UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2000

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

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

PAPER C389=I3.24

NEURAL NETWORKS

Wednesday 10 May 2000, 14:00
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions

1 Assume you have a set of 1296 pattern-target examples from a manufacturing process. For each product i a label $t^i \in \{0,1\}$ is assigned to a feature vector $p^i \in \mathbb{R}^4$ of this product. $t^i = 1$ stands for "the product i passes an expensive quality test" and $t^i = 0$ stands for "the product i fails the test". You may think of the products being light bulbs of one kind, the quality test being whether a light bulb burns for more than 1000 hrs and the feature data being production line data and measurements.

a Assume further that the underlying classification is linearly separable. Draw a diagram with the simplest neural-network solution that is capable of learning the problem. Write down the network function. Explain the concept of learning by example.

b How can the trained neural network be used to predict from unseen production data p^u whether the product is likely to pass the test? Where is the knowledge stored? How would you define and estimate the generalisation error in this case?

c Now assume that the underlying problem is *not* linearly separable. Which artificial neural network (architecture, activation functions, cost function) would you suggest to learn to predict the class label t^i? Which learning algorithm would you suggest? Name a method which finds the proper dimensions of the neural network.

The three parts carry, respectively, 40%, 20%, 40% of the marks.

2 A helicopter drops sensors $i \in \{1,2,\ldots,400\}$ onto a desert. Ideally, the sensors would fall onto perfect grid positions {(1,1), (1,2), ..., (1,20), (2,1), (2,2), ..., (2,20), ..., (20,20)}. However, the sensors are slightly off target and fall into known nearby positions (x_i, y_i). The sensors measure and record data $\phi(x_i, y_i) \in \mathbb{R}^n$. The data are supposed to change continuously with the location.

a How can you use a neural network to estimate the data vector ϕ at the perfect grid positions? Detail your approach (architecture, activation functions, cost function, learning) and justify the complete decision with an argument from learning theory. Draw a picture for $n = 3$.

b In another case most of the sensors were lost in the sand and their position (x_i, y_i) could not be recorded. Nevertheless data were received from these sensors (including their identification i) and the data are sufficiently distinctive from each other. Explain briefly how you could use a self-organising network to roughly estimate a plausible grid area where the data came from?

The two parts carry, respectively, 75%, 25% of the marks.

3 A neural network shall be trained to complete a time series of length 5 (y^1, y^2, y^3, y^4, y^5) given the initial value y^1 and given another time series of length 5 (p^1, p^2, p^3, p^4, p^5). The network architecture is given as a general recurrent network with two nodes $\{a, b\}$ and synchronous node updates.

a Show how weight sharing can help to transform the recurrent network into an equivalent feedforward network that can be trained with supervised learning: draw a figure with the feedforward network detailing how a time series is input to the network and how the output time series is obtained when the network is running. State the necessary constraints on the weights in the feedforward network. Assume that gradient descent is used to train the feedforward network. How are the weight constraints fulfilled during learning?

b Use the multi-dimensional chain rule to devise a general mechanism for ensuring that more general weight constraints are fulfilled during gradient descent training. Draw a diagram for your process.

c Compute the update rule Δw for a three-dimensional weight space in the case of gradient descent learning with the constraints $w_3 = w_2 + w_1$ and $w_2 = {w_1}^2$. You may assume that the gradient $\nabla_w \mathrm{Err(w)}$ of the network error function is known, eg, through the backpropagation algorithm.

The three parts carry, respectively, 40%, 40%, 20% of the marks.

4a What is a point of attraction of a Hopfield network? How do points of attraction relate to the auto-associative task of pattern completion? Define an energy function for a deterministic attractor network with Glauber dynamics and symmetric weights without self-interaction. Show that this energy function is minimised in a point of attraction.

b Show that the asynchronous update rule for Boltzmann machines corresponds to the update rule of a Hopfield network in the case of very small temperatures, ie, Hopfield networks may be considered as special cases of Boltzmann machines.

c A Boltzmann machine without hidden nodes and without input nodes is to be used for storing pictures ξ^1, ξ^2, ... Which term of the (supervised) Boltzmann learning rule can be identified as the Hebb rule $w_{ab} = \sum_i \xi^i_a \xi^i_b / |N|$ of the (unsupervised) Hopfield model? How could the other term be implemented into Hopfield learning in order to improve its performance?

The three parts carry, respectively, 55%, 20%, 25% of the marks.