Answer THREE questions.

1.

a) Give mathematical definitions of

(i) the step function $\theta(x)$

(ii) the sigmoid firing function $f(x)$

as functions of their real-valued argument x, and sketch each of the functions.

[7 marks]

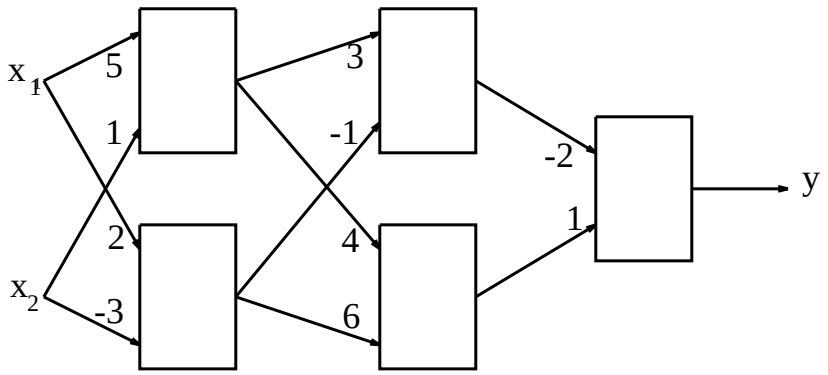
b) The replacement of the step function by the sigmoid firing function was a crucial factor in the development of a successful multilayer perceptron learning algorithm. Explain why.

[6 marks]

c) Write down an expression which relates the firing state y of a binary decision neuron (BDN) to the signals x_i, i = 1..n, on each of its n input lines.

[4 marks]

d) The figure below shows a 3-layer BDN network, where the numbers attached to the input lines indicate the weights associated with them.



Construct a truth table listing the final layer output y for each possible input pair (x_1, x_2) in the case that

(i) all thresholds are zero

[8 marks]

(ii) thresholds equal to +1 and -1 are applied to the top and bottom neurons of the first hidden layer, and thresholds equal to -1 and +2 are applied to the top and bottom neurons of the second hidden layer.

[8 marks]

TURN OVER

2.

a) Give a careful definition of each of the following terms

(i) linear separability

(ii) feature detector

(iii) generalisation

(iv) momentum

[16 marks]

b) Write down an expression for Δw_{ij}^{l}, the update to the weight associated with the jth input line to neuron i in layer *l*, for an error backpropagation network in the case that

(i) *l* is the output layer

(ii) *l* is a hidden layer

[7 marks]

c) An error backpropagation network with sigmoidal transforms applied to all units is to be trained to classify upper case characters using one exemplar for each of the 26 classes. There are 26 output units, and the desired response to the ith alphabetic character is for the ith output unit to produce a '1' and all others to produce a '0'.

(i) Should the number of hidden units be set to 26? Why or why not?

[5 marks]

(ii) Explain why, for *any* number of hidden units, it would be unreasonable to expect the network to achieve zero classification error.

[5 marks]

CONTINUED

3.

a) Give examples of *two* neural networks in which network *learning* is described by a process of gradient descent. In each case write down an expression for the error function which is thus reduced, defining all terms used.

[8 marks]

b) Give an example of a neural network in which the *evolution of firing states* is described by a process of gradient descent. Write down an expression for the quantity thus reduced in this case, defining all terms used.

[4 marks]

c) Why are gradient descent algorithms vulnerable to trapping in *local minima*? In each of the cases a) and b) above, explain what type of situation (in training or recall) trapping in an undesirable local minimum would correspond to.

[8 marks]

d) A multilayer feedforward network is to be trained using the A_{RP} reinforcement rule rather than the more conventional error backpropagation.

(i) Using the result that

$$< \Delta w_{ij}^{l}(t) > \; = \; \eta \frac{\partial}{\partial w_{ij}^{l}} < r(t) >$$

show that a suitable definition of reward probability allows the A_{RP}-trained network to perform a process of stochastic gradient descent in error-weight space.

