NUI Galway
OÉ Gaillimh

# ***Semester 1 Examinations 2019/2020***

| | |
|---|---|
| **Exam Code(s)** | 1MAO, 1MAI, 1CSD, 4BCT1, 4BP1, 4BS2, SPE |
| **Exam(s)** | MSc in Computer Science - Artificial Intelligence - Online<br>MSc in Computer Science - Artificial Intelligence<br>MSc in Computer Science - Data Analytics<br>Fourth BSc in Computer Science & Information Technology<br>Fourth BE in Electronic & Computer Engineering<br>Fourth Bachelor of Science (Honours) (CS Pathway)<br>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<br>Dr Patrick Mannion |
| **Instructions:** | Answer two questions from Section A and two questions from Section B.<br>Use a separate answer book for each section.<br>All questions carry equal marks. |
| **Duration** | 2 hours |
| **No. of Pages** | 5 |
| **Discipline(s)** | Computer Science |
| **Course Co-ordinator(s)** | Dr Desmond Chambers (BCT), Dr Enda Howley (CSD), Dr Michael Schukat (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]**

# SECTION A

**1.** (a) Chess is a two-player strategy board game that is played on a checkered board with 64 squares arranged in an 8x8 grid. The game does not have any hidden information (i.e. both players are aware of the full state of the game board). Each player begins the game with 16 pieces; one of these pieces is called the king. The goal of the game is to checkmate the opponent's king by placing it under an inescapable threat of capture. A game of chess ends when a checkmate is reached, when one of the players resigns, or if a draw is declared. Games of chess often have a long duration (e.g. 100+ moves).

A fellow student is working on an agent to play chess, and has asked you to look over their code. You see that they have implemented the Q-learning algorithm, with the following parameter values:

- Learning rate = 0.1
- Discount factor = 0.0
- Exploration rate = 0.7

Evaluate the student's choices for each of the parameters above. If appropriate, recommend alternative parameter values. Justify your evaluation and recommended alternative parameter values (if any) by explaining the impact of each parameter on the Q-learning process. [12]

(b) Explain what is meant by the **Markov property**. Use the game of chess as an example in your answer. [3]

(c) You have received the following email message from a non-technical manager who works for a company that runs an online platform for discussion forums. Prepare a detailed reply.

*"My company wishes to develop a system to identify forum posts that contain spam. The content of spam posts on our forums changes quite frequently, and currently our moderators spent a large amount of time reviewing forum posts. I have heard that machine learning models could be useful for automatic identification of suspicious posts. A colleague mentioned that machine learning algorithms may be categorised as* ***lazy*** *or* ***eager*** *– could you tell me what the difference is? What are the main strengths and weaknesses of each of these two categories? Which category of algorithm would be more suitable for our application? It would also be helpful if you could provide a specific example of an algorithm from each category."* [10]

**2.** (a) Explain the difference between the following machine learning tasks: **classification**, **regression**. For each task type, list a specific application, and an algorithm which may be used to learn a suitable model. [7]

(b) Recommend a procedure to select a suitable value of the parameter *k* when applying the k Nearest Neighbours algorithm to a regression task. Your answer should also discuss the impact of using a weighting scheme. [6]

(c) A colleague with very little prior experience of machine learning has asked for your advice on selecting an appropriate **hypothesis language**. Describe to your colleague with the aid of diagrams **what a hypothesis language is**, and include discussions of relevant issues (including **underfitting** and **overfitting**). [12]

[PTO]

**3.** (a) Explain what is meant by the term **Curse of Dimensionality** in the context of machine learning tasks.

Discuss how **heuristics** may be used to mitigate the effects of the Curse of Dimensionality. Your answer should include a specific example of a heuristic used in conjunction with an ML algorithm. [6]

(b) Identify a **scale invariant** method to measure similarity between instances in a dataset. Explain how this method may be applied using the aid of a diagram and an appropriate equation. [4]

(c) Briefly explain what is meant by the term **pure inductive learning**.

Describe in detail an algorithm which can be used for pure inductive learning of decision trees. Use detailed pseudocode and a diagram showing a sample decision tree to aid your explanation. [12]

(d) Decision tree learning can be "unstable". Explain what this means and why it happens. Briefly outline a modification that can reduce instability. [3]

[END OF SECTION A]

[PTO]

# SECTION B

**4.** (a) A botanist is working on classifying trees as deciduous (lose their leaves in winter) or not. Describe in detail the process of constructing a learning curve. As part of the description, provide an example of a learning curve, ensuring that axes and all other parts are clearly labelled. In addition, explain for the botanist what can be determined from comparing two learning curves. [7]

(b) Are learning curves applicable to: (1) multi-class classification problems; (2) regression problems; (3) non-probabilistic classifiers? Explain your reasoning. [3]

(c) Referring back to the example of Part (a), describe in detail the process of constructing a ROC curve. Provide an example of a ROC curve, with axes and all other parts of the curve clearly labelled. Explain for the botanist what can be determined from comparing two ROC curves. [8]

(d) Are ROC curves applicable to: (1) multi-class classification problems; (2) regression problems; (3) non-probabilistic classifiers? Explain your reasoning. [3]

(e) