

EXAMINATIONS — 2010

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 appendix on the last sheet can be removed for reference for questions 4-6.

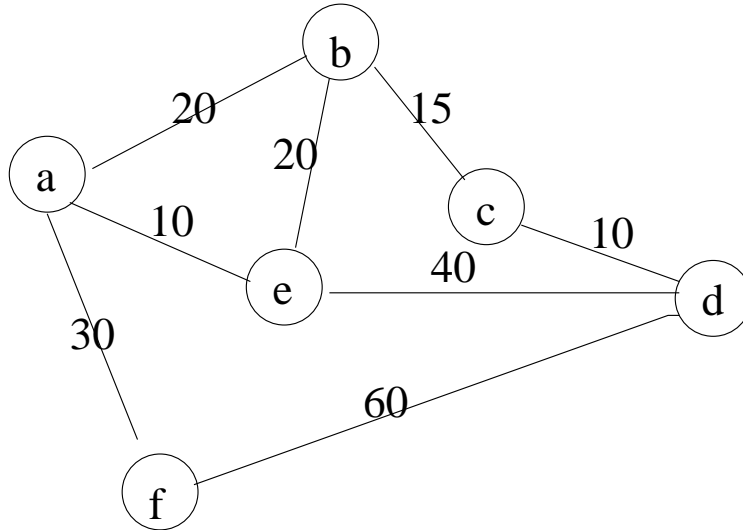
Questions

- | | |
|--------------------------------|------|
| 1. Prolog | [30] |
| 2. Search | [26] |
| 3. Knowledge Based Systems | [14] |
| 4. Machine Learning Basics | [43] |
| 5. Neural Networks | [15] |
| 6. Evolutionary Computation | [32] |
| 7. Natural Language Processing | [20] |

Question 1. Prolog and search

[30 marks]

Suppose there are six cities a, b, c, d, e and f. The figure below shows the roads between the cities and the lengths of the roads.



(a) [6 marks] Represent the road information in the above figure as Prolog facts (or rules). All the roads are two-way roads.

(b) [10 marks] Write a Prolog program to find the distance between two cities. For example, if you ask a query

```
|?-distance(a, c, L).
```

it should print

```
L=35
```

Please note that you don't need to find the shortest distance. So any distance such as 45, 60 and 100, etc. is also acceptable. In fact, if you use ";" to force a backtracking, it should find another distance if there is another path, or say "no" if there is no more answer. For example,

```
L = 45    ?;  
L = 60    ?;  
L = 100   ?;  
.....  
no.
```

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.

(c) [10 marks] Write a new version of your program so that it not only finds the distance, but also prints the cities on the path in the right order, for example,

```
|-find_a_path(a, c, L).  
a b c  
L = 35
```

Please note that you are only required to find one of the paths. So it is acceptable to print any other paths such as

```
a e d c  
L = 60
```

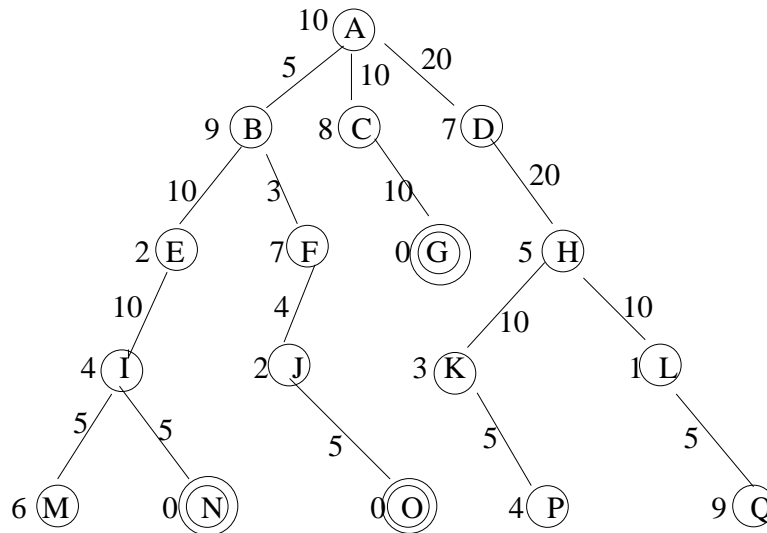
You may use the list library if you need to.

(d) [4 marks] What search strategy were you using to find a path in this question? State the main problems and limitations of this search strategy for this path finding problem.

Question 2. Search

[26 marks]

(a) [20 marks] Consider the following search tree. The start state is A and the three goal state nodes are N, O and G. The number next to the node is the heuristics $h(n)$ for each state which is the estimated cost from this node to a goal state node. The number on the edge is the cost $g(n)$ between the two nodes linked by the edge.



