Tutorial 1

1. An example of the logistic function is defined by

   $$\varphi (v) = 1/ 1+ \exp(-av)$$

   Whose limiting values are 0 and 1. Show that the derivative of $\varphi(v)$ with respect to v is given by

   $$d\varphi/dv = a\varphi(v) [1 - \varphi(v)]$$

   What is the value of this derivative at the origin?
2. An odd sigmoid function is defined by

   $$\varphi (v) = (1- \exp(-av))/ 1+ \exp (-av) = \tanh(av/2)$$

   Where tanh denotes a hyperbolic tangent. The limiting values of this second sigmoid function are –1 and +1. Show that the derivative of $\varphi(v)$ with respect to v is given by $d\varphi/dv = a/2[1-\varphi^2(v)]$

   What is the value of the derivative at the origin? Suppose that the slope parameter a is made infinitely large. What is the resulting form of $\varphi(v)$?
3. An odd sigmoid function is the algebraic sigmoid

   $$\varphi (v) = v/(1+v^2)^{1/2}$$

   Whose limiting values are –1 and +1. Show that the derivative of $\varphi(v)$ with respect to v is given

   by $d\varphi/dv = \varphi^3(v)/v^3$

   What is the value of this derivative at the origin?
4. consider the following two functions
   (i) $\varphi(v) = 1/(2\Pi)^{1/2} \int \exp (- x^2/2) dx$
   (ii) $\varphi(v) = 2/ \Pi \tan^{-1}(v)$

   Explain why both of these functions fit the requirements of a sigmoid function. How these two functions differ from each other.
5. Which of the five sigmoid functions described in above problems will qualify as probability distribution functions. Justify.
6. Consider the pseudo linear activation function $\varphi(v)$ shown in fig. 1 & 2 for both
   a) Formulate $\varphi(v)$ as a function of v.
   b) What happens to a if a is allowed to approach zero.
7. A neuron has an activation function $\varphi(v)$ defined by the logistic function

   $$\varphi (v) = 1/ 1+ \exp(-av)$$

   Where v is the induced local field and the slope parameter a is available for adjustment. Let x1,x2,….xm denote the input signals applied to the source nodes of the neuron, and b denotes the bias. For convenience of presentation we would like to absorb the slope parameter a in the induced local field v by writing

$\varphi (v) = 1/ 1+ \exp(-v)$

How would you modify the inputs x1,x2,…….xm to produce the same output as before. Justify

8. A neuron j receives inputs from four other neurons whose activity levels are 10, -20 , 4 and –2. The respective synaptic weights of neuron j are 0.8,0.2, -1.0 and 0.9. Calculate the output of neuron j for the following two situations

   a) the neuron is linear
   b) the neuron is represented by the McCulloch Pitts Model

   Assume that the bias applied to the neuron is zero.

9. Repeat question 8 with the neuron model based on the logistic function

$$\varphi\,(v) = 1/\,1+\exp(-v)$$

10. A fully connected feed forward network has 10 source nodes, 2 hidden layers, one with 4 neurons and the other with 3 neurons, and a single output neuron. Construct an architectural graph of this network.

11. Construct a fully recurrent network with 5 neurons but with no self-feed back.

12 A recurrent network has 3 source nodes 2 hidden neurons and 4 output neurons. Construct an architectural graph that describes such a network.

13 Figure shows the signal flow graph of a 2-2-2-1 feed forward network. The function φ(.) denotes a logical function. Write the input output mapping defined by this network.

14 Suppose that the output neuron in the signal flow graph operates in its linear region. Write an input-output mapping defined by this new network.

15 Suppose that we add biases equal to –1and +1 to the top and bottom neurons of the first layer and +1 and –2 are applied to the top and bottom neurons of the second hidden layer. Write the new form of input output mapping defined by this new network.

16 Consider a multiplayer feed forward network, all the neurons of which operate in their linear regions. Justify that such network is equivalent to a single layer feed forward network.

17 A useful form of preprocessing is based on the autoregressive model described by the difference equation (for real valued data)

$$Y\,(n) = w_1 y(n-1) + w_2 y(n-2) + \ldots\ldots + w_M y(n-M) + v(n)$$

Where y(n) is the model output v(n) is a sample drawn from a white noise process of zero mean and some prescribed variance. $W_1, w_2, \ldots\ldots, w_M$ are the AR model coefficients and M is the model order. Show that the use of this model provides two forms of geometric invariance (a) scale (b) time translation.

