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.

[6 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.

[7 marks]

c) In the multilayer perceptron the sigmoid function is used to denote a neural firing rate. Give two examples of networks in which this same function is used to denote a neural firing probability.

[4 marks]

d)

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

[3 marks]

(ii) Show that by introducing the idea of a 'bias input' and 'bias weight' that the BDN output rule of (i) can be written in a format which avoids the use of a special notation for the threshold variable. Why is this alternative format frequently preferred?

[5 marks]

e) The boolean functions NAND and NOR are defined by

$$\mathrm{NAND}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2 = 1 \\ 1 & \text{otherwise} \end{cases} \qquad \mathrm{NOR}(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 = x_2 = 0 \\ 0 & \text{otherwise} \end{cases}$$

Obtain for each of these functions a weight vector $\underline{w} = (w_1, w_2)$ and threshold s that would enable the function to be computed by a BDN.

[8 marks]

2.

a) Explain the essential differences between programming a computer and training a neural network. What kind of information is needed for each task? What are the advantages and disadvantages of the conventional rule-based and neural network approaches? Give examples of the kinds of problems you think would be best suited to each approach.

[12 marks]

b) There are three broad classes of neural learning algorithm, each appropriate to different circumstances. For each of the following problems, suggest a suitable type of algorithm, justifying your choice with a careful argument:

(i) Training a neural network to recognise the letters of the alphabet from a set of handwritten samples.

(ii) Teaching a robot to juggle.

(iii) Discovering the number of different speakers represented in a database containing many examples of the same spoken phrase.

[12 marks]

c) Explain why training algorithms based on a process of gradient descent, such as error backpropagation, can be vulnerable to being trapped in local minima.

[5 marks]

(d) Boltzmann training is also a form of gradient descent, but Boltzmann nets are much less likely to be trapped in local minima than error backpropagation networks. Explain why this is so.

[4 marks]

CONTINUED

3.

a) It is desired to store a set of P N-bit binary patterns

$$\underline{x}^{(p)} = (x_1^{(p)}, x_2^{(p)}, ..., x_N^{(p)}) \quad p = 1..P$$

in an N-node Hopfield net.

(i) Write down the simplest storage prescription for the $\frac{N}{2}(N-1)$ weights w_{ij}, linking each neuron symmetrically to its N-1 neighbours, which will achieve this (assume that the neuron thresholds are to be set to zero).

[4 marks]

(ii) What are the disadvantages of setting the thresholds to zero, as in the storage prescription above?

[4 marks]

(iii) Write down an extension to the storage prescription of (i) which allows the problems associated with zero thresholds to be overcome.

[4 marks]

b) Write down an output update rule for Hopfield neurons, which gives the new binary state of a neuron in terms of its own and its neighbours' previous states (assume that the thresholds are not to be set to zero).

[4 marks]

c) It is desired to store the single binary pattern (1,0) in a 2-node Hopfield net.

(i) Using the storage prescription which sets thresholds to non-zero values, show that in this case the Hopfield energy function is given by

$$H(x_1, x_2) = x_1x_2 - x_1 + x_2$$

Assuming *asynchronous* update, draw a state transition diagram, labelling all transitions with their probabilities and showing the energy levels of the system

[10 marks]

(ii) Suppose in the above 2-node network that the nodes now change their states *simultaneously*, ie that *synchronous* update is now used. Draw a state transition diagram for the new system. Comment on any change in the dynamical behaviour of the system.

[7 marks]

4.

a) Describe how a Hopfield net with suitably chosen energy function may be used to obtain good solutions to difficult optimisation problems. Explain how the use of a stochastic firing rule helps the net avoid being trapped in firing states which correspond to poor solutions.

[10 marks]

b) Explain the significance of the Markov transition matrix M, whose elements are denoted $M_{\underline{x},\underline{y}}$ (where the n-bit binary vectors $\underline{x}$ and $\underline{y}$ denote firing states of an N-neuron system). Why is $\sum_{\underline{x}} M_{\underline{x},\underline{y}} = 1$? Which elements of M are *always* zero for a stochastic Hopfield net, and why?

[7 marks]

c) In a stochastic Hopfield net $p_i(t)$, the probability of unit i being in firing state $x_i = 1$ at time t, is given in terms of the firing states $x_j(t-1) \in \{0,1\}$ $(j \neq i)$, unit i's weights and threshold, and the parameter β by

$$p_i(t) = \frac{1}{1 + e^{-\beta\left(\sum_{j \neq i} w_{ij} x_j(t-1) - s_i\right)}}$$

Consider a 2-neuron net with weights $w_{12} = w_{21} = -1$ and thresholds $s_1 = s_2 = 0$.

(i) Write down expressions for $p_1(t)$ and $p_2(t)$ for general values of the parameter β.

[4 marks]

(ii) Work out the transition matrix component $M_{\underline{x},(01)}$ for each of $\underline{x}$ = (00), (01), (10), (11)

- for general values of β
- for 'temperature' T=1
- for 'temperature' T=0.1

[12 marks]

CONTINUED

5.

a)

(i) Explain what is meant by *convergence in the mean* for a stochastic learning system such as the associative reward-penalty (A_{RP}) model.

[5 marks]

(ii) An A_{RP} system is presented with a set of R context vectors $X = \{\underline{x}^{(1)}, \underline{x}^{(2)}, \ldots, \underline{x}^{(R)}\}$, occurring with probabilities $\xi^{(i)}$, $i = 1..R$. State the conditions which must hold in order for the learning process to properly converge.

[8 marks]

b) The output of an A_{RP} unit depends stochastically on its input $\underline{x}$ and on its weight vector $\underline{w}$. Write down an expression for the firing probability of an A_{RP} unit with n external inputs $x_1,..,x_n$ and weights $w_0,..,w_n$.

[4 marks]

c) Consider the reinforcement task for a single 2-input A_{RP} neuron defined by the table below:

x_1	x_2	$d_0(\underline{x})$	$d_1(\underline{x})$
0	0	0.9	0.2
0	1	0.1	0.7
1	0	0.3	0.8
1	1	0.8	0.4

$d_y(\underline{x})$ is the probability of the neuron receiving an environmental reward for action y in context $\underline{x}$. Assume that initially the neuron has the parameter vector $\underline{w} = (0,0,0)$ and that all four patterns are equally likely to be seen.

(i) Work out the value of the initial performance measure M_{init} in this case.

[4 marks]

(ii) What is the value of the maximal performance measure M_{max}?

[4 marks]

(iii) If A_{RP} training was used to solve this task, how would the value of the learning parameter λ affect the neuron's ability to reach a final performance level approaching M_{max}? Why would it *not* be a good idea to set $\lambda = 0$?

[8 marks]

END OF PAPER