Programme Code: TU856, DT844
Module Code: CMPU 3051
CRN: 30102

TECHNOLOGICAL UNIVERSITY DUBLIN

Grangegorman

TU856 - BSc. Degree in Computer Science

Year 3

SEMESTER 1 EXAMINATIONS 2022/2023

CMPU3051 Introduction to Artificial Intelligence

Internal Examiner(s):
Mr. Richard Lawlor
Dr. Paul Doyle

External Examiner:
Ms. Pamela O'Brien

Instructions To Candidates:
Attempt 4 out of 5 questions. All questions carry equal marks.

Exam Duration: 2 hours

Special Instructions /Handouts/ Materials Required: None

1 (a) Explain what is meant by *state space search*.

In state space search, discuss the difference in performance between depth first and breadth first search. Assume each state has **b** successor or branching states. Is there a way to combine the advantages of both?

(9 marks)

(b) Represent the following statements in Predicate Calculus:

i. "Everyone who loves all animals is loved by someone". Use the predicates ***loves(X ,Y)***, ***animal(W)***.

ii. "The dog who lives with Bugsbunny does not chase cats". Use the predicates ***dog(X), cat(X), livesWith(A, B), chases(U, V)***.

(8 marks)

(c) Prolog uses backward chaining. Explain what this means with the help of a proof tree in terms of the following Prolog program when the Prolog interpreter is asked to prove `k(X)`. Make reference to *unification* in your answer.

```
f(a).
f(b).

g(a).
g(b).

h(b).

k(X) :- f(X), g(X), h(X).
```

(8 marks)

2 (a) Write down Bayes' rule.

Then use it to calculate P(m | s) in the following situation where s is the proposition that a patient has a stiff neck and m is the proposition that a patient has meningitis. The doctor knows some unconditional facts, namely: prior probability of a patient having meningitis is 1/50,000, prior probability of any patient having a stiff neck is 1%. The doctor also knows that having meningitis causes a stiff neck 70% of the time.

(7 marks)

(b) Give three reasons why Bayesian reasoning is difficult to use in an expert system. Also, briefly describe two advantages in using certainty factors as opposed to Bayesian reasoning.

Briefly mention two well-known historical expert systems, one of which used Bayesian reasoning and the other which used certainty factors.

(10 marks)

(c) Given the forecasting rules below, show how they would fire to forecast tomorrow's weather when provided with the following information: there is rain today and the rainfall is low with a certainty factor of 0.8, and also it is cold with a certainty factor of 0.9. This formula may be useful:

$$cf(cf_1, cf_2) = cf_1 + cf_2 \times (1 - cf_1)$$

Rule: 1
if today is rain
then tomorrow is rain {cf 0.5}

Rule: 2
if today is dry
then tomorrow is dry {cf 0.5}

Rule: 3
if today is rain
and rainfall is low
then tomorrow is dry {cf 0.6}

Rule: 4
if today is rain
and rainfall is low
and temperature is cold
then tomorrow is dry {cf 0.7}

Rule: 5
if today is dry
and temperature is warm
then tomorrow is rain {cf 0.65}

Rule: 6
if today is dry
and temperature is warm
and sky is overcast
then tomorrow is rain {cf 0.55}

(8 marks)

3 (a) Express the following Predicate Calculus function and statements in Prolog:

minincome(Y) = 15000 + 4000 * Y

savings_account(adequate) $\wedge$ income(adequate) $\Rightarrow$ investment(stocks).

$\forall$ X earnings(X, steady) $\wedge$ $\forall$ Y dependents(Y) $\wedge$ greater(X, minincome(Y))
$\Rightarrow$ income(adequate).

(5 marks)

(b) Write a recursive Prolog predicate **canGoFromTo(A, B)** to express the following: One can travel from A to B if there is a bus or train from A to B. Or one can travel from A to B if there is a bus or train from A to C and one can travel from C to B.

(5 marks)

(c) Write a Python function myDotProd(a,b) which computes the dot-product of 2 vectors of equal size using a loop. You are not to use the numpy.dot() operation.

(5 marks)

(d) Write Python code for a simple perceptron decision function **buyCocktail(x)**, which decides on whether to buy an expensive cocktail or not based on the following inputs where x= [x1, x2, x3] .

x1 = 1 means you have plenty of cash, x1 = 0 means you do not. x2 = 1 means you want to impress a new girlfriend, 0 means you don't. x3 = 1 means it's a busy exciting night on the town, 0 means it is not.

Choose the weights and threshold (or bias) so that if you want to impress someone, you will buy cocktails no matter what. However, if you are not out to impress but it's an exciting night and you have plenty of cash, you will buy a cocktail. Otherwise you will not buy a cocktail.

(5 marks)

(e) Suppose you were to train a perceptron to behave as described in part(d) instead of handcrafting the weights. Write down an appropriate list of input training vectors and the corresponding list of desired outputs which could be used to train such a perceptron.

(5 marks)

4 (a) What is a perceptron? Draw a diagram for a single two-input perceptron. Explain what is meant by linear separability. How many categories can it classify inputs into? Mention any issues involved in using a perceptron to compute logical AND, OR and XOR.

(7 marks)

(b) Illustrate the single perceptron training algorithm for an OR gate by computing the weight adjustments for the first three epochs in Table 4(b) below.

(5 marks)

(c) Distinguish between a sigmoid neuron and a perceptron, and outline the difference in the training formula for Δw_i.

Why might it be preferable to use sigmoid neurons in a neural network rather than perceptrons?

(8 marks)

(d) Provide a Python code segment which implements the perceptron training algorithm.

(5 marks)

Perceptron Training for logical OR

Threshold $\theta = 0.2$, learning rate $\alpha = 0.1$.

Epoch	Inputs		Desired output	Weights		Weighted Sum	Actual Output	Error	Weight adjustments	
	x_1	x_2	Y_d	w_1	w_2	z	a	e	Δw_1	Δw_2
1	0	0	0	03	-0.2					
	0	1	1							
	1	0	1							
	1	1	1							
2	0	0	0							
	0	1	1							
	1	0	1							
	1	1	1							
3	0	0	0							

Table 4(b)

5 (a) Suggest a network architecture appropriate for MNIST and explain why the cross-entropy cost function might be preferable to the mean-squared-error when the sigmoid activation function is used.

(5 marks)

(b) Explain the problem of overtraining a neural network and how it could be alleviated. How does MNIST address this problem?

(5 marks)

(c) Briefly explain what gradient descent means and how it relates to the cost or loss function. Explain what each of the following symbols represents and the relationship between them in gradient descent.

$$\Delta w_{ij}^{l} \text{ and } \frac{\partial C}{\partial w_{ij}^{l}} ?$$

(5 marks)

(d) At the core of the backpropagation algorithm are the matrix formulae:

$$\frac{\partial C}{\partial w_{jk}^{l}} = \delta_j^l \, a_k^{l-1}$$

$$\delta^L = (a^L - y) \odot \sigma'(z^L) \qquad \delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$$

Briefly explain what each term means and which one in particular relates to backpropagation.

Also, for a three layer network, write a Python code fragment which includes each of these formulae.

(10 marks)