synaptic cleft. These chemical transmitters diffuse across the gap and join to the postsynaptic membrane of the receptor site. The membrane of the postsynaptic cell gathers the chemical transmitters. This causes either a decrease or an increase in the soma potential, called graded potantial, depending on the type of the chemicals released in to the synaptic cleft. The kind of synpases encouraging depolarization is called excitatory and the others discouraging it are called inhibitory synpases. If the decrease in the polarization is adequate to exceed a threshold then the post-synaptic neuron fires. The arrival of impulses to excitatory synapses adds to the depolarization of soma while inhibtory effect tends to cancel out the depolarizing effect of excitatory impulse. In general, although the depolarization due to a single synapse is not enough to fire the neuron, if some other areas of the membrane are depolarized at the same time by the arrival of nerve impulses through other synapses, it may be adequate to exceed the threshold and fire.

Q. 2. Discuss the following learning rules for AHH's :

(i) Widrow-Hoff learning rule,

(ii) Correction learning rule.

Ans. (i) Widrow-Hoff Learning Rule :

The window-hoff learning rule is applicable for the supervised training of NN. It is independent of the activation function of neurons used since it minimizes the squared error between the desired output value d_i and the neuron's activations value $net_1 = w_i^t x$. The learning signals for this rule is defined as follows :

$$r \underline{\underline{\Delta}} d_i - w_i^t x$$

The weight vector increment under. This learning rule is :

$$\Delta w_i = c\left[d_i - w_i^t x\right]x$$

Or for the signal weigh the adjustment is,

$$\Delta w_{ij} = c\left[d_i - w_i^t x\right]x_j, \text{ for } j = 1, 2, \ldots.. n$$

This rule can be considered a special case of the delta learning rule. This rule is sometimes called least mean square [LMS] learning rule. Weights are initialized at any value in this method.

(ii) Correction Learning Rule :

By substituting $r = d_i$ into

$$\Delta w_i(t) = C_r\left[w_i(t), x(t), d_i(t)\right]x(t)$$

We obtain correlation learning rule. The adjustment for the weight vector and the single weights resp, are.

$$\Delta w_i = cd_i x$$

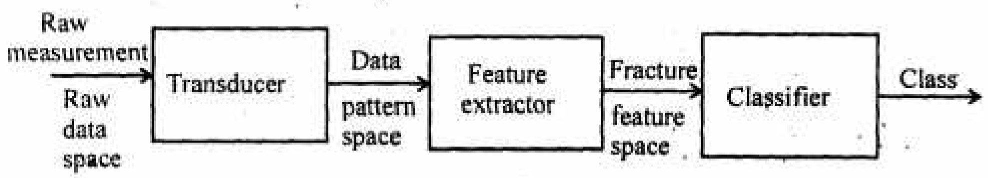
$$\Delta w_{ij} = cd_i x_j \text{ for } j = 1, 2, \ldots\ldots n.$$

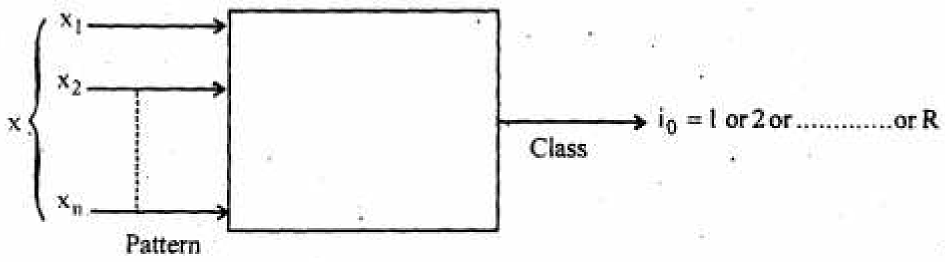
This simple rule states that if d_i is the desired response due to x_i the corresponding weight increase is proportional to their product. The rule typically applies to recording data in memory networks with binary response neurons.