For each search strategy listed below, give a solution and list all the expanded nodes in the order that they are expanded. The first one is done for you as an example.

Depth first search

Solution path: A, B, E, I, N
 All expanded nodes: A, B, E, I, M, N

Breadth first search

Solution path:
 All expanded nodes:

Iterative deepening search

Solution path:
 All expanded nodes:

Uniform cost search

| |
|---------------------------------------|
| Solution path: All expanded nodes: |
|---------------------------------------|

Greedy best first search

| |
|---------------------------------------|
| Solution path: All expanded nodes: |
|---------------------------------------|

A* search

| |
|---------------------------------------|
| Solution path: All expanded nodes: |
|---------------------------------------|

(b) [6 marks] Suppose you are browsing our school web site to look for some specific information, for example, you want to find out what text book is used in COM307. This problem can be formalized as a state space search problem: the start state is our school home page; the goal state is a page with your required information, for example, the goal state for the text book example is “the comp307 course outline”; the operators are the page links, and each page is linked to may other web pages; a solution for the example problem is “school home page - courses - comp307 home page - course outline”.

What is a good search strategy for this problem? Justify your answer. (Hints: You will need to consider the size of the search space and whether it is finite. You also need to consider whether you have a good heuristic for selecting the good links and whether it is admissible.)



Student ID:

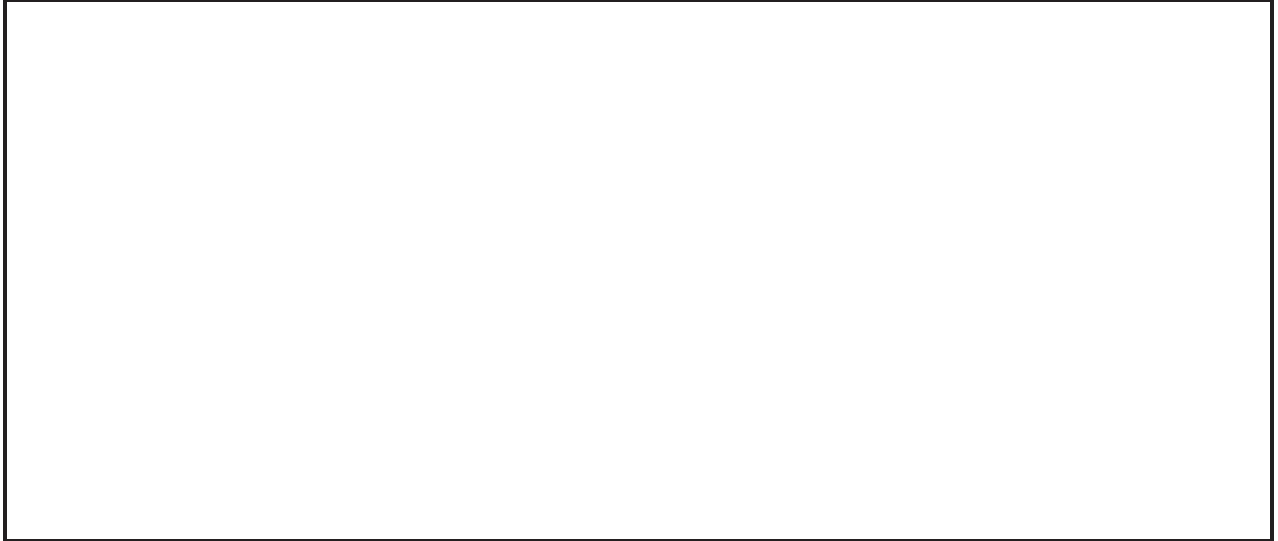
SPARE PAGE FOR EXTRA ANSWERS

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

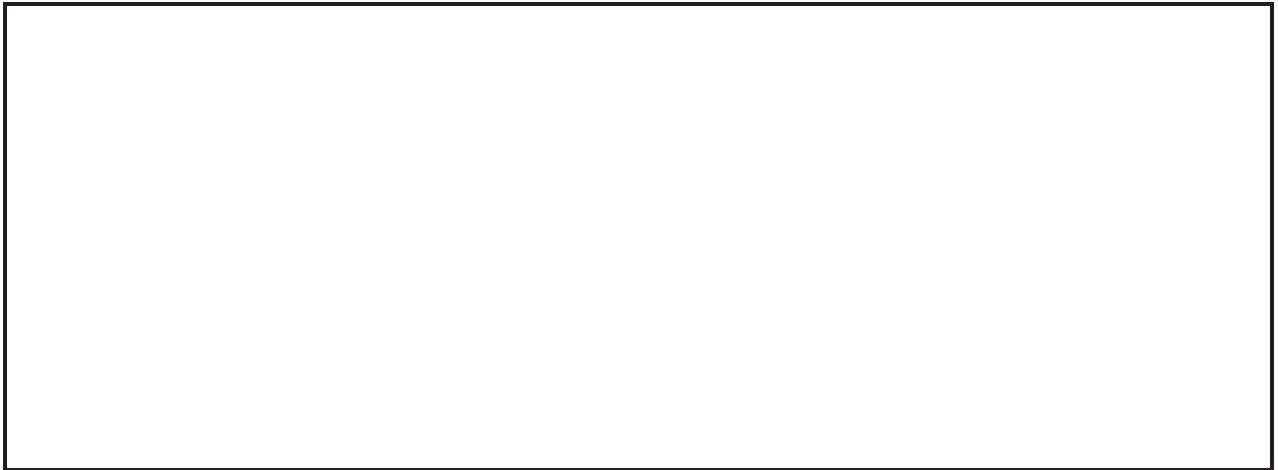
Question 3. Knowledge Based Systems

[14 marks]

(a) [5 marks] What are the two main components of a knowledge based system? Explain the main functions of the two components.



(b) [5 marks] State the main differences between backward chaining and forward chaining.



(c) [4 marks] State four limitations or challenges in knowledge based systems.

Question 4. Machine Learning Basics

[43 marks]

(a) [5 marks] Give the name of an algorithm or approach used in each of the following machine learning paradigms:

- (i) Case based learning
- (ii) Induction learning
- (iii) Statistical learning
- (iv) Connectionist learning
- (v) Evolutionary learning

(b) [4 marks] The data sets are typically separated into a *training set* and a *test set* in machine learning systems. Briefly describe the role for each of them.

(c) [4 marks] Briefly describe the term *overfitting* in the context of machine learning. State two methods to avoid (or reduce) overfitting.

(d) [4 marks] Briefly describe the *Nearest neighbour method* widely used in machine learning (classification).

(e) [6 marks] The *k-mean* method is widely used in machine learning.

(i) State whether this method is typically used for *classification*, *clustering*, or *association rules*.

(ii) This method has an important assumption. State this assumption.

(iii) Briefly describe the process of this method.

(f) [2 marks] When using the naïve Bayes method for classification, there is an important assumption on the attributes for the given task. State the assumption.

(g) [6 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: “game”, and “assignment” to characterise each message, and have counted how many of the messages contain each word:

| | spam | | email (non-spam) | |
|--------------|--------------|------------------|------------------|------------------|
| | word present | word not present | word present | word not present |
| “game” | 300 | 100 | 10 | 90 |
| “assignment” | 5 | 395 | 60 | 40 |
| Total count | 400 | | 100 | |

If your spam filter was presented with a new message that contained both words “game” and “assignment”, would your spam filter classify the message as spam or as email (non-spam)? Show your working.

(h) [8 marks] Consider the following data set describing 10 instances for a group of friends' parties, of which 5 were *BBQ* (a party outside) and 5 were *HomeParty* (at home) depending on the weather conditions. They are described by three attributes: whether it is clear or shower, whether the wind is strong or weak, and whether the temperature is hot, good, or cold.

| Instance | Weather | Wind | Temp | Class |
|----------|---------|--------|------|-----------|
| 1 | clear | weak | good | BBQ |
| 2 | clear | weak | hot | BBQ |
| 3 | clear | strong | good | BBQ |
| 4 | clear | weak | cold | BBQ |
| 5 | shower | weak | hot | BBQ |
| 6 | clear | strong | cold | HomeParty |
| 7 | shower | strong | hot | HomeParty |
| 8 | shower | strong | cold | HomeParty |
| 9 | shower | weak | cold | HomeParty |
| 10 | clear | strong | hot | HomeParty |

The group of friends want to build a decision tree to help making decisions for whether to have the party outside for BBQ or stay inside for a home party. Which attribute should be chosen for the root of the decision tree if the impurity function $p(\text{BBQ})p(\text{HomeParty})$ is used? Show your working.

(i) [4 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 | Output Class |
|----------------------------|----------------------------|-------------------------|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

In the perceptron, he used two input nodes and one output node corresponding to the two input features and the output class. He used the standard threshold transfer function and the simple perceptron learning algorithm discussed in our lectures. However, his perceptron could not converge no matter how he changed the learning parameters.

- (i) Explain why John's perceptron was not successful.
- (ii) Suggest two improvements that would enable the instances to be learned successfully.

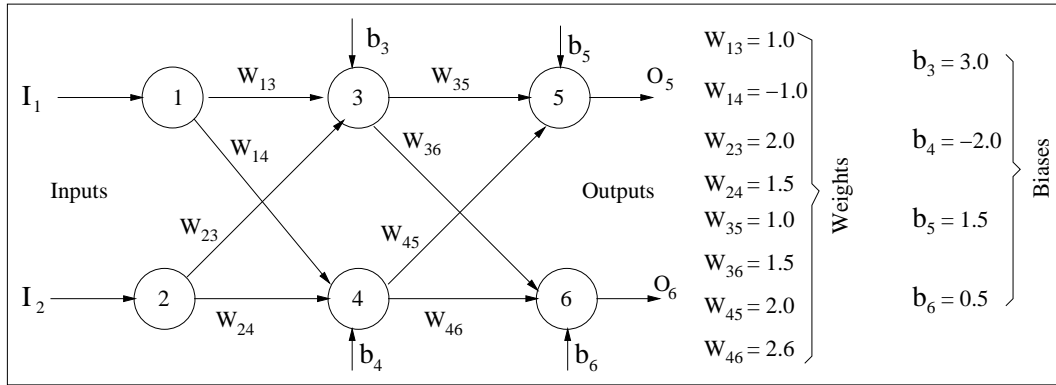
SPARE PAGE FOR EXTRA ANSWERS

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

Question 5. Neural Networks

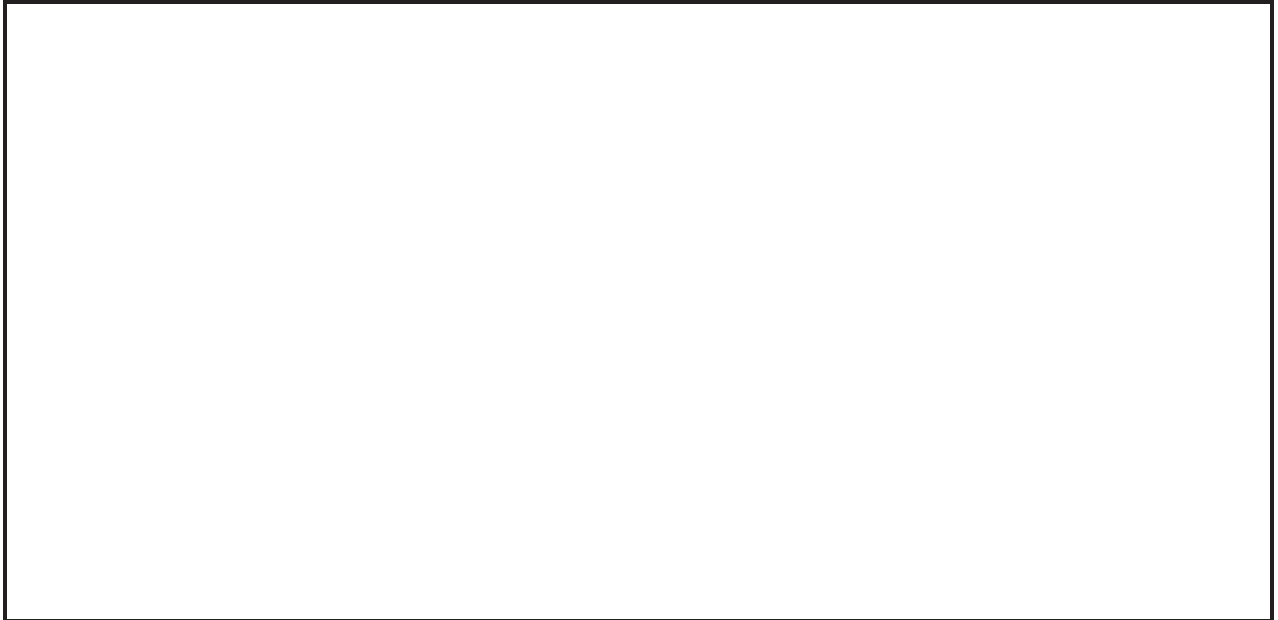
[15 marks]

(a) [9 marks] Consider the following feed forward neural network which uses the sigmoid/logistic transfer function (see Appendix B),



- (i) What will be the output of node 5 (O_5) for the input vector (0.0, 0.0)?
- (ii) What will be the new value of weight w_{35} after one epoch of training using the back propagation algorithm? Assume that the training set consists of only the vector (0.0, 0.0, 0.0, 0.0) for both input and output nodes and that the learning rate η is 0.3.

