R228/419C

DUBLIN INSTITUTE OF TECHNOLOGY
KEVIN STREET, DUBLIN 8

BSc. (Hons) in Computer Science

Stage 4

SUPPLEMENTAL EXAMINATIONS 2014

ARTIFICIAL INTELLIGENCE II

Dr. John Kelleher
Dr. Deirdre. Lillis
Mr. P. Collins

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) Distinguish between **supervised** and **unsupervised** learning.

(5 marks)

(c) Distinguish between **classification learning** and **regression learning**.

(5 marks)

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

(5 marks)

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

(10 marks)

(f) Explain how do machine learning algorithms deal with the fact that machine learning is ill posed.

(10 marks)

Table 1: Dataset for the 3-Nearest Neighbor question

ID	Feature1	Feature2	Target
101	4	180000	C1
102	3	120000	C2
103	7	360000	C2
104	5	420000	C1
105	8	480000	C2

Table 2: Query instance for the 3-Nearest Neighbor question.

ID	Feature1	Feature2	Target
250	4	240000	?

2. (a) Table 1 lists a dataset containing examples described by two descriptive features, **Feature1** and **Feature2**, and labelled with a target class **Target**. Table 2 lists the details of a query for which we want to predict the target label. We have decided to use a **3-Nearest Neighbor** model for this prediction and we will use Euclidean distance as our distance metric:

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^{n} \left((x_1.f_i - x_2.f_i)^2 \right)}$$

(i) What target class (**C1** or **C2**) will our **3-Nearest Neighbor** model label the query with? Provide an explanation for you answer.

(5 marks)

(ii) There is a large variation in range between **Feature1** and **Feature2**. To account for this we decide to normalize the data. Compute the normalized data to four decimal points of precision using range normalization

$$x_i.f' = \frac{x_i.f - min(f)}{max(f) - min(f)}$$

(5 marks)

(iii) Assuming we use the normalized dataset as input, what target class (**C1** or **C2**) will our **3-Nearest Neighbor** model label the query with? Provide an explanation for you answer.

(5 marks)

(b) Table 3 (on the next page) lists a data set with of 6 examples described in terms of 3 binary descriptive features (**A**, **B**, and **C**) and a target feature (**Target**). You are asked to create a decision tree model using this data. **Which of the descriptive features will the ID3 decision tree induction algorithm choose as the feature for the root node of the decision tree?** Support you anwer with appropriate calculations and dicussions of your results. Note that Table 4 (also on the next page) lists some equations that you may find useful for this question.

(15 marks)

ID	A	B	C	Target
1	1	0	1	C1
2	1	1	1	C2
3	1	0	1	C1
4	0	1	1	C2
5	0	1	0	C1
6	0	1	1	C2

Table 3: Dataset for the ID3 Algorithm Question

$$
\begin{aligned}
\text{Entropy(DS)} &= -\sum_{i=1}^{k} p_i \times log_2(p_i) \\
\text{Remainder(F)} &= \sum_{v \in Domain(F)} \frac{|DS_v|}{|DS|} Entropy(DS_v) \\
\text{InformationGain(F,DS)} &= Entropy(DS) - Remainder(F)
\end{aligned}
$$

Table 4: Equations from information theory.

Table 5: Query Document

Machine learning is fun

Table 6: Document counts from the corpus for the words in the query.

Document counts for the **dislike** data set

fun	is	machine	learning
415	695	35	70

Document counts for the **like** data set

fun	is	machine	learning
200	295	120	105

3. Lets assume we are given a set of **700** training documents that a friend has classified as **dislike** and another **300** documents that they have classified as **like**. We are now given a new document and asked to classify it. Table 5 lists the content of this query document and Table 6 gives the number of documents from each class (**dislike** and **like**) that the words in the query document occurred in. **What class will a Naive Bayes prediction model label the query document as belonging to?** (You must support your answer by showing the calculations that a Naive Bayes model will make.)

(30 marks)

4. (a) The following model is commonly used for continuous prediction tasks:

$$y(x) = w_0 + w_1 x_1 + \ldots + w_D x_D$$

(i) Provide the name for this model and explain all of the terms that it contains.

(5 marks)

(ii) Explain how the following model can overcome some of the limitations of the model given above.

(8 marks)

$$y(x) = \sum_{j=0}^{M-1} w_j \phi_j(x)$$

(b) Explain why a **linear threshold perceptron** can represent the AND and OR functions but not the XOR function.

(7 marks)

(c) Describe how perceptron training can be viewed as gradient descent search through an error landscape.

(10 marks)