S228/401C

DUBLIN INSTITUTE OF TECHNOLOGY
KEVIN STREET, DUBLIN 8

BSc. (Hons) in Computer Science

Stage 4

SEMESTER 2 EXAMINATIONS 2011/2012

ARTIFICIAL INTELLIGENCE II

Dr. John Kelleher
Dr. Deirdre. Lillis
Mr. D. Tracey

Monday
14^{th} May 2012
4:00 p.m to 6:00 p.m

Question 1 is **compulsory**

Answer Question 1 (40 marks) **and**

any 2 Other Questions (30 marks each).

1. (a) Explain what is meant by **inductive learning**.

(5 marks)

(b) For some data sets it is possible to devise multiple hypotheses that are consistent with the data. Describe a heuristic for choosing among multiple consistent hypotheses and explain why your heuristic is reasonable.

(5 marks)

(c) Describe the differences between **lazy learners** and **eager learners**, giving examples of each.

(10 marks)

(d) Let us say we have three classification algorithms. How can we order these three from best to worst?

(20 marks)

Table 1: Class Exam Results: a 1 indicates the student possesses the feature listed in the column and 0 indicates that they do not. The final column lists whether or not the student was awarded a 1^{st} this year

Student	1^{st} last year	Works hard?	Blonde	1^{st} this year
X	1	1	0	1
Y	0	1	1	0
Z	0	1	0	0

Table 2: The attributes of the student whose script was not marked. A 1 indicates the student possesses the feature listed in the column and 0 indicates that they do not. The column on the right contains a ? because they have not been graded yet.

Student	1^{st} last year	Works hard?	Blond	1^{st} this year
U	1	0	0	?

2. (a) Discuss the advantages and disadvantages of ***k*-Nearest Neighbour** classification.

(5 marks)

(b) Just before an exam board a lecturer finds an exam script by student U that the lecturer had forgotten to mark. The lecturer does not have time to correct the script before the exam board, so they decide to use a nearest-neighbour approach to decide whether or not to award student U a 1^{st}. The case base of results the lecturer used is listed in Table 1 and the attributes of student U are listed in Table 2 .

(i) Assuming the lecturer uses Euclidean distance

$$d(x_1, x_2) = \sqrt{\sum_{r=1}^{n}(a_r(x_1) - a_r(x_2))^2}$$

as their distance metric, compute the distance between the student U and each of the students in the case base.

(5 marks)

(ii) Given that the lecturer used ***1*-NN** classification was student U awarded a 1^{st}?

(5 marks)

(iii) If the lecturer used **3-NN** classification would student U be awarded a 1^{st}?

(5 marks)

(c) Table 3, on the next page lists a classification dataset. Each instance in the dataset has two explanatory attributes (attribute 1 and attribute 2) and is classified as either a positive (+) or a negative(-) example.

(i) Calculate the classification **entropy** for this dataset.

(5 marks)

(ii) Calculate the **information gain** for attribute 1 and attribute 2.

(5 marks)

Attribute 1	Attribute 2	Classification
T	T	+
T	F	-
T	F	+
T	T	+
F	T	-

Table 3: Classification Dataset

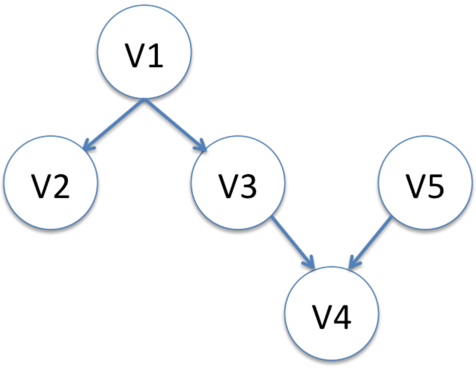

Figure 1: An example Bayesian network.

Table 4: Full joint distribution for a dentist visit

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

3. (a) Given the full joint distribution shown in Table 4, calculate $\mathbf{P}(Toothache|cavity)$.

(5 marks)

(b) Express the joint probability distribution for the Bayesian network shown in Figure 1 using the chain rule.

(10 marks)

(c) Consider the following time keeping patterns of the lecturers in your college:

- 25% of lecturers start 75% of their lectures on time and 25% late.
- 50% of lecturers start 50% of their lectures on time and 50% late.
- 25% of lecturers start 25% of their lectures on time and 75% late.

(i) Given that both the 1^{st} and 2^{nd} Artificial Intelligence lectures of the year started on time, compute the posterior probability that your Artificial Intelligence lecturer follows each of the three time keeping patterns.

(10 marks)

(ii) Given that both the 1^{st} and 2^{nd} Artificial Intelligence lectures of the year started on time, what is the Bayesian Prediction that the 3^{rd} Artificial Intelligence lecture will start on time?

(5 marks)

x	2	4	6	8
y	2	5	5	8

Table 5: Example Dataset for Linear Regression Question

4. (a) Assuming a domain with one explanatory variable x and one dependent variable y linear regression uses the following formula to model the relationship between the explanatory and dependent variable:

$$f(x) = w_1 x + w0$$

where $w1$ and $w0$ are computed using the following formulae (where M is number of data points in the dataset):

$$w_1 = \frac{(M \sum_{i=1}^{M} x_i y_i) - (\sum_{i=1}^{M} x_i \sum_{i=1}^{M} y_i)}{(M \sum_{i=1}^{M} x_i^2) - (\sum_{i=1}^{M} x_i)^2}$$

$$w_0 = \left(\frac{1}{M} \sum_{i=1}^{M} y_i\right) - \left(\frac{w_1}{M} \sum_{i=1}^{M} x_i\right)$$

Using the data in Table 5 compute the values of w_0 and w_1 that provide the best linear fit to the data.

(10 marks)

(b) Figure 2 shows a backprogation network that is currently processing the training vector $[1.0, 0.9, 0.9]$ which has an associated target vector $[0.1, 0.9, 0.1]$. Given that the output from unit B is 0.6 and from C is 0.8, and assuming that the activation function used at all nodes in the network is the logistic function (i.e., $f(x) = \frac{1}{1+\exp^{-x}}$):

(i) Calculate the actual output vector (to 3 decimal places).

(5 marks)

(ii) Calculate the error for each output unit.

(5 marks)

(iii) Calculate the error for each hidden unit B and C.

(10 marks)

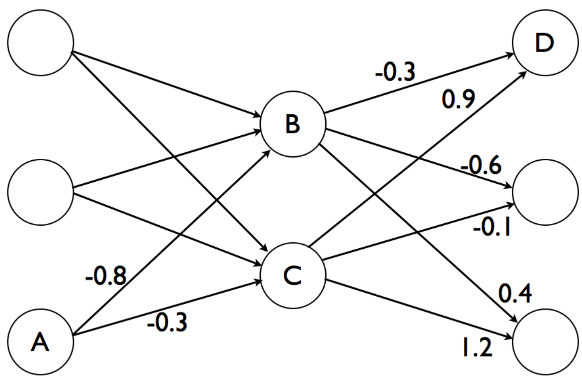

Figure 2: Example Neural Net