(b) [6 marks] Peter Smith has developed a neural network classifier for detecting potential *terrorists* from normal people using finger-print images. The process involves the extraction of 6 features from the recorded finger-print images and the use of the standard multilayer feed forward neural networks with the back propagation algorithm for classification. There are 300 examples in total, of which 50 are used for network training and 250 for testing. The network architecture he used is 6-25-2. After training for 20,000 epochs, the network classifier still performed badly on the test set. Suggest three good ways for improving the performance.



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.

Question 6. Evolutionary Computation and Learning

[32 marks]

(a) [3 marks] Genetic algorithms and genetic programming are two techniques in evolutionary computation and learning. State six additional techniques in evolutionary computation and learning.

(b) [6 marks] In evolutionary algorithms, the three main genetic operators, *elitism*, *crossover* and *mutation*, are often used to generate new candidate solutions. Briefly describe what each operator does and the main purpose for each of them.

(c) [2 marks] In evolutionary computation, there is a selection method called *tournament selection*. Briefly describe this method.

(d) [3 marks] Suppose that you want to use genetic programming (GP) for a particular task and that a GP package is available for you to use. Suggest three criteria that are commonly used for terminating a GP experiment run.

(e) [4 marks] Briefly describe the terms *sufficiency* and *closure* in the context of creating a primitive set in genetic programming.

(f) [6 marks] Genetic Programming (GP) is widely used for symbolic regression tasks. In assignment 4, you used GP to evolve a mathematical function to model the relationship between the output variable and the input variable(s) from a (training) set of instances.

Suppose your task is to use GP to map a single input variable x to a single output variable y from the following data set (20 points).

| | | | | | | | | | | |
|---|-------|-------|-------|-------|-------|-------|-------|-------|------|------|
| x | -2.0 | -1.75 | -1.50 | -1.25 | -1.00 | -0.75 | -0.50 | -0.25 | 0.00 | 0.25 |
| y | 37.00 | 24.16 | 15.06 | 8.91 | 5.00 | 2.72 | 1.56 | 1.10 | 1.00 | 1.04 |

| | | | | | | | | | | |
|---|------|------|------|------|------|------|------|------|-------|-------|
| x | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 | 1.75 | 2.00 | 2.25 | 2.50 | 2.75 |
| y | 1.06 | 1.04 | 1.00 | 1.10 | 1.56 | 2.72 | 5.00 | 8.91 | 15.06 | 24.16 |

- (i) Suggest a good terminal set.
- (ii) Suggest a good function set.
- (iii) Suggest a good fitness function.
- (iv) Describe two characteristics of the GP approach to such regression tasks compared with the conventional statistical regression method.

(g) [8 marks] The standard tree based GP approach has been applied to many classification tasks. In this approach, each evolved program typically returns a single floating point number. One of the key issues here is to use a *strategy* to translate the single output value of an evolved classifier program into a set of class labels.

(i) In assignment 4, GP was used to evolve a classifier to categorise the 699 instances in the Wisconsin medical data set into either the *benign* class or the *malignant* class. Suggest a strategy (rule) for translating the single program output into the above two classes.

(ii) Assuming you are going to use tree-based GP for a classification problem with four classes: *class1*, *class2*, *class3* and *class4*. Suggest a mapping strategy that can translate the single program output value into the four classes above.

(iii) State the advantages and limitations of your strategy for part (ii).

Question 7. Natural Language Processing

[20 marks]

(a) [10 marks] Write a DCG program to extract the *size* (number of bedrooms) and *price* (the rent) from the following rental property advertisements:

```
Karori 2br 2009 flat $300pw sunny big  
kelburn 3brms only $200pweek call me on 463 4756
```

Each advertisement is a character list e.g. [k, a, r, o, r, i, , 2, b, r, ...]. You may use two variables to save the size and the price, and show them as the output, for example,

```
S = [2]  
P = [3,0,0]
```

You may make assumptions to simply the problem. For example, you may assume only `br` or `brms` are used for *size* and only `pw` or `pweek` for *price*. But remember that any words (noise) can appear anywhere in the advertisements.

(b) [5 marks] State the main challenges in Natural Language Processing.

(c) [5 marks] State the main applications of Natural Language Processing.

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

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

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

$$O_i = f(I_i) = f\left(\sum_k w_{k \rightarrow i} \cdot o_k + b_i\right) \quad (3)$$

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

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

$$\beta_z = d_z - o_z \quad (6)$$

B Sigmoid/Logistic Function

