UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

B.Sc. Examination 2003

COMPUTING AND INFORMATION SYSTEMS

IS53002A (CIS311) Neural Networks

Duration: 2 hours 15 minutes

Date and time:

- *Full marks will be awarded for complete answers to FOUR questions. Do not attempt more than FOUR questions on this paper.*
- *Electronic calculators may be used. The make and model should be specified on the script. The calculator must not be programmed prior to the examination. Calculators which display graphics, text or algebraic equations are not allowed.*

THIS EXAMINATION PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM

Question 1.

a) Describe how a batch training algorithm for neural networks operates. What is the alternative kind of a training algorithm for neural networks and how does this alternative training mode suggest processing the examples? **[6]**

b) What is the main limitation of the classification performance of a Perceptron network using a linear activation function? What should be changed in order to overcome this limitation of the Perceptron? **[6]**

c) Which of the following boolean functions can be realised by a single layer Perceptron: AND, NOT, XOR, OR ? Explain your answer. **[4]**

d) Implement the AND function with a Perceptron network. Assume that the network has three inputs and a bias input, and demonstrate which combination of weights accurately classifies all eight examples of the AND function (without training the network). (use the threshold function: $f(s) = 0$ if $s \leq 0$ and $f(s) = 1$ if $s > 0$). **[9]**

Question 2.

a) Define the incremental gradient descent training rule for single-layer Perceptron networks. Explain the meaning of every term in the formulae. **[5]**

b) Consider a single layer Perceptron network with seven inputs and a threshold x_0. This network may learn to recognize the digits from 0,9,8,7,6,5,4,3,2,1 represented by seven segments each associated with an input. When the network is trained to learn a digit the output should be greater than one only for this digit *Threshold*(s) = 1 if $s > 1$. Assume that the learning rate is one.

Training can be performed using the following example vectors, provided sequentially:

x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
1	0	1	1	1	1	1	1	0
1	1	1	1	1	1	1	0	9
1	1	1	1	1	1	1	1	8
1	0	0	1	1	1	0	0	7
1	1	1	1	0	1	1	1	6
1	1	1	1	0	1	1	0	5
1	1	1	0	1	1	0	0	4
1	1	0	1	1	1	1	0	3
1	1	0	1	1	0	1	1	2
1	0	0	0	1	1	0	0	1

How many weight updates are necessary to learn the digit 0 starting with weights that are all zero: (w_0, w_1, w_2, w_3, w_4, w_5, w_6, w_7)=(0,0,0,0,0,0,0,0)? Demonstrate the training step by step: what is the network output after each example and what should happen to the weights after this example? **[20]**

Question 3.

a) What function does the Probabilistic Neural Network learn to approximate? **[2]**

b) Which activation function is used in the pattern layer nodes of the probabilistic network? **[4]**

c) Consider a probabilistic neural network with four inputs (x_1, x_2, x_3, x_4), that serves for classification into negative 0 and positive 1 using parameter $\sigma=0.5$. Demonstrate how this probabilistic neural network will classify the following example: [`0.32 0.44 0.11 0.66`] assuming that there are eight training examples available:

```
[0.8 0.71 0.21 0.8] negative
[0.9 0.85 0.22 0.7] negative
[0.7 0.77 0.23 0.8] negative
[0.6 0.75 0.24 0.7] negative
[0.2 0.55 0.41 0.8] positive
[0.1 0.5  0.42 0.7] positive
[0.3 0.55 0.43 0.8] positive
[0.4 0.6  0.44 0.7] positive
```

Show the computation of each output and take a classification decision without using and computing the exponent function (because: exp(z) > exp(y) when $z > y$). **[19]**

Question 4.

a) How many backpropagations does the gradient-descent training algorithm for feedforward multilayer perceptron networks require? **[5]**

b) Can it happen that the backpropagation algorithm for training multilayer neural networks becomes unstable or gets stuck at suboptimal solutions? Explain which network solution is considered suboptimal? **[5]**

c) Can we implement the the backpropagation algorithm so as to perform network training in stages by presenting consecutively different subsets of the training examples? **[2]**

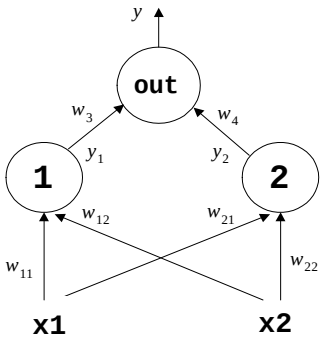
d) Can we train the multilayer perceptron network to recognise multiple classes simultaneously? Explain how this can be achieved? **[5]**

e) How many weights in the multilayer perceptron network: more or less can lead to overfitting? **[3]**

f) Is network overfitting advantageous or not when addressing real world problems? Motivate your answer. **[5]**

Question 5.

Let a two layer Perceptron network with two hidden nodes and one output node using the sigmoidal activation function be given. Assume that the input variables are x_1 and x_2, and the weights on their connections feeding the hidden nodes are w_{11}, w_{21}, and w_{12}, w_{22}. The hidden to output connections are weighted by w_3, w_4, and these links feed signals y_1, y_2, as illustrated in the plot below. Assume that there are no bias connections, that is the weights w_{01}, w_{02}, w_{0out} are zero.



Suppose that you have to train this multilayer network using the backpropagation algorithm.

a) Explain how the forward propagation pass computes y_1 and y_2, and what will be the network output y in terms of these variables (without evaluating the exponent function)? **[6]**

b) Give the formula for performing the backward propagation pass and explain how the hidden to output weight updates w'_3 and w'_4 could be obtained. **[8]**

c) How do we compute during the backward pass the errors and the weight updates w'_{11}, w'_{21} , w'_{12} and w'_{22} on connections from the input to the hidden nodes? Explain the meanings of all terms in the equation. **[11]**

Question 6.

a) What are the similarities between Radial Basis Function networks and Multilayer Perceptrons? **[4]**

b) Which are the five essential differences between Radial Basis Function networks and Multilayer Perceptrons? **[10]**

c) Explain how many layers does a generalised radial basis function network have and what is the purpose of each particular layer? **[6]**

d) What is the rationale for using Radial Basis Function n to perform classification and regression tasks? **[5]**