TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON

Student ID: .....

EXAMINATIONS — 2008

MID-YEAR

COMP 307

INTRODUCTION TO ARTIFICIAL INTELLIGENCE

Time Allowed: 3 Hours

Instructions: There are a total of 180 marks on this exam.
Attempt all questions.
Calculators may be used.
Non-electronic foreign language translation dictionaries may be used.
The appendices can be removed for reference.

Questions

1. Prolog [25]
2. Rule Based Systems [30]
3. Search [30]
4. Machine Learning Basics [35]
5. Neural Networks and Evolutionary Computing [34]
6. Clustering and Filtering [26]

Question 1. Prolog [25 marks]

(a) [8 marks] Define the following function in Prolog.

$$f(x) = \begin{cases} 2 * f(x-1) + 1 & \text{if } x > 0 \\ 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Your program should be able to calculate the $f(x)$ value for any number x.

(b) [8 marks] Write a program for `union(L1, L2, L)` such that `L` is the union of `L1` and `L2`. For example,

`union([a,b,c],[x,y,b], [a, c, x, y, b])` is true, whereas

`union([a,b,c],[x,y,b], [a, b, c, x, y, b])` is false, and

`union([a, b, c], [x, y, b], L)` is true with `L=[ a, c, x, y, b]`.

You may assume there are no duplicates in `L1` and `L2`.

(c) [9 marks] Write a program for `tree_member(X,L)` where `X` occurs in the list structure `L` at any level of nesting. For example,

`tree_member(a,[x,[[a],b],y])` is true, and

`tree_member([[a],b],[x,[[a],b],y])` is true, whereas

`tree_member([a,b],[x,[[a],b],y])` is false.

Question 2. Rule based systems [30 marks]

(a) [10 marks] Develop a simple set of rules for diagnosing diseases given patient symptoms, using the following knowledge of typical symptoms.

- Influenza: Symptoms include a persistent dry cough and a feeling of general malaise.
- Hey-fever: Symptoms include a runny nose and sneezing. The patient will show a positive reaction to allergens, such as dust or pollen.
- Laryngitis: Symptoms include a fever, a dry cough, and a feeling of general malaise. A 'laryngoscopy' will reveal that the person has an inflamed larynx.
- Asthma: Symptoms include breathlessness and wheezing. If it is triggered by an allergen, such as dust or pollen, it is likely to be "extrinsic asthma". "Intrinsic asthma" tends to be triggered by exercise, smoke or a respiratory infection.

(b) [8 marks] Describe how a simple backward chaining interpreter could be used to go through the possible diagnosis, asking the user questions about their symptoms.

(c) [8 marks] What do you think are the main problems and limitations of the rule-set developed?

(d) [4 marks] What are the main differences between a backward chainer and a forward chainer?

Question 3. Search

[30 marks]

(a) [10 marks] Briefly describe each of the following search algorithms:

(i) Iterative deepening

(ii) Hill climbing

(Question 3 continued on next page)

(Question 3 continued)

(b) [10 marks] Considering all the search algorithms introduced in the lectures, including *Breadth first, Uniform cost, Depth first, Depth limited, Iterative deepening, Bidirectional, Greedy search, A* search, Hill climbing, Beam search, Branch and bound*, identify the algorithms that are suitable for finding the best solution and the algorithms that are suitable for finding one solution quickly. Justify your answer.

(Question 3 continued on next page)

(Question 3 continued)

(c) [10 marks] Route Finding

There are four cities, A, B, C, D, which are all connected to each other. The distances between pairs of cities are as follows:

	A	B	C	D
A		10	15	20
B			12	9
C				5
D				

(i) What is the shortest route between A and D, **and** what search algorithm could you use to avoid exploring every path?

(ii) What would be the maximum number of paths in the search space for six cities all connected to each other?

(iii) Considering the search problem with many cities all connected to each other, what is a good search algorithm and how could you implement it? For example, if you choose to use heuristic search, what is your heuristic function?

Question 4. Machine Learning Basics [35 marks]

(a) [4 marks] Briefly describe the main difference between *classification* and *clustering* in terms of learning schemes, and data sets.

