NUI Galway
OÉ Gaillimh

Semester 1 Examinations SAMPLE

Exam Code(s)	1MAO, 1MAI, 1CSD, 4BCT1, 4BP1, 4BS2, SPE
Exam(s)	MSc in Computer Science - Artificial Intelligence - Online MSc in Computer Science - Artificial Intelligence MSc in Computer Science - Data Analytics Fourth BSc in Computer Science & Information Technology Fourth BE in Electronic & Computer Engineering Fourth Bachelor of Science (Honours) (CS Pathway) Structured PhD
Module Code(s)	CT4101, CT5143
Module(s)	Machine Learning
Paper No.	1
Repeat Paper	No
External Examiner(s)	Dr Jacob Howe
Internal Examiner(s)	*Prof. Michael Madden Dr Patrick Mannion

Instructions:

This is an open book exam.
During the first 30 minutes of the exam period, you must complete an MCQ (worth 25%) on Blackboard.
During the remaining 90 minutes of the exam period, you must answer any three questions on this exam paper.
All questions on this exam paper carry equal marks (25% per question).
[THIS SAMPLE ONLY HAS 2 QUESTIONS, NOT 4.]

Duration	2 hours
No. of Pages	4
Discipline(s)	Computer Science
Course Co-ordinator(s)	Dr Desmond Chambers (BCT), Dr Frank Glavin (CSD), Dr Matthias Nickles (MAI), Dr James McDermott (MAO)

Requirements:

MCQ	Release to Library: Yes
Handout	None
Statistical/ Log Tables	None
Cambridge Tables	None
Graph Paper	None
Log Graph Paper	None
Other Materials	None
Graphic material in colour	No

[PTO]

1.

ID	Births Live Young	Lays Eggs	Feeds Offspring Own Milk	Warm-blooded	Cold-Blooded	Land and Water Based	Has Hair	Has Feathers	Class
1	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	mammal
2	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	amphibian
3	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	mammal
4	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	bird
Q	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	???

The dataset above contains four examples of animals that have been classified as mammals, amphibians or birds, based on their features. You are required to apply similarity-based learning to this dataset to develop a model that can be used to classify newly discovered species (such as the unclassified species ID = Q above).

i. Propose a suitable method to measure similarity between the instances in this dataset (i.e. a distance metric or a similarity index covered in the lecture notes). Provide a detailed description of your chosen method in your own words, and justify your choice of method by explaining why it is more suitable than any other two methods covered in the lecture notes. [8]

ii. Using your chosen method from part i. above, compute the similarity/distance between the query Q and each of the four instances in the training set. [8]

iii. If you chose a 1-NN model with uniform weighting, what class would be assigned to the query Q? [2]

iv. If you chose a 4-NN model with uniform weighting, what class would be assigned to the query Q? [2]

v. In your opinion, is the 1-NN or the 4-NN model more suitable for this dataset? Provide an explanation in your own words to support your opinion. [5]

[25]

[PTO]

2. (a) You have received the following message from a biomedical scientist. Prepare a comprehensive reply in your own words, making reference to concepts including bias, variance, and hypothesis complexity.
"*I am trying to predict whether or not a drug will be successful for treating the symptoms of COVID19, based on the characteristics of the drug. I have done an initial study with data divided into a training set and a testing set. First I used linear regression and found that the model's root mean squared error was very high, even on the training data. Then I used a polynomial regression with terms to the power of 3 and 4, and found that the model worked better on the training data, but very poorly on the test data. Can you explain why I got these two results? Following on from this, can you give me at least two recommendations on what I should try next, with explanations for what you propose?* [10]

(b) *"The cost function for multiple linear regression has a global optimum."* Explain in your own words what this means. What are the implications of this for applying techniques such as gradient descent to linear regression? [3]

(c) The two main ways of generating a Bayesian network structure are: (1) construct it by hand; (2) learn it from data. Outline the basic idea of these two approaches, in your own words. [6]

(d) Explain in your own words how to determine probabilities associated with a Bayesian network structure, after the structure has been determined. As part of your answer, explain what a conditional probability table (CPT) is. Referring back to the earlier parts of this question, provide an example of a single CPT, with probability values that you have made up. [6]

[END]