S228/406

DUBLIN INSTITUTE OF TECHNOLOGY
KEVIN STREET, DUBLIN 8

BSc (Hons) in Computer Science

Stage 4

SEMESTER 2 EXAMINATIONS 2009

ARTIFICIAL INTELLIGENCE 2

Dr. John Kelleher
Dr. D. Lillis
Dr. I. Arena

Duration: 2 Hours

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

any 2 Other Questions (30 marks each).

Table 1: Joint Distribution for X and Y

	$X = x_1$	$X = x_2$
$Y = y_1$	0.02	0.30
$Y = y_2$	0.14	0.32
$Y = y_3$	0.10	0.12

1. (a) Given the joint distribution for X and Y listed in Table 1 calculate $P(Y = y_2)$

(5 marks)

(b) Given the joint distribution for X and Y listed in Table 1 calculate $P(Y = y_2|X = x_1)$

(5 marks)

(c) In the context of machine learning, explain what is meant by **overfitting** the training data.

(5 marks)

(d) In the context of inductive learning explain what is meant by a **consistent hypothesis**.

(5 marks)

(e) What is the aim of **inductive logic learning**?

(5 marks)

(f) In the context of inductive logic learning, what is meant by the **extension** of a hypothesis?

(5 marks)

(g) In the context of machine learning distinguish between **false negatives** and **false positives**.

(5 marks)

(h) In the context of machine learning, what does it mean if two classes C_1 and C_2 are described as **linearly separable**?

(5 marks)

2. (a) In your local power station, there is an alarm that senses when a temperature gauge exceeds a given threshold. The gauge measures the temperature of the core. Consider the Boolean variables A (alarm sounds), F_A (alarm is faulty), and F_G (gauge is faulty); and multivalued nodes G (gauge reading) and T (actual core temperature).

   (i) Draw a Bayesian network for this domain, given that the gauge is more likely to fail when the core temperature gets too high.

   (5 marks)

   (ii) Suppose there are just two possible actual and measured temperatures: normal and high. The probability that the gauge gives the correct temperature is x when it is working, but y when it is faulty. Give the conditional probability table associated with node G.

   (5 marks)

   (b) Suppose you are a security guard at some secret underground installation. You want to know whether it's raining today, but your only access to the outside world occurs each morning when you see the director coming in with, or without, an umbrella. For each day t, the set $\mathbf{E}_t$ contains a single evidence variables U_t (whether the umbrella appears), and the set $\mathbf{X}_t$ contains a single state variable R_t (whether it is raining). Figure 1 provides the Bayesian network structure and conditional distributions that describe this scenario.

   (i) Assuming that you have a prior belief about whether it rained on day 0, just before the observation sequence begins of: $\mathbf{P}(R_0) =< 0.5, 0.5 >$ and that the umbrella appears on day 1, so $U_1 = true$, **compute the probability that it rained on day 1**, (i.e., compute $\mathbf{P}(R_1|u_1)$).

   (10 marks)

   (ii) Using the value for $\mathbf{P}(R_1|u_1)$ that you computed in part (i) of this question, and assuming that the umbrella appears on day 2 (i.e. $U_2 = true$), **compute the probability that it rained on day 2** (i.e., compute $\mathbf{P}(R_2|u_1, u_2)$).

   (10 marks)

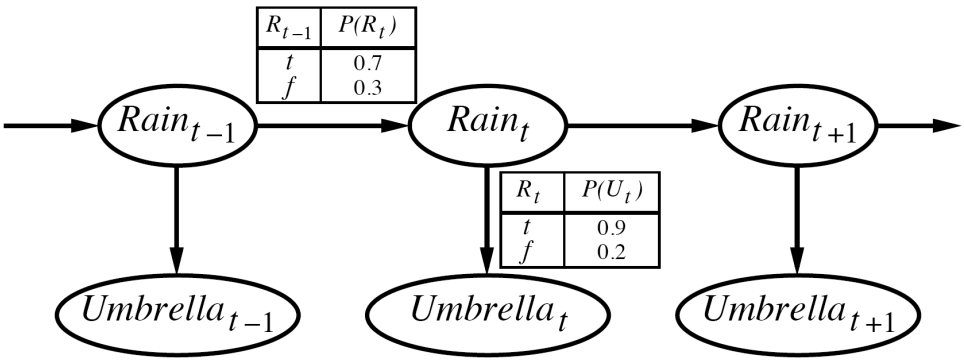

Figure 1: Bayesian network structure and conditional distribution describing the umbrella world. The transition model is $P(Rain_t|Rain_{t-1})$ and the sensor model is $P(Umbrella_t|Rain_t)$

X	Y	Class
T	T	+
T	F	-
T	F	+
T	T	+
F	T	-

Table 2: X and Y Classification Data

3. (a) In the context of machine learning, distinguish between **supervised** and **unsupervised** learning.

(5 marks)

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

(10 marks)

(c) Table 2 provides a classification for a data set of X Y pairs.

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

(5 marks)

(ii) Calculate the **information gain** for X and Y.

(10 marks)

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

(10 marks)

(b) Figure 2 shows a backprogation network, with weights as shown and all biases set to 0, 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)

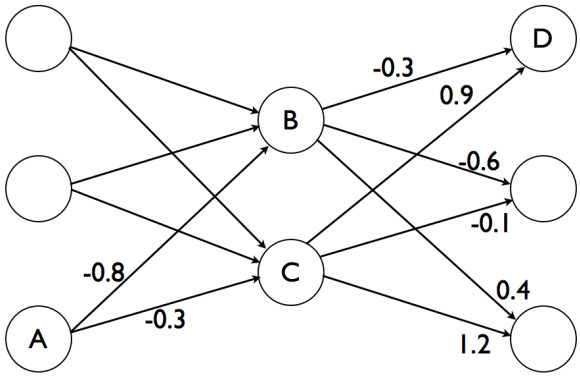

Figure 2: Example Neural Net