

<b>FAMILY NAME</b>	
<b>OTHER NAMES</b>	
<b>STUDENT ID</b>	
<b>SIGNATURE</b>	

**Sample Exam**  
**COMP 9444 NEURAL NETWORKS**  
**COMP 9844 EXTENDED NEURAL NETWORKS**  
**Solutions**

- (1) TIME ALLOWED – 3 HOURS
- (2) TOTAL NUMBER OF QUESTIONS – **13**
- (3) **COMP9444** CANDIDATES SHOULD ATTEMPT QUESTIONS **1-12**  
**COMP9844** CANDIDATES SHOULD ATTEMPT **ALL** QUESTIONS
- (4) QUESTIONS ARE NOT OF EQUAL VALUE
- (5) THIS PAPER MAY NOT BE RETAINED BY THE CANDIDATE
- (6) QUESTIONS CARRY THE NUMBER OF MARKS INDICATED. THE TOTAL NUMBER OF MARKS FOR COMP9444 IS 180, AND FOR COMP9844 STUDENTS IS 195.

**PLEASE ANSWER PART A IN YOUR EXAMINATION BOOKLET**  
**PLEASE ANSWER PART B IN THE SPACES ON THIS QUESTION PAPER**

ANSWERS MUST BE WRITTEN CLEARLY IN INK. EXCEPT WHERE THEY ARE EXPRESSLY REQUIRED, PENCILS MAY ONLY BE USED FOR DRAWING, SKETCHING OR GRAPHICAL WORK.

**PART A - Answer in your exam booklet.**

**(Ask Achim)**

PART B - Answer in the spaces on this question paper.

**Question 8 – Temporal Processing (18 marks)**

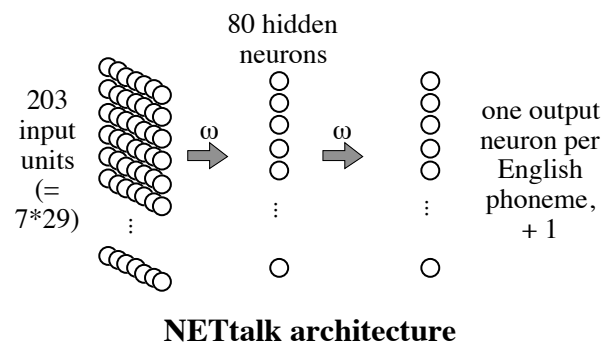
(a) With a supervised learning algorithm, we can specify target output values, but we may never get close to those targets at the end of learning. Give two reasons why this might happen.

**Answer:**

- (i) data may be valid, and inconsistency results from a stochastic aspect of the task (or some aspect of the task is not modelled by the input data collected);
- (ii) the data may contain errors - e.g. measurement errors or typographical errors

8(b) Describe the *architecture* and the *computational task* of the NetTalk neural network.

**Answer:** *architecture:*



Each group of 29 input units represents a letter, so inputs together represent seven letters. Right-hand caption should say “one output per English phoneme +2” (not + 1).

The *computational task* is to output the representation of the phoneme corresponding to the middle letter of the seven

8(c) Why does a time-delay neural network (TDNN) have the same set of incoming weights for each column of hidden units?

**Answer:**

To provide temporal translation invariance. *or*

So that the TDNN will be able to identify the input sound, no matter which frame the input sound begins in.

8(d) Distinguish between a *feedforward network* and a *recurrent network*.

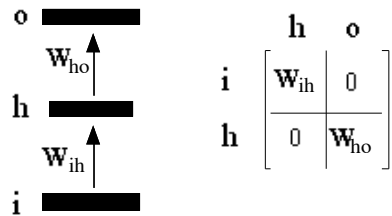
**Answer:**

A feedforward network has no cyclic activation flows.

**Question 9 – Elman and Jordan nets, etc. (18 marks)**

- (a) Draw the weight matrix for a feedforward network, showing the partitioning. You can assume that the weight matrix for connections from the input layer to the hidden layer is  $\mathbf{W}_{ih}$ , and that the weight matrix for connections from the hidden layer to the output layer is  $\mathbf{W}_{ho}$ .