How could these two invariances be used in neural networks?

18 The delta rule described by equation $\Delta w_{kj}(n) = \eta\, e_k(n) x_j(n)$ and hebb's rule described by equation $\Delta w_{kj}(n) = \eta\, y_k(n) x_j(n)$ represent two methods of learning. List the features that distinguish these two rules from each other.

19 The error correction learning rule may be implemented by using inhibition to subtract the desired response (target value) from the output and then applying the anti Hebbian Rule. Discus this interpretation of the error correction learning.

20 Consider a group of people whose collective opinion on a topic of interest is defined as the weighted average of the individual opinion of its members. Suppose that if over the course of time, the opinion of a member in the group tends to agree with the collective opinion of the group the opinion of that member is given more weight if on the other hand the particular member disagrees with the collective opinion of the group that member's opinion is given less weight. This form of weighting is equivalent to positive feedback control, which has the effect of producing a consensus of opinion among the group. Discuss the analogy between the situation described and Hebb's postulate of learning.

21 An input signal of unit amplitude is applied repeatedly to a synaptic connection whose initial value is also unity. Calculate the variation in the synaptic weight with time using the following two rules

a) The simple form of Hebb's rule assuming the learning rate parameter $\eta = 0.1$.
b) The covariance rule assuming that the presynaptic activity is zero & postsynaptic activity is one.

22 Formulate the expression for the output y_j of neuron j in the network of figure. We can use the

Following x_i = ith input signal

W_{ji} = synaptic weight from input I to neuron j

C_{kj} = weight of lateral connection from neuron k to neuron j

V_j = induced local field of neuron j

$Yj = \varphi(v_j)$

What is the condition that would have to be satisfied for neuron j to be the winning neuron?

23 Unsupervised learning can be implemented in an off-line or on-line fashion. Discuss the physical implications of these two possibilities.

24 Consider the difficulties that a learning machine faces in assigning credit for the outcome of a game of chess. Discuss the notions of temporal credit assignment and structural credit assignment in the context of this game.

25 A supervised learning task may be viewed as a reinforcement learning task by using the reinforcement signal some measure of the closeness of each response of the system to the desired response. Discuss the relationship between supervised learning and reinforcement learning.

26 Consider the following orthonormal set of key patterns, applied to a correlation matrix memory

$X1 = [1, 0, 0, 0]^T$

$X2 = [0, 1, 0, 0]^T$

$X3 = [0, 0, 1, 0]^T$

The respective stored patterns are

$Y1 = [5, 1, 0]^T$

$Y2= [-2,1, 6]^T$

$Y3= [-2, 4 ,3]^T$

Calculate the memory matrix M.

27 Show the memory associates for the above matrix perfectly.

28 Consider again the correlation matrix memory . the stimulus applied to the memory is noisy version of the key pattern x_1 as shown by

$X = [0.8, - 0.15 , 0.15 , -0.20]^T$

(a) Calculate the memory response y.

(b) Show that the response y is closest to the stored pattern y_1 in Euclidean sense.

29 To which of the two paradigms, learning with a teacher & learning without a teacher, do the following algorithms belong? Justify

a) nearest neighbor rule

b) K-nearest neighbor rule

30 Whether Hebbian learning and/or Boltzmann learning belong to supervised or unsupervised learning. Justify?

31 Explore the method of steepest descent involving a single weight w considering the following cost function $(w) = ⅓\sigma^2 - r_{xd}w + ½ r_x w^2$ where σ, r_{xd}, r_x are constants

32 Consider the cost function $(w) = 1/2\sigma^2 - r^T_{xd}w + ½ wR_x w^T$

Where σ^2 is some constant , and $r_{xd} =$ .8182

0.354

$R_x =$ 1 .8182

.8182 1

(a) find the optimum value w* for which C(w) reaches its minimum value.

(b) Use the method of steepest descent to compute w* for the following two values of learning rate parameter $\eta = 0.3 , 1.0$

33. The correlation matrix R_x of the input vector x(n) in the LMS algorithm is defined by

$$R_x = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Define the range of values for the learning rate parameter η of the LMS algorithm for it to be convergent in the mean square.