Q. 3. (a) What do you mean by a pattern? Describe any one technique for pattern classification.

Ans. Pattern :

A pattern is a quantitative description of an object, event or phenomenon. The classification may involve spatial and temporal patterns. e.g. spatial patterns are pictures, videos images of ships, characters etc. The goal of pattern classification is to assign a physical object, event or phenomenon to one of the prespecified classes. Despite the lack of any formal theory of pattern perception and classification, human beings and animals have performed these tasks since. The beginning of their existence. The classifying system consists of an input transducer providing. The input pattern data to the feature extractor. Typically, input to the feature extractor are set of data vectors that belong to a certain category. Assume that each such set member consists of real numbers corresponding to measurement results for a given physical situation. Usually the converted data at the output of the transducer can be compressed while still maintaining. The same level of machine performance. The compressed data are called features. The feature extractor at the input of classifier shown in figure performs. The reduction of dimensionality. The feature space dimensionality is postulated to be much smaller than dimensionality of pattern space. The feature vectors retain the minimum number of data dimensions while maintaining the probability of correct classification thus making handling data easier.





Q. 3. (b) Write and explain single. Discrete preceptor training algorithm.

Ans. Single Discrete Perception Training Algo given are P training pairs,

$$\{x_1, d_1, x_2, d_2, \ldots\ldots x_p, d_p\} \text{ where } x_i \text{ is } (n \times 1),\ d_i \text{ is } (1 \times 1),\ i = 1, 2 \ldots\ldots p$$

The argumented input vectors are used :

$$y_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix}, \text{ for } i = 1, 2, \ldots\ldots p.$$

In the following, k denotes the training step and p denotes the step counter within training cycle.

Step 1 : $c > 0$ is chosen.

Step 2 : Weights are initialized at W at small random values, W is $(n + 1) \times 1$. Counters and error are initialized.

$$K \leftarrow 1,\ P \leftarrow 1,\ E \leftarrow 0$$

Step 3 : The training cycle begins here. Input is presented and output completed :

$$y \leftarrow y_p,\ d \leftarrow d_p,\ 0 \leftarrow \text{sign}\left(w^t g\right)$$

$$w \leftarrow w + \frac{1}{2}c[d-0]y$$

Step 5 : Cycle error is computed

$$E \leftarrow \frac{1}{2}(d-0)^2 + E$$

Step 6 : If $p < P$ $p \leftarrow p+1$, $k \leftarrow k+1$ and go to step 3 otherwise go to step 7.

Step 7 : The training cycle is completed. For $E = 0$ terminate the training session. Output weights and k.

If $E > 0$, then $E \leftarrow 0$, $P \leftarrow 1$ and enter the new training cycle by going to step 3.

Q. 4. 'Vrite and explain the "Winner-Take-All" algorithm for cluster classification? What are its separability limitations?

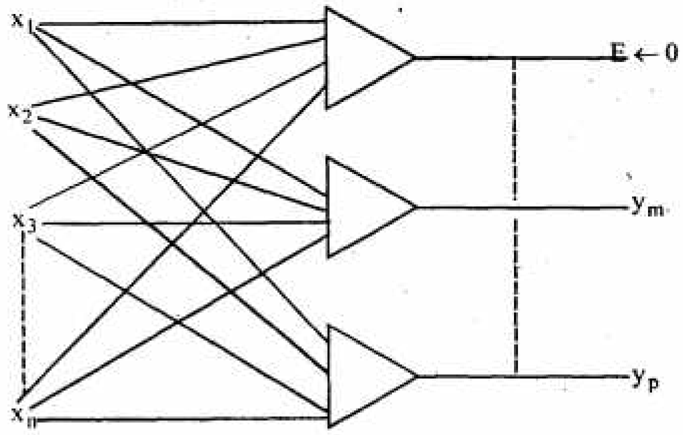
Ans. Winner-Take All Learning :

This network classifies input vectors into one of the specified number of p categories according to the cluster detected in the training set $\{x_1, x_2 .... x_N\}$. The training is performed in an unsupervised mode and the network undergoes the self-organisation process. During the training dissimilar vectors are rejected and only one. The most similar is accepted for weight building. It is possible in training method to assign network nodes to specified input classes in advance. It is equally impossible to predict which neurons will be activated by members of particular cluster at the beginning of the training. This node to cluster assignment is however easily done by calibrating the network after training. The network to be trained is called kohonen network shown in figure. The processing of input data x from training set $\{x_1, x_2 ..... x_n\}$ which represents p clusters follows the customary expression;

$$y = \Gamma[w_x]$$

With diagonal elements of operator Γ being continuous activation functions operating component wise on entries of vector W_x. The processing by the layer of neurons is instantaneous and feed forward. Rearranging matrix W shows.