[7 marks]

(ii) Explain why a multilayer network trained in the above way may be less vulnerable to trapping in local minima than one trained by error backpropagation.

[6 marks]

4.

a)

(i) Suppose that the operation of a recurrent N-node net is given by the system of equations $\underline{x}(t+1) = F(\underline{x}(t))$, where F is the "next state" function. What is meant by the statement that $\underline{\tilde{x}} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N)$ is a fixed point of the system?

[3 marks]

(ii) If the fixed point $\underline{\tilde{x}}$ is to be regarded as an output, what variable(s) could play the role of an input?

[3 marks]

b) In the context of the Hopfield net, define the terms

(i) energy function

(ii) basin of attraction

(iii) content-addressable memory (CAM)

[9 marks]

c) The Hopfield net firing rule is given by

$$x_i(t+1) = h\Big(\sum_{i \neq j} w_{ij} x_j(t) - s_i, x_i(t)\Big)$$

where

$$h(a, x) = \begin{cases} 1 & \text{if } a > 0 \\ x & \text{if } a = 0 \\ 0 & \text{if } a < 0 \end{cases}$$

It is usually applied *asynchronously*, giving each of the N neurons an equal probability of 1/N of being allowed to update their state at each time step.

Consider the 2-node Hopfield net with single weight

$$w = w_{12} = w_{21} = 2$$

and thresholds

$$s_1 = 1, \quad s_2 = 1$$

(i) Using the usual asynchronous firing state update rule, draw a state transition diagram, labelling transitions with their probabilities and showing the energy levels of the system.

[8 marks]

(ii) Suppose now that the nodes change their states *simultaneously*. Draw a state transition diagram for the new system. Comment on the change in CAM properties caused by the use of simultaneous update in this network.

[10 marks]

CONTINUED

5.

a) Give a careful definition of each of the following terms

(i) self-organisation

(ii) "winning" node (in a competitive learning algorithm)

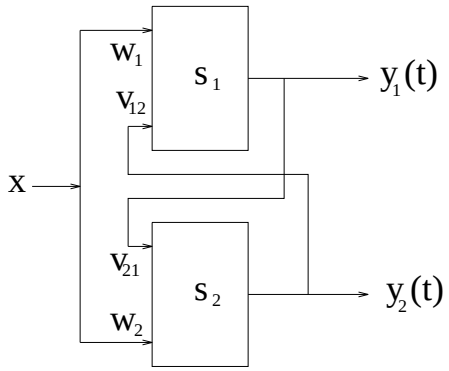
(iii) topographic map

[9 marks]

b) In the Kohonen map, why is it important that neighbouring neurons (those within a certain spatial distance of the winner) also have their weights updated in a similar way? Why must the size of this update neighbourhood be decreased as training progresses?

[7 marks]

c) The figure below shows a simple example of a competitive network



where

s_i = threshold of neuron i

w_i = synaptic weight from input to neuron i

v_{ij} = weight of lateral connection from neuron j to neuron i

a_i = linear activation function of neuron i

$y_i = a_i$ = *linear* output function of neuron i

(i) Write down expressions for $y_1(t)$, $y_2(t)$, the outputs of neurons 1 and 2 at time t, in terms of their external and lateral inputs, weights and thresholds.

[4 marks]

(ii) Suppose that after a certain time y_1 and y_2 reach stable values for a given input x. By solving a pair of simultaneous linear equations obtain expressions relating the y_1, y_2 to the input x and to the output function parameters (the w_i, v_{ij} and s_i).

[5 marks]

Suppose that $x = 1$, $s_1 = s_2 = 0$, $w_1 = 1$, $w_2 = -1$.

(iii) Using the fixed point output solutions obtained in (ii), derive a condition on the lateral weights which would allow neuron 2 to be the "winner" for this input.

[3 marks]

(iv) Assume additionally that $y_1(0) = y_2(0) = 0$. What would happen if $v_{12} = v_{21} = 1$?

[5 marks]

END OF PAPER