34 Consider a linear predictor with its input vector made up of samples x(n-1), x(n-2),…… x(n-m), where m is the prediction order. The requirement is to use the prediction x(n) of the input sample x(n). Set up the recursions that may be used to compute the tap weight w1, w2,…. W_m of the predictor.

35 Use the back propagation algorithm for a set of synaptic weights and bias levels for a neural network to solve the XOR problem. Assume the use of a logistic function for non linearity.

36 We assume that a multi layer perceptron classifier's outputs provides estimate of the posteriori class probabilities. This property assumes that the size of the training set is a large enough and the back propagation algorithm used to train the network does not get stuck at a local minimum. Fill in the mathematical detail of this property

37 What you mean by soft computing. Explore the various areas & do their comparison.

38 What is Fuzzy computing? Describe the fuzzy set theory with the help of membership function.

39 Find the weights required to perform the following classifications : Vectors (1,1,1,1) and ((1,-1,-1,1) are not members of the class (and have target value –1). Using each of training vectors as input, test response of the net using Hebb rule.

40 Using perceptron learning rule Find the weights required to perform the following classifications : Vectors (1,1,1,1) and ((1,-1,-1,1) are not members of the class (and have target value –1). Using each of the training vectors as input, test the response of the net.

41 Design a McCulloch to model the **perception of simple musical patterns**. Use three inputs to correspond to three pitches, do, re and mi. Assume that only one pitch is presented at any time. Use two outputs to correspond to perception of an upscale segment and a downscale segment, specifically:

a. the pattern of inputs do at time t = 1, re at time t=2 and mi at time t=3 should elicit a positive response from the upscale segment unit

b. the pattern of inputs mi at time t=1, re at time t=2, and do at time t=3 should elicit a positive response from the downscale segment unit; any other patterns of inputs should generate no response

42 It may be argued that cross validation is a case study in structural risk minimization. Describe a neural network example using cross validation that supports this argument.

43 In multifold cross validation there is no clear separation between the training data & test data as there is in the hold out method. Is it possible for the use of multifold cross validation to produce a biased estimate? Justify your answer.

44 What is the difference between neural net computer & a Von _Neumann Machine?

45 What are the problems to be addressed in the recurrent neural networks?

46 How you ca generalize learning. Give some innovative ways.

47 In a linear associator, if input/output prototype vectors are $I_1= [\ 1\ \ -1\ \ 1\ \ -1]^T$, $t_1 = [1\ –1]^T$, $I_2 = [1\ 1\ –1\ –1]^T$ $t_2 = [\ 1\ 1]^T$. Use the Hebb's rule to find the appropriate weight matrix for this linear associator.

48 Consider following prototype patterns shown below: Design an auto associator for these patterns. Use Hebb's Rule : $I_1 = [\ 1\ \ 1\ \ -1\ \ 1\ \ -1\ \ -1]^T$, $I_2 = [\ -1\ \ 1\ \ 1\ \ 1\ \ 1\ \ -1]^T$.

49 If $I_1= [\ 1\ 1]^T$, $t_1 = 1$ and $p_2 = [\ 1\ -1]^T$, $t_2 = -1$ represents input & target pairs. Train the network using the delta rule with the initial guess set to zero and a learning rate of 0.5.

50 Consider a 3-layer perceptron with three inputs, three hidden layers & one output unit. Given the initial weight matrix for hidden and output nodes as:

$$W_H = [2\ \ 1\ \ \ 0\ \ \ \ \ 1\ \ \ 2\ \ \ \ 2\ \ \ 0\ \ \ 3\ \ \ \ \ 1] \text{ and } W_o = [-1\ \ \ 1\ \ \ 2\ ]$$

If input vector is I = (3   4 0), calculate the output.

51 A neural network is trained by using the capital letters T and L and their rotated versions. Each letter is represented on a figure with 9 squares arranged in 3 rows and 3 columns to form a large square. Each of these smaller squares is either filled or unfilled as needed to represent the letter. Each letter is rotated clockwise through 90^0, 180^0 and 270^0. Taking such 8 exemplars, design a feed forward neural network with one hidden layer so that the network is trained to produce an output value 1 when the input pattern is either T or any other its rotated versions and 0 otherwise.