$$W = \begin{bmatrix} w_1^t \\ w_2^t \\ \vdots \\ w_p^t \end{bmatrix}$$



Prior to learning, the normalization of all weights vectors is required.

$$W_i \triangleq \frac{w_i}{\|w_i\|} \text{ for } i = 1, 2, \ldots.. P.$$

Separability Limitations :

The weights responsible for clustering are marked w_1^f, w_2^f and w_3^f. The scalar product metric based clustering is impossible for distributed patterns. Since only a single neuron has a stroqgest response. The winner-take-all rule produces that same neuron would at same time respond if pattern are on the negative side of the partitioning hyperplane. Adding trainable threshold to neuron's inputs may, improve separability conditions. Even, then, the two neuron network will not be able to learn the solutions of an XOR problem or similar linearly non-separable tasks. An excessive number of neurons created in the winner-take-all layer could certainly also be of significance when learning difficult clustering.

Q. 5. Assume that a linear associator has been designed using the cross-correlation matrix for hetero associative association of p orthogonal patterns. Subsequently another orthogonal patterns $s^{(p+1)}$ associated with $f^{(p+1)}$ must be started. An incremental change in the weight matrix needs to be per formed using the cross-correlation concept. Prove that the association $s^{(p+1)} \rightarrow f^{(p+1)}$ results in no noise term present at O/ϕ.

Ans. The patterns are $\left\{s^{(p+1)}, f^{(p+1)}\right\}$ suppose 'p' association is to be stored in the linear associator. So the vector pair $\left\{s^{(p+1)}, f^{(p+1)}\right\}$ denotes the stored memory stimuli for P = 1, 2, ... ϕ. Since the memory is unidirectional these terms are self explanatory. So we have n-tuple stimuli and m tuple response vectors of $(p+1)^{+n}$ pair :

$$s^{(p+1)} = \left\{s_1^{(p+1)}\, s_2^{(p+1)} \text{-----} s_n^{(p+1)}\right\}^t$$

$$f^{(p+1)} = \left\{f_1^{(p+1)}\, f_2^{(p+1)} \text{-----} f_m^{(p+1)}\right\}^t$$

The objective of linear association is to implement the mapping as follows :

$$f^{(p+1)} = \eta^{(p+1)} = w_s^{(p+1)}$$

Or using the mapping symbol 1

$$s^{(p+1)} \rightarrow f^{(p+1)} + \eta^{(p+1)} \text{ for } P \rightarrow 1, 2, \ldots P$$

Such that the length of noise term vector denoted as η^{p+1} is minimized. The weight update rule for $(p+1)^{th}$ output node and j^{th} input node can be expressed as,

$$w^1_{(p+1)} = w_{(p+1)j} + f_{(p+1)} + s_{(p+1)} \text{ For } \quad P = 1, 2, \ldots,$$

$$j = 1, 2, \ldots n$$

By using outer product formula.

$$w' = w + fS^t$$

W is the weight matrix before update. Initializing weights in unbaised position $W_0 = 0$ we obtain outer product learning rule,

$$w' = f^{(p+1)} s^{(p+1)t}$$

Since there are p pairs to be learned it will be,

$$w' = \sum_{i=1}^{p} f^{(p+1)} s^{(p+1)t}$$

Where w' forms a cross-co-relation matrix so

$$w' = F\delta^t$$

Where F and s are matrices containing vectors of forced response and stimuli and defined as

$$F \underline{\underline{\Delta}} \left[f^{(p+1)}\, f^{(p+2)} - - - - f^{(p+p)} \right]$$

$$S \underline{\underline{\Delta}} \left[s^{(p+1)}\, s^{(p+2)} - - - - s^{(p+p)} \right]$$

The retrieval formula will be

$$V\left(\sum_{i=1}^{P} f^{(p+1)} s^{(p+1)t} \right) S^j$$

Expanding sum of p terms yields

$$V = f^{(p+1)} S^{(p+1)t} S^{(j)} + \ldots\ldots f^{(j)} s^{(j)t} S^{(j)} + \ldots\ldots + f^{(2p+1)} s^{(2p+1)t} s^j$$

According to mapping criteria, the ideal mapping $s^j \rightarrow f^j$ such that no noise term is present and would require

$$V' = f^{(j)}$$

The ideal mapping can be achieved when'v for $P+1 \neq j$

$$S^{(j)t} S^{j} = 1$$

Thus, the orthogonal set of p-input stimuli vectors $\{S^{(p+1)}, S^{(p+2)} ------ S^{(p+1)}\}$ ensures perfect mapping.

