Answer THREE questions.

1.

a) Explain the difference 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.

[10 marks]

b) If the processing units of an artificial neural network compute their outputs according to a clearly-defined mathematical function, why is it that the operation of neural networks has been frequently referred to as "non-algorithmic"? If the *learning law* as well as the output function is supplied, should the network *then* be considered an algorithmic computing machine? Are biological neural networks algorithmic computers?

[7 marks]

c)

(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.

[4 marks]

d) The Widrow-Hoff delta rule, which relates the change to be made to the jth weight after presentation of the n-dimensional pattern $\underline{x}_p$ during training epoch t to the difference between the desired output d_p and the actual output $y_p(t)$, is given by

$$\Delta w_{j,p}(t) = \eta(d_p - y_p(t))x_{j,p}$$

Consider a single-neuron network with initial (epoch 0) weight vector $\underline{w}(0) = (0,1,0)$, which is required to respond to a 2-pattern training set in the following way:

$$\underline{x}_1 = (1,0) : \text{desired output } d_1 = 0$$

$$\underline{x}_2 = (0,1) : \text{desired output } d_2 = 1$$

Using the BDN firing rule to compute outputs and the Widrow-Hoff rule with $\eta = 1$ to change the neuron's weights, repeat the training sequence $\{\underline{x}_1, d_1\}$, $\{\underline{x}_2, d_2\}$ until two consecutive correct responses are achieved. After how many epochs is the overall error reduced to zero? What is the final weight vector?

[9 marks]

2.

a) Explain using a geometrical argument why *single-layer* perceptrons can only compute a limited range of functions. Use the 2-input Boolean function XOR to illustrate your argument.

[12 marks]

b) Why are *multi-layer* perceptrons able to overcome this restriction?

[5 marks]

c) Consider the 2-dimensional pattern set

$$\underline{x}_1 = (0.6, 0.9),\ \ \underline{x}_2 = (0.8, 0.8),\ \ \underline{x}_3 = (0.6, 0.6),\ \ \underline{x}_4 = (0.7, 0.4)$$

$$\underline{x}_5 = (0.2, 0.4),\ \ \underline{x}_6 = (0.4, 0.3),\ \ \underline{x}_7 = (0.2, 0.1),\ \ \underline{x}_8 = (0.5, 0.1)$$

Patterns $\underline{x}_1$ to $\underline{x}_4$ are to be classified as Class A, patterns $\underline{x}_5$ to $\underline{x}_8$ as Class B.

(i) Demonstrate using a graphical argument that a single 2-input binary decision neuron (BDN) with real-valued inputs $\underline{x}_i \in [0,1]^2$ and a binary output $y \in \{0,1\}$ would be able to successfully separate these two pattern classes.

[8 marks]

(ii) Suggest a set of weights w_1, w_2 and a threshold s for the BDN neuron that would enable this pattern separation to be accomplished. Assume that Class A corresponds to $y = 1$ and Class B to $y = 0$.

[4 marks]

(iii) Consider the test patterns

$$\underline{x}_9 = (0.2, 0.6),\ \ \underline{x}_{10} = (0.3, 0.8)$$

Using the weights and threshold you have chosen, decide whether these patterns would be assigned to Class A or to Class B.

[4 marks]

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).

[3 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.

[3 marks]

b)

(i) 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.

[3 marks]

(ii) Explain why it is preferable that only one neuron be allowed to update its state at a time (asynchronous update).

[4 marks]

c) It is desired to store the 3-bit binary patterns (0,0,1), (0,1,0) and (1,1,1) in a 3-node Hopfield net.

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

$$H(x_1, x_2, x_3) = -x_1x_2 - x_1x_3 + x_2x_3 + x_1 - x_2 - x_3$$

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

[12 marks]

(ii) What would happen if one tried to store a fourth pattern, equal to (1,0,0), as well?

[4 marks]

4.

a) Explain what is meant by

(i) A gradient descent procedure.

[5 marks]

(ii) A local minimum.

[5 marks]

b) Explain why local minima of the energy function are desirable if a Hopfield network is to be used as a content-addressable memory, but undesirable if it is to be used to solve optimisation problems.

[5 marks]

c) How does the addition of 'noise' help a network escape from local minima?

[5 marks]

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

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

For a 2-node net, the probability of the net changing state from (x_1, x_2) at time t to the configuration in which neuron i has changed its state from x_i to $1 - x_i$ at time t+1 is given by

$$\frac{1}{2}[\, x_i(t)(1 - p_i(t+1)) + (1 - x_i(t))p_i(t+1) \,]$$

Consider the 2-node stochastic Hopfield net with weight

$$w = w_{12} = w_{21} = -1$$

and thresholds

$$s_1 = s_2 = 0$$

Draw a state transition diagram for this network at temperature T=1, labelling all transitions with their probabilities and showing the energy levels of the system.

[13 marks]

5.

a) Describe the differences between supervised, reinforced and unsupervised learning algorithms. How would you choose which kind of training was appropriate for a given problem? (Hint: what kind of information does each type of training algorithm need?) Give *one* example of a problem which is best suited to each of the types of training procedure, and justify your choice.

[15 marks]

b) The output of an associative reward-penalty (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 $\lambda = 0$?

[6 marks]

END OF PAPER