(b) [3 marks] Due to its simplicity, the *nearest neighbour method* is often used as a classifier. Briefly describe this method.

(c) [3 marks] The decision tree learning algorithm uses an impurity measure to choose between attributes. Explain why the $p(A)p(B)$ formula is a reasonable impurity measure for a set of instances belonging to two different classes A and B, but is not good for three or more possible classes that the decision tree must distinguish.

(Question 4 continued on next page)

(Question 4 continued)

(d) [9 marks] Suppose you are building a Naïve Bayes spam filter to distinguish *spam* messages from real email messages (*non-spam*). You have picked two key words: "winner" and "donation" to characterise each message, and have counted how many of the messages contain each word:

	spam		**non-spam**	
	word present	word not present	word present	word not present
"winner"	40	360	10	190
"donation"	5	395	40	160
Total count	400		200	

If your spam filter was presented with a new message that contained the word "winner" but did not contain "donation", would your spam filter classify the message as spam or as non-spam? Show your working.

(Question 4 continued on next page)

(e) [10 marks] Consider the following data set describing 10 loan applications at a bank, of which 5 were approved and 5 were rejected. They are described by three attributes: whether the applicants have a job or not, whether their deposits are low or high, and whether their credit records are very good, good or bad.

Instance	**Job**	**Deposit**	**Credit**	**Class(loan decision)**
1	true	low	very good	Approve
2	true	low	good	Approve
3	true	high	very good	Approve
4	true	high	good	Approve
5	true	high	good	Approve
6	false	low	good	Reject
7	false	low	bad	Reject
8	true	low	bad	Reject
9	false	low	very good	Reject
20	false	high	bad	Reject

The bank wants to build a decision tree to help making loan decisions. Which attribute should the bank choose for the *root* of the decision tree if they use the impurity function $p(\text{Approve})p(\text{Reject})$. Show your working.

(Question 4 continued on next page)

(Question 4 continued)

(f) [6 marks] John Smith used a perceptron (linear threshold unit) to solve a binary classification task with the following labelled instances:

Input Feature 1	Input Feature 2	Input Feature 3	Output Class
0	1	0	0
0	1	1	1
1	1	0	1
1	1	1	0

His perceptron used three input nodes and one output node, which included a bias weight. It was trained using the usual perceptron learning rule, but the weights did not converge no matter how he changed the learning parameters.

(i) Explain why John's perceptron was not successful.

(ii) Suggest an improvement that would enable the instances to be learned successfully.

Question 5. Neural Networks and Evolutionary Computing [34 marks]

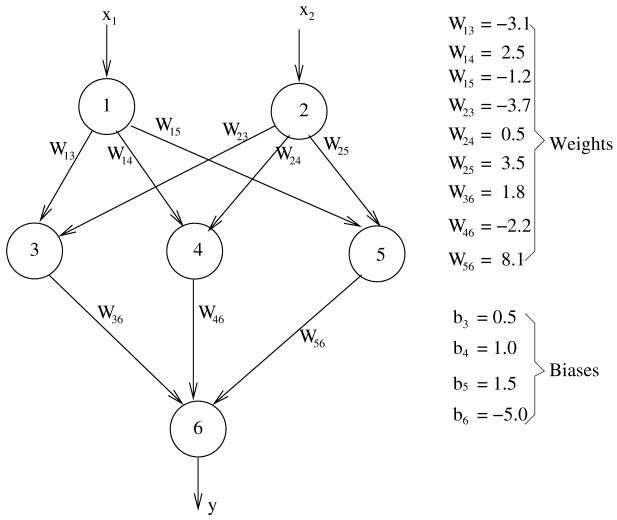
(a) [9 marks] Percy Smith has developed a classifier for distinguishing cancer cells from normal cells. The process involves the extraction of 5 features from images of cells and the use of a standard multilayer feed forward neural network, trained by back propagation, for classification. There are 500 examples in total, of which 100 are used for network training and 400 for testing. The network architecture he used is 5-25-1. After training for 10,000 epochs, the network classifier still performs badly on the test set.

Suggest **three** ways for improving the performance, and in each case explain briefly why it will help.

