RS249/401C

DUBLIN INSTITUTE OF TECHNOLOGY
KEVIN STREET, DUBLIN 8

BSc. (Hons) in Computer Science

Stage 4

SUPPLEMENTAL EXAMINATIONS 2012

ARTIFICIAL INTELLIGENCE II

Dr. John Kelleher
Dr. Deirdre. Lillis
Mr. Ray Walshe

Duration: 2 Hours

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) Explain what can go wrong when a machine learning classification algorithm uses the wrong inductive bias.

(5 marks)

(c) Inductive machine learning is often referred to as an **ill-posed problem**. What is meant by this description?

(10 marks)

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

(20 marks)

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

(5 marks)

(b) IT-Tunes is on online music sales service that is trying to build an inductive machine learning system to recommend new songs to its users. They have collected a dataset that records details of songs they sell and details of ratings that users have given these songs (missing values in the ratings columns indicate that a user has not rated a particular song). The table below shows an extract from this dataset showing details of 5 songs and associated ratings for 5 users (note that in the full dataset there are many more songs and many more users).

ID	001	004	007	011	022
Title	Jeremy	One	Please	Fool's Game	Help!
Album	Ten	Achtung Baby	Pop	M. Bolton	Help!
Artist	Pearl Jam	U2	U2	M. Bolton	The Beatles
Year	1990	1992	1997	1983	1965
Genre	Rock	Rock	Rock	Soul	Pop
Best Chart Pos	10	1	6	25	1
User1-Score	3	2	2	5	4
User2-Score	5	-	5	-	4
User3-Score	5	2	2	1	1
User4-Score	1	-	-	-	1
User5-Score	-	1	1	5	3

(i) In order to build a case based reasoning (CBR) classification system to predict whether a user would like a particular song, or not, a case representation is required. Describe a case structure for the dataset shown in the table above that could be used to build a CBR system that would predict the rating a particular user would be likely to give to a particular song. Justify any decisions that you make and explain any assumptions used.

(5 marks)

(ii) A good similarity measure is crucial for any CBR system. Describe a similarity measure that would be appropriate for comparing cases in the structure described in the answer part (i) of this question. Justify any decisions that you make and explain any assumptions used.

(5 marks)

(iii) Is **case-based reasoning** the most appropriate inductive machine learning technique to use for this prediction problem?

(5 marks)

(c) In the context of Decision Tree Learning define what is meant by the following terms:

(i) entropy

(5 marks)

(ii) information gain

(5 marks)

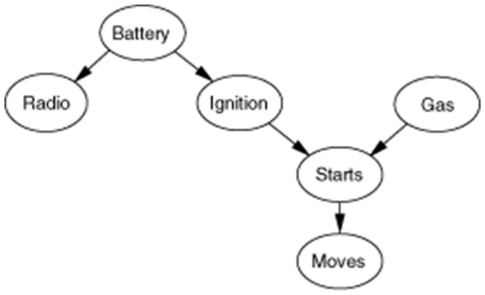

Figure 1: A Bayesian network of a car's electrical system and engine. Each variable is boolean and the true value indicates that the corresponding aspect of the vehicle is in working order.

Table 1: Full joint probability 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 1, calculate:

$$\mathbf{P}(Cavity|toothache \vee catch)$$

(5 marks)

(b) Consider the network for car diagnosis shown in Figure 1.

(i) Extend the network with the Boolean variables *IcyWeather* and *StarterMotor*; assume that a *StarterMotor* effects whether or not the car *Starts* and that *Icy Weather* effects the *Battery* and the *StarterMotor*,

(5 marks)

(ii) How many independent values are contained in the full joint probability distribution?

(5 marks)

(c) You are on holidays on Fisher Island. The yearly weather on Fisher Island comes in five different varieties:

- there is a 10% chance that there will be rain everyday of the year.
- there is a 20% chance that there will be rain on 75% of the days of the year.
- there is a 40% chance that there will be rain on 50% of the days of the year.
- there is a 20% chance that there will be rain on 25% of the days of the year.
- there is a 10% chance that there will be no rain on any day of the year.

(i) Given that it has rained on day 1 and 2 of the year compute the posterior probability of each of the 5 yearly weather patterns on day 2 of the year. Give your answer rounded to four places of precision.

(10 marks)

(ii) Given that after the first 10 days of the year the weather has been such that the posterior probabilities of each of the 5 varieties of the yearly weather on Fisher Island are:

- there is now a 90% chance that there will be rain everyday for the rest of the year;
- a 7% chance that there will be rain on 75% of the rest of the days of the year;
- a 2% chance that there will be rain on 50% of the rest of the days of the year;
- a 1% chance that there will be rain on 25% of the rest of the days of the year;
- and there is a 0% chance that there will be no rain for the rest of the year.

What is the Maximum a Posterior (MAP) probability of rain on day 11?

(5 marks)

x	0	1	2	3	4
y	3	6	7	8	11

Table 2: 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 2 compute the values of w_0 and w_1 that provide the best linear fit to the data.

(10 marks)

(b) What does it mean if two classes C_1 and C_2 are described as **linearly separable**?

(5 marks)

(c) Describe the processing stages of a McCulloch-Pits "unit".

(7 marks)

(d) Figure 2 is a schematic of a 3 input perceptron. Input a_0 is fixed at $a_0 = -1$, inputs a_1 and a_2 are binary. The perceptron uses a threshold activation function that outputs a 1 if the weighted sum of inputs is greater than 0 and a 0 otherwise. Define the **truth-table of the function** that this perceptron implements *and* identify the **name of the function**.

(8 marks)

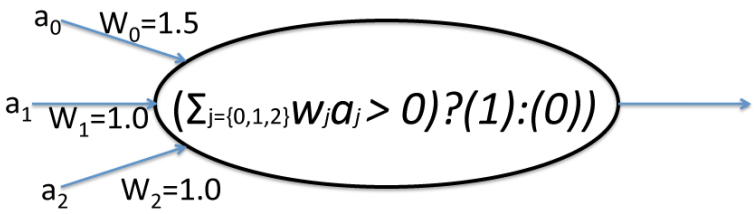

Figure 2: A 3 input perceptron. Input $a_0 = -1$, inputs a_1 and a_2 are binary. The perceptron uses a threshold activation function that outputs a 1 if the weighted sum of inputs is greater than 0 and a 0 otherwise.