**Answer:**

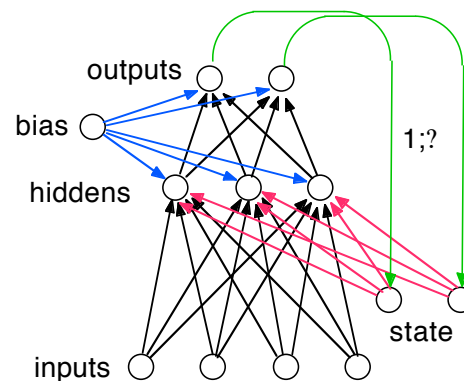


The left-hand part of the diagram isn't necessary for the answer to this question.

- 9(b) In a Jordan network with  $i$  input neurons,  $h$  hidden layer neurons, and  $o$  output neurons: (a) how many neurons will there be in the state vector, and (b) if  $i = 4$ ,  $h = 3$ , and  $o = 2$ , draw a diagram showing the connectivity of the network. Do not forget the bias unit.

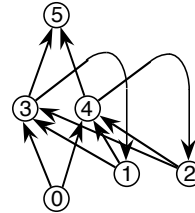
**Answer:** (a)  $o$  neurons in state vector (same as output vector – that's letter  $o$ , not zero)

(b)



9(c) Draw a diagram illustrating the architecture of Elman's simple recurrent network that performs a temporal version of the XOR task. How are the two inputs to XOR provided to this network?

**Answer:**



The inputs are passed sequentially to the single input unit (0) of the temporal XOR net.

9(d) Briefly describe the use of cluster analysis in Elman's lexical class discovery experiments, and one of his conclusions from this.

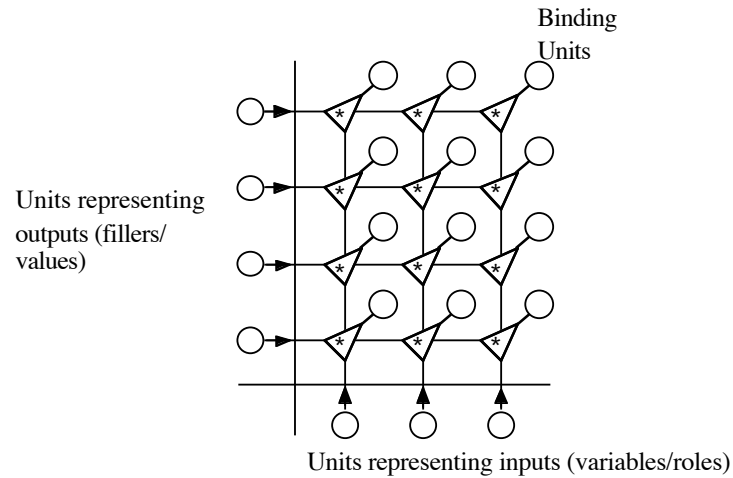
**Answer:**

Elman clustered hidden unit activation patterns corresponding to different input vectors and different sequences of input units. He found that the clusters corresponded well to the grammatical contexts in which the inputs (or input sequences) occurred, and thus concluded that the network had in effect learned the grammar.

**Question 10 – Tensor Product Networks (18 marks)**

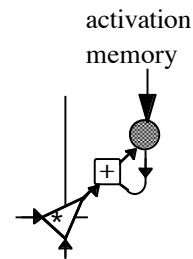
- (a) Draw an architectural diagram of a rank 2 tensor product network where the dimensions of the input/output vectors are 3 and 4. You do not need to show the detailed internal structure of the binding units.

**Answer:**



- 10(b) Draw a diagram of a *single binding unit* in a rank 2 tensor product network illustrating the internal operation of the binding unit in *teaching* mode.

**Answer:**



10(c) Define the concepts of *dense* and *sparse random representations*. How do their properties compare with those of an orthonormal set of representation vectors.