(Question 5 continued on next page)

(Question 5 continued)

(b) [10 marks] Consider the following feed forward network which applies inputs x_1 and x_2 to nodes 1 and 2 directly, and uses the sigmoid/logistic transfer function (see Appendix B) for the other nodes (3 to 6).



(i) What will be the output (y) of node 6 be, if the vector (0.0, 0.0) is presented to the input of this network?

(ii) What will be the new value of weight w_{56} after one epoch of training using the back propagation algorithm? Assume that the training set consists of only the single exemplar $x = (0.0, 0.0)$, $y = 0.0$, and that the learning rate η is 0.2.

Show your working.

(c) [4 marks] Briefly compare genetic algorithms (GAs) with neural networks (NNs) in terms of the representation of solutions and the search techniques.

(d) [4 marks] In the context of genetic algorithms and genetic programming, briefly explain why the *mutation* operator is usually needed (in addition to crossover) and why it is only set to a small rate (compared to the crossover operator).

(e) [3 marks] Briefly describe the term *sufficiency* in the context of creating a primitive set in genetic programming.

(Question 5 continued on next page)

(Question 5 continued)

(f) [4 marks]

Suppose your task is to use Genetic Programming to map a single input variable x to a single output variable y, based on the following data set of 20 points:

x	-2.0	-1.75	-1.50	-1.25	-1.00	-0.75	-0.50	-0.25	0.00	0.25
y	37.0000	24.1602	15.0625	8.9102	5.0000	2.7227	1.5625	1.0977	1.0000	1.0352

x	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75
y	1.0625	1.0352	1.0000	1.0977	1.5625	2.7227	5.0000	8.9102	15.0625	24.1602

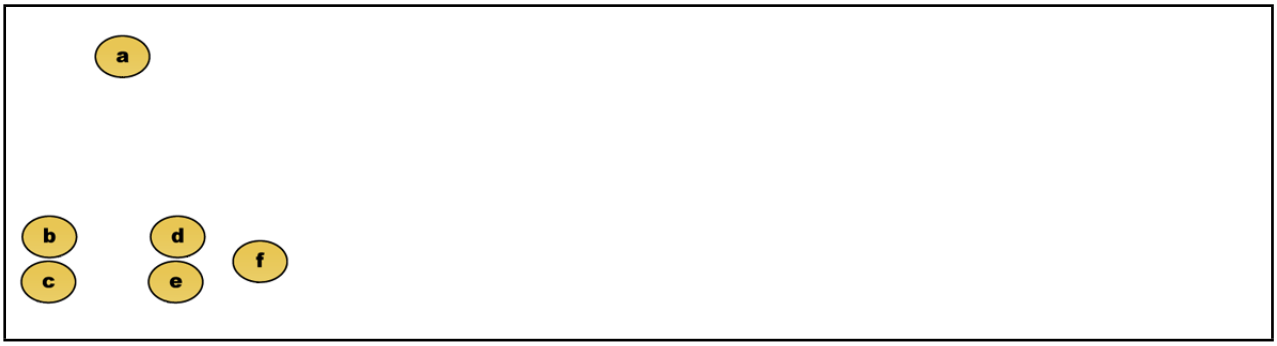
(i) Suggest a good terminal set to use.

(ii) Suggest a good fitness function to use.

