Answer THREE questions.

1.

a) Explain the difference between a local and distributed representation of information, and the way in which the latter of these is implemented within a neural network. What are the advantages and disadvantages of each storage method?

[8 marks]

b) Discuss the role of the binary decision neuron (BDN) and single-layer perceptron network (SLP) in the early history of neural computing. What problems could arise when training SLPs? How were they handled by workers of that time? What was the underlying reason for these difficulties, and how were they later fully resolved?

[9 marks]

c) Consider the 2-dimensional pattern set

$$\underline{x}_1 = (2, 2),\ \underline{x}_2 = (2, 4),\ \underline{x}_3 = (4, 1)$$

$$\underline{x}_4 = (4, 7),\ \underline{x}_5 = (8, 6),\ \underline{x}_6 = (7, 4)$$

Patterns $\underline{x}_1$ to $\underline{x}_3$ are to be classified as Class A, patterns $\underline{x}_4$ to $\underline{x}_6$ 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.

[6 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

$$\underline{x}_7 = (4, 3),\ \underline{x}_8 = (6, 5)\ \ \underline{x}_9 = (6, 7)$$

To which classes would your network, with the weights and threshold you have chosen, assign these patterns? If necessary, could you modify your choice of parameters so as to assign $\underline{x}_7$ and $\underline{a}_8$ to Class A, $\underline{x}_9$ to Class B?

[6 marks]

TURN OVER

2.

a) Write down the expression that typically describes a neuron's activation as a function of its inputs.

[3 marks]

b) Why is the activation value usually then passed through a firing function?

[8 marks]

c) Consider the sigmoid firing function. Give the mathematical form of this function when applied to an activation value *a*. Discuss the role of the sigmoid function in at least two different kinds of neural networks. Pay particular attention to the interpretation of the numerical value produced by the firing function, and the use which may be made of the gain parameter β.

[8 marks]

d) 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

[5 marks]

e) In a multilayer perceptron network to be trained using error backpropagation it is usual for the weights to initially be set to small random values. Suppose that instead, all the weights (including bias weights) in a 2-layer net with external inputs $x_1,..,x_n$ were to be set to *exactly zero*. By referring to the weight-change rules of d), explain why this would not in general produce a desirable result. What kinds of problems could this network still successfully solve?

[9 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} which will achieve this.

[4 marks]

(ii) Show, by splitting the activation of unit i in a pattern s into 'signal' and 'noise' terms, that the choice of weights prescribed by (i) can be expected to result in stable pattern storage (for suitably small P).

[8 marks]

b) It is desired to store the 3-bit binary patterns (011) and (111) in a 3-node Hopfield net.

(i) Using the simple pattern storage prescription of a)(i), show that in this case the Hopfield energy function is given by

$$H(x_1, x_2, x_3) = -2x_2x_3$$

Draw a state transition diagram, labelling all transitions with their probabilities (assuming one node is updated randomly at each time step) and showing the energy levels of the system.

[9 marks]

(ii) Write down a prescription for choosing thresholds that will break the symmetry of the storage algorithm of a)(i) with respect to complement states ($x_i \rightarrow 1-x_i$).

[3 marks]

(iii) Show how the energy function and state transition diagram of b)(i) are modified in this case.

[9 marks]

4.

a) Write down the error function appropriate to Boltzmann Machine training. Explain the meaning of all the quantities involved in the expression.

[5 marks]

b) A Boltzmann Machine, with three 'visible' units numbered i, j, k, is trained to perform the XOR function. During training neurons i and j are designated 'inputs', neuron k the 'output'.

(i) Assuming that all the weights start with values close to zero, what would be the initial probability of seeing the pattern (0, 0, 0) on the visible units when the network is in *free-running mode*? What would this probability become after successful training?

[4 marks]

(ii) After training the network is mistakenly used with i and k as the input neurons. What difference would this make?

[3 marks]

c) What are the strengths and weaknesses of the Boltzmann Machine in comparison with a multilayer perceptron trained using error backpropagation, and what are the reasons for these?

[6 marks]

d) The A_{RP} neuron is a 2-action ($y = 0$ or 1) stochastic learning automaton.

(i) Write down an update rule for the parameters w_{ij} in the presence of an externally generated reinforcement signal r. In what way does the 'penalty parameter' λ in this expression affect the behaviour of the neural automaton?

[7 marks]

(ii) Show by considering binary reinforcement signals $r \in \{0, 1\}$ as well as binary actions that for any action/reinforcement combination the above rule implements an effective strategy for dealing with an initially unknown and unpredictable environment.

[8 marks]

CONTINUED

5.

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

(i) self-organisation
(ii) topographic map
(iii) dimensional reduction
(iv) winning node (in a Kohonen map)

[12 marks]

b) Write down the Kohonen update rule for the change in weights to a winning node and its neighbours in an m×m output grid. Define the neighbourhood size d(t) on such a grid. Explain how (and why) the neighbourhood size and training rate must be decremented in the course of training.

[8 marks]

c) Assuuming that the weight vectors $\underline{w}_i$ and input vectors $\underline{x}$ in a Kohonen net are *normalised*

$$\sum_j w_{ij}^2 = 1, \quad \sum_j x_j^2 = 1$$

explain how a process of local excitation and lateral inhibition could be used to pick out a winning node and its neighbours. How could neighbourhood size be modified in this more biologically plausible version of the Kohonen algorithm?

[7 marks]

d) Suggest some areas of application for a Kohonen network. Explain whay in these areas an unsupervised net would be a more appropriate choice than a supervised or reinforcement-trained model.

[6 marks]

END OF PAPER