Q. 6. For bidirectional associative memory discuss the following :

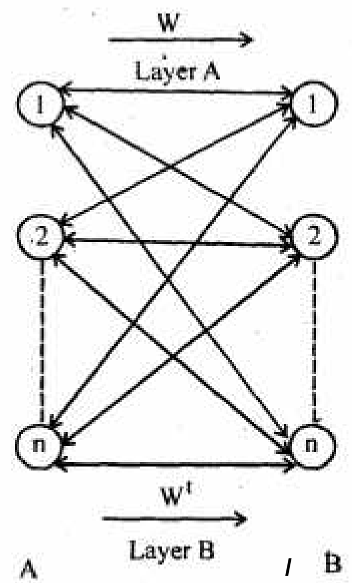
(i) Memory architecture,

(ii) Association Encoding and Decoding,

Ans. (i) Memory Architecture :

The diagram for bi-directional associative memory is shown below. Let us assume that an initialising vector b is applied at input to layer A of neurons. The neurons are assumed to be bipolar binary. The input is processed through the linear connection layer and then through they bipolar threshold functions as follows :

$$a' = \Gamma[Wb]$$



Where $\Gamma[.]$ is a non-linear operator. This pass consists of matrix multiplication and a bipolar threshol operation so that the i'th output is,

$$a_i' = \operatorname{sgn}\left(\sum_{j=1}^{m} w_{ij} b_j\right) \text{ for } i = 1, 2, \ldots\ldots n$$

Assuming threshold in both equations is synchronous and vector a' now feeds the layer B of neurons. It is now processed in layer B through similar matrix multiplication and bipolar thresholds but processing now uses transposed matrix W^t of layer B.

$$b' = \Gamma\left[W^t a'\right]$$

Or

$$b_j' = \operatorname{sgn}\left(\sum_{i=1}^{n} w_{ij} . a_i'\right) \text{ for } j = 1, 2, \ldots. m.$$

(ii) Association Encoding and Decoding :

To coding of information into bidirectional associative memory is done using the customary outer product rule or by addition p cross-correlation matrices. The formula for the weight matrix is

$$W = \sum_{i=1}^{P} a^{(i)} b^{(i)t}$$

Where $a^{(i)}$ and $b^{(i)}$ are bipolar binary vectors which are members of the i^{th} pair.

The weight values will be,

$$W_{ij} = \sum_{m=1}^{P} a_i^{(m)} b_j^{(m)}$$

Suppose one of the stored pattern $a^{m'}$ is presented to memory. The retrieval proceeds as follows :

$$b = \Gamma\left[\sum_{m=1}^{P}\left(b^{(m)} a^{(m)t}\right) a^{(m')}\right]$$

Which further reduces to

$$b = \Gamma\left[nb^{(m')} + \sum_{m \neq m'}^{P} b^{(m)} a^{(m)t} a^{(m')} \right]$$

The net_b vector inside the bracket contains a single term $nb^{(m')}$ additive with the nose term η of value

$$\eta = \sum_{m \neq m'}^{P} b^{(m)} \left(a^{(m)t} a^{(m')} \right)$$

Q. 7. (a) The weight matrix W for a network with bipolar discrete binary neurons is given as :

$$W = \begin{bmatrix} 0 & 1 & -1 & -1 & -3 \\ 1 & 0 & 1 & 1 & -1 \\ -1 & 1 & 0 & 3 & 1 \\ -1 & 1 & 3 & 0 & 1 \\ -3 & -1 & 1 & 1 & 0 \end{bmatrix} \Omega^{-1}$$

Knowing that the thresholds are external inputs of neurons are zero, compute the values of energy, for $v = [-11111]^t$ and $v = [-1-11-1-1]^t$.