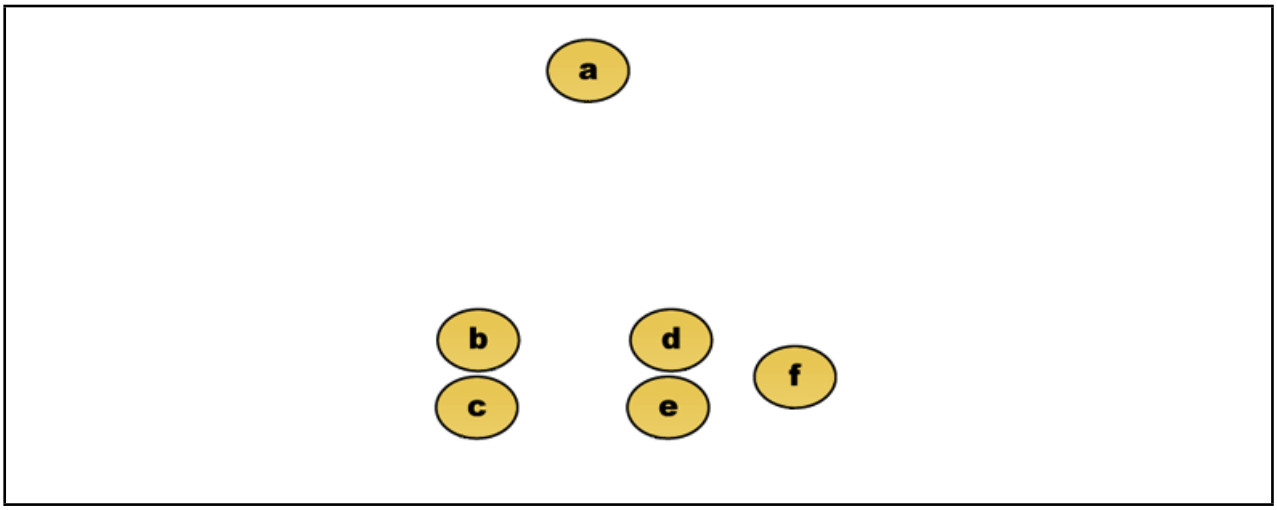
Question 6. Clustering and Filtering [26 marks]

(a) [2 marks]

Consider the data points labelled "a" to "f" on the left-hand side of the box below. On the right-hand side of the box show the hierarchical clustering (dendrogram) of the points that would result from a bottom-up clustering using the Euclidean distance measure.



(b) [3 marks] Consider using the K-means algorithm on the same data. By giving an example on the picture below, demonstrate that at least one initial choice of cluster centers can lead to a sub-optimal final clustering.

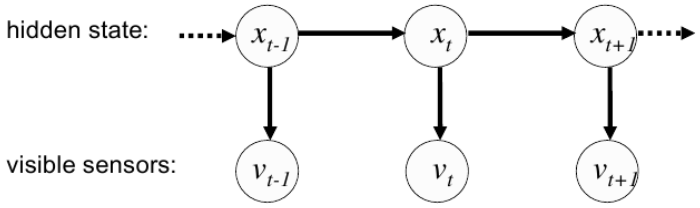


(c) [4 marks] How are the `E` and `M` steps of the `soft K-means` clustering algorithm related to the corresponding steps of conventional K-means?

(d) [4 marks] Suggest **two** approaches for coping with the existence of local optima (*i.e.* sub-optimal solutions that are stable) in the `soft K-means` algorithm.

(e) [5 marks] Discuss the role "Occam's razor" might play in complexity control for machine learning, such as in deciding what value of `K` to use in `soft K-means` clustering.

(f) [8 marks] The figure below shows a schematic model of sequential data. In this model there is a "hidden" state x that changes over time, and which results in "visible" sensor values v. For example, in tracking we represent uncertainty about the location of the target by a probability distribution $P(x_t)$ over possible states x at time t, and update this distribution to $P(x_t)$ based on both a transition and a sensor model.



A crucial requirement of such a system is that the representation of $P(x_{t+1})$ should be of the same *form* as its predecessor $P(x_t)$. Explain how this is achieved by the `particle filtering` algorithm.

****

Student ID: .....

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Student ID: .....

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Appendix for COMP307 exam

(You may tear off this page if you wish.)

A Some Formulae You Might Find Useful

$$p(C|D) = \frac{p(D|C)p(C)}{p(D)} \tag{1}$$

$$f(x_i) = \frac{1}{1+e^{-x_i}} \tag{2}$$

$$O_i = f(I_i) = f(\sum_k w_{k\to i} \cdot o_k + b_i) \tag{3}$$

$$\Delta w_{i\to j} = \eta o_i o_j (1-o_j)\beta_j \tag{4}$$

$$\beta_j = \sum_k w_{j\to k} o_k (1-o_k)\beta_k \tag{5}$$

$$\beta_z = d_z - o_z \tag{6}$$

B Sigmoid/Logistic Function