**Answer:**

In a dense random representation, each vector component is chosen at random from a uniform distribution over say  $[-1, +1]$ . In a sparse random representation, the non-zero components are chosen in this way, but most components are chosen (at random) to be zero. In both cases, the vectors are normalised so that they have length 1.

Members of orthonormal sets of vectors have length one, and are orthogonal to one another. Vectors in dense and sparse random representations are “orthogonal on average” – their inner products have a mean of zero.

10(d) What is a Hadamard matrix? Describe how a Hadamard matrix can be used to produce suitable distributed concept representation vectors for a tensor product network. What are the properties of the Hadamard matrix that makes the associated vectors suitable?

**Answer:**

A Hadamard matrix  $H$  is a square matrix of side  $n$ , all of whose entries are  $\pm 1$ , which satisfies  $HH^T = nI_n$  ...where  $I_n$  is the identity matrix of side  $n$ .

The rows of a Hadamard matrix, once normalised, can be used as distributed representation vectors in a tensor product network.

This is because the rows are orthogonal to each other, and have no zero-components.

### Question 11 - Self-Organising Systems (18 marks)

(a) In a 2-D self-organising map with input vectors of dimension  $m$ , and  $k$  neurons in the map, how many weights will there be?

**Answer:**

$$mk$$

11(b) Describe the *competitive* process of the Self-Organising Map algorithm.

**Answer:**

Let  $m$  denote the dimension of the input pattern

$$\mathbf{x} = [x_1, x_2, \dots, x_m]^T$$

The weight vector for each of the neurons in the SOM also has dimension  $m$ . So for neuron  $j$ , the weight vector will be:

$$\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jm}]^T$$

For an input pattern  $\mathbf{x}$ , compute the norm difference  $\|\mathbf{w}_j - \mathbf{x}\|$  for each map neuron  $j$ , and choose the smallest value: this is the winning neuron. Let  $i(\mathbf{x})$  denote the index of the winning neuron (and also the output of a trained SOM).

11(c) Briefly explain the concept of a Voronoi cell.

**Answer:**

Given a set of vectors  $X$ , the Voronoi cells about those vectors are the ones that partition the space they lie in, according to the nearest-neighbour rule. That is, the Voronoi cell that a vector lies in is that belonging to the  $\mathbf{x} \in X$  to which it is closest.

11(d) Briefly explain the term *code book* in the context of vector quantisation.

**Answer:**

When compressing data by representing vectors by the labels of a relatively small set of *reconstruction vectors*, the set of reconstruction vectors is called the *code book*.

**Question 12 – Hopfield Nets, BSB (18 marks)**

(a) Write down the *energy function* of a discrete Hopfield net.

**Answer:**

$$E = -0.5 \sum_{i=1}^N \sum_{j=1}^N w_{ji} x_j x_i$$

12(b) Compute the weight matrix for a 4-neuron Hopfield net with the single fundamental memory  $\xi_1 = [1, -1, -1, 1]$  stored in it.

**Answer:**

$$\mathbf{W} = \frac{1}{4} (\xi_1^T \xi_1 - \mathbf{I}_4) = \frac{1}{4} \begin{bmatrix} 0 & -1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & -1 \\ 1 & -1 & -1 & 0 \end{bmatrix}$$

12(c) Briefly describe two types of *attractor* in a dynamical system.

**Answer:**

An attractor is a bounded subset of space to which non-trivial regions of initial conditions converge at time passes. *Pick two of ...*

- point attractor: system converges to a single point
- limit cycle: system converges to a cyclic path
- chaotic attractor: stays within a bounded region of space, but no predictable cyclic path

12(d) Write down the energy function of a BSB network with weight matrix  $\mathbf{W}$ , feedback constant  $\beta$ , and activation vector  $\mathbf{x}$ .

**Answer:**

$$E = -\left(\frac{\beta}{2}\right) \sum_{ij} w_{ij} x_i x_j$$

or

$$E = -\left(\frac{\beta}{2}\right) \mathbf{x}^T \mathbf{W} \mathbf{x}$$

**Question 13 –COMP9844 Students Only** (*15 marks*)

Solutions for these are found in a separate document.