UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 1999

BEng Honours Degree in Computing Part III
BEng Honours Degree in Information Systems Engineering Part III
MEng Honours Degree in Information Systems Engineering Part III
BSc Honours Degree in Mathematics and Computer Science Part III
MSci Honours Degree in Mathematics and Computer Science Part III
MSc Degree in Computing Science
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the
Diploma of Membership of Imperial College
Associateship of the City and Guilds of London Institute
Associateship of the Royal College of Science

PAPER 3.89 / I 3.24

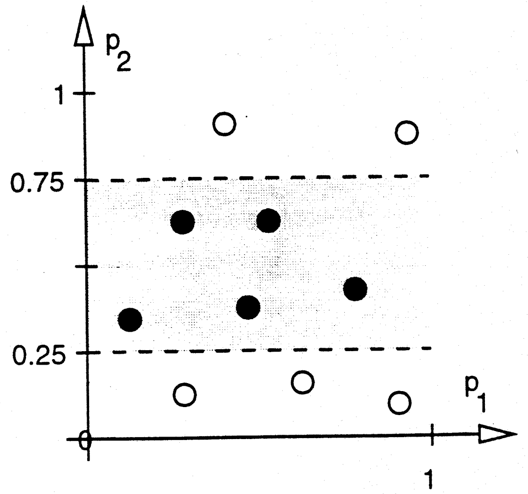
NEURAL NETWORKS

Thursday, April 29th 1999, 2.00 – 4.00

Answer THREE questions

For admin. only:
paper contains 4 questions

1a The following figure represents a classification problem in the unit square $[0, 1]^2$.



Let

$$f(p) = \Theta(\sum_{i=1}^{n} p_i w_i + w_0) \tag{1}$$

be the function that threshold perceptrons with $n+1$ weights w_0, ..., w_n compute from n inputs p_1, ..., p_n. Θ denotes the step function.

Why can no single threshold perceptron with two inputs, p_1 and p_2, solve this problem? Support your reasoning by a *geometric* argument derived from (1).

b Solve this classification problem by using two threshold perceptrons. Both perceptrons have the same 2-dimensional input vector from the unit square. Compute weights for the two perceptrons such that

- both output one if a point lies in the gray area and
- at least one of the perceptrons outputs zero otherwise.

c Construct a network by adding a third threshold perceptron such that the network outputs one if and only if its input is from the gray area. Draw a figure and label all weights.

d Assume the input point $(p_1, p_2) \in [0, 1]^2$ is described with the 3-dimensional vector

$$(p_1, p_2, (p_2)^2) \in [0, 1]^3.$$

How could a *single* threshold perceptron with these *three* inputs be used to solve the classification problem? Draw a figure.

The four parts carry, respectively, 25%, 30%, 20%, 25% of the marks.

2a Define Artificial Neural Networks in terms of nodes a, edges (a, b), local connections, activation functions f_a, node outputs o_a, weights w_{ab} and network input p. Draw a figure of a node and its related local quantities.

b Specialise to the case of backpropagation feedforward networks. Draw a figure of a feedforward network displaying the role of its layers, input patterns p and network output $\text{out}_w(p)$. Explain the concept of "learning by examples" using gradient descent.

c Formulate an algorithm for computing the gradient of the neural network error with respect to the weight vector w given a dataset of pattern target pairs (p^i, t^i). For simplicity assume that
- input and output nodes have a linear activation (the identity)
- hidden nodes use the tanh activation function
- the following cost function is used: $(y, t) \mapsto 3(y - t)^2$

Reduce the gradient computation to simple operations such as multiplication, summation and tanh computation. Do not prove the algorithm correct.

The three parts carry, respectively, 30%, 30%, 40% of the marks.

3a Let N be a deterministic attractor network with symmetric weights ($w_{ab} = w_{ba}$) without biases ($w_{0a} = 0$) driven by Glauber dynamics. Show that if $S^* \in \{-1, 1\}^N$ is a point of attraction then $-S^*$ is also a point of attraction.

b Specialise the network of part (a) to one with three neurons without self-interaction ($N = \{1, 2, 3\}$, $w_{aa} = 0$).

i) Draw the network graph with all relevant weights. The aim is to store one pattern $\xi := (-1, 1, 1)$. Set the weights to achieve this aim according to Hebb's rule, $w_{ab} = \sum_i \xi_a^i \xi_b^i / |N|$, and label the graph accordingly. Prove that ξ is a stable point of attraction in this network.

ii) Draw a picture of the state space S using the axes S_1, S_2 and S_3. Mark all discrete states with their energy (in this special case, the energy is $E = -\sum_{a<b} w_{ab} S_a S_b$). For which state(s) is the energy minimal?

iii) Show that the introduction of bias weights $w_{0a} := \xi_a/3$ helps to break the symmetry $S \mapsto -S$ and, hence, enables the energy of ξ to drop below the energy of $-\xi$. Show that now ξ is the only state attaining the absolute energy minimum of all states. Plot the states as nodes in an energy graph, ie, states with the same energy appear on the same horizontal line. Mark all possible transitions (according to the Glauber dynamics) with an arrow.

The two parts carry, respectively, 15%, 85% of the marks.

4a Explain the concept of unsupervised learning. Give three examples of what can be achieved with unsupervised learning in general.

b Describe the basic learning algorithm for self-organising feature maps. How is a trained network applied to new data?

c For each of the main objectives of self-organising maps (ie, clustering, dimensionality reduction and neighbourhood preservation) identify the key-parts of the learning algorithm and architecture that support it.

The three parts carry, respectively, 25%, 60%, 15% of the marks.

[End of Paper