Ans. The energy function for

$$E(v) = E(-v) = \frac{-1}{2} V^t W_v$$

So when, $v = [-11111]^t$ the energy values will be,

$$E(v) = -\frac{1}{2} \begin{bmatrix} 0 & 1 & -1 & -1 & -3 \\ 1 & 0 & 1 & 1 & -1 \\ -1 & 1 & 3 & 0 & 1 \\ -1 & 1 & 3 & 0 & 1 \\ -3 & -1 & 1 & 1 & 0 \end{bmatrix}^{[-11111]}$$

After evaluating

$$= -\frac{1}{2}[-4 \quad 0 \quad 6 \quad 6 \quad 4]$$

$$= [2 \quad 9 \quad -3 \quad -3 \quad -2]$$

$$v = [-1 \quad -1 \quad 1 \quad -1 \quad -1]^t$$

Energy value will be $-\frac{1}{2}[2 \quad 0 \quad -4 \quad 2 \quad 4]$

$$= [-1 \quad 0 \quad 2 \quad -1 \quad -2].$$

Q. 7. (b) What do you mean by an attractor? What are its types?

Ans. Attractor :

An attractor is a state towards which the system evolves in time starting from certain initial conditions. Each attractor has its set of initial conditions which initiates the evolution terminating in that attractor. This set of conditions for an attractor is called the basin of attractions. If an attractor is a unique point in state space then it is called a fixed point. If an attractor may consist of a periodic sequence of states, in which case it is called the limit cycle or it may have a more complicated structure.

Q. 8. Write the algorithm and also draw the flowchart for "Error Back Propagation Training Algorithm" for two layer network.

Ans. The algorithm for error-back propagation training is as follows :

Given are P training pairs $\{z_1, d_1, z_2, d_2 \text{-----} z_p, d_p\}$, where z_i is $(I \times 1)$, d_i is $(K \times 1)$ and i = 1, 2,..... P. Note that I'th component of each z_i is of value -1 since input vectors have been argumented. Size J-1 of the hidden layer having output y is selected. The j^{th} component of y is value -1 since hidden layer outputs have also been argumented y is $(J \times 1)$ and 0 is $(K \times 1)$.

Step 1 : $\eta > 0$, E_{max} chosen. Weights W and V are initialized at small random values W is $(K \times J)$, V is $(J \times I)$.

$$q \leftarrow 1, \; P \leftarrow 1, \; E \leftarrow 0$$

Step 2 : Input is presented and layers output computed.

$$z \leftarrow z_p,\ d \leftarrow d_p$$

$$y_j \leftarrow f\left(V_j^+ z\right) \text{ for } j = 1, 2, \ldots\ldots J$$

Where V_j a column vector, is the k'th row of w.

Step 3 : Error value is computed.

$$E \leftarrow \frac{1}{2}(d_k - o_k)^2 + K, \text{ for } K = 1, 2, \ldots K.$$

Step 4 : Error signal vectors δ_o and δ_y of both layers are computed. Vector δ_o is $(K \times 1)$ δ_y is $(J \times L)$. The error signal terms of the output layer in this step are :

$$\delta_{ok} = \frac{1}{2}(d_k - o_k)\left(1 - o_k^2\right), \text{ for } K = 1, 2, \ldots\ldots K.$$

The error signal terms of the hidden layer in this step are :

$$\delta_{yj} = \frac{1}{2}\left(1 - y_j^2\right)\sum_{k=1}^{k} \delta_{ok} W_{kj} \text{ for } J = 1, 2, \ldots. J.$$

Step 5 : Output layer weights are adjusted $W_{kj} \leftarrow W_{kj} + \eta \delta_{ok} y_i$ for $K = 1, 2, \ldots\ldots K$ and $j = 1, 2, \ldots\ldots J$

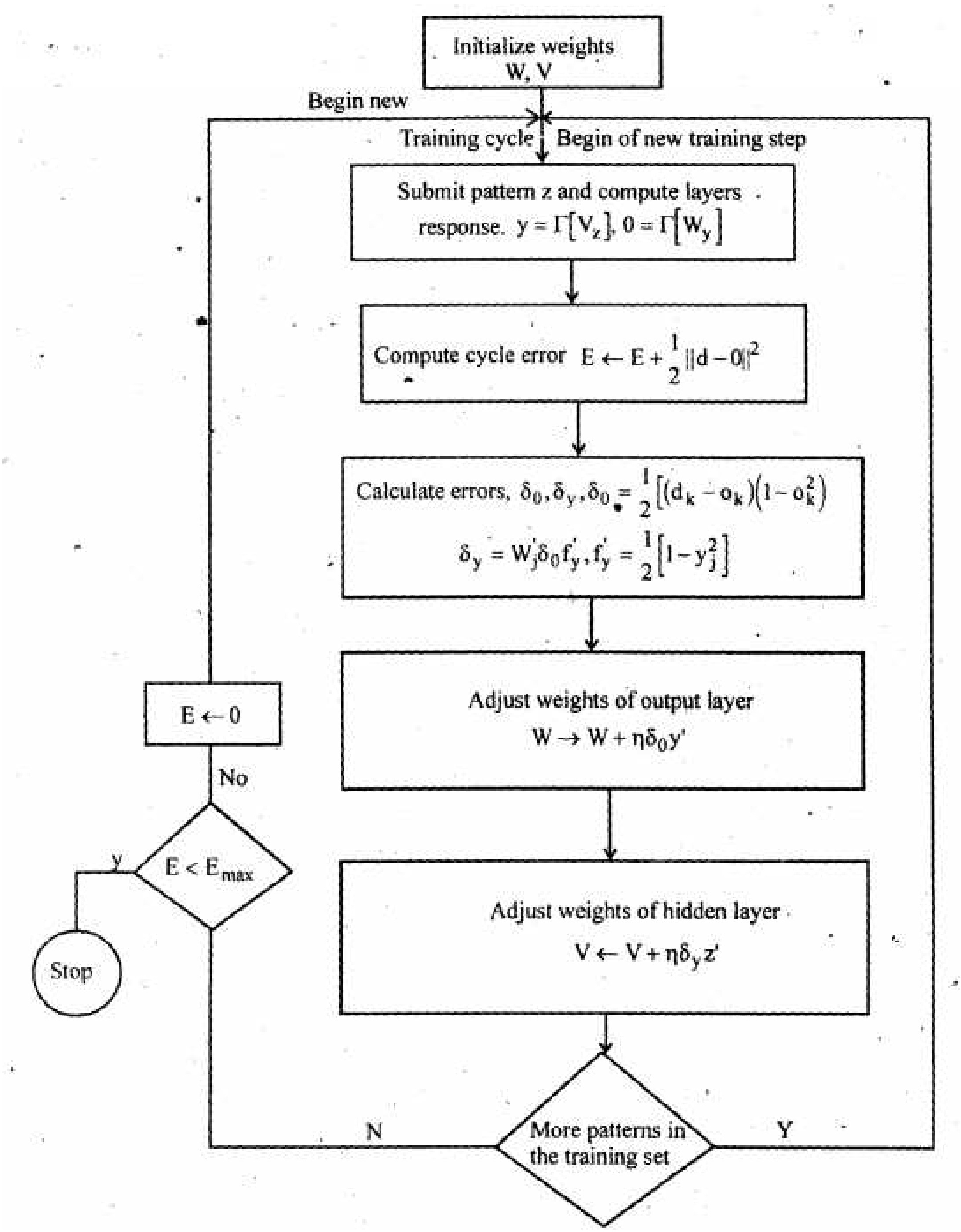
Step 6 : Hidden layer weights are adjusted

$$V_{ji} = V_{ji} + \eta \delta_{yj} z_i \text{ for } j = 1, 2, \ldots\ldots J \text{ and } i = 1, 2, \ldots\ldots I.$$

Step 7 : If $p < P$ then $p \leftarrow p + 1$. $q \leftarrow q + 1$ and go to step 2 otherwise go to step 8.

Step 8 : The training cycle is completed. For $E < E_{max}$ terminate the training session. Output weights W, V, q and E. If $E > E_{max}$ then $E \leftarrow 0$, $P \leftarrow 1$ and initiate the new training cycle by going to step 2.

Flow chart :



Initialize weights W, V
Begin new
Training cycle
Begin of new training step
Submit pattern z and compute layers response. $y = \Gamma[V_z], 0 = \Gamma[W_y]$
Compute cycle error $E \leftarrow E + \frac{1}{2}\|d - 0\|^2$
Calculate errors, $\delta_0, \delta_y, \delta_0 = \frac{1}{2}\left[(d_k - o_k)(1 - o_k^2)\right)$
$\delta_y = W_j'\delta_0 f_y', f_y' = \frac{1}{2}\left[1 - y_j^2\right]$
Adjust weights of output layer
$W \rightarrow W + \eta\delta_0 y'$
Adjust weights of hidden layer
$V \leftarrow V + \eta\delta_y z'$
More patterns in the training set
N
Y
$E \leftarrow 0$
No
$E < E_{max}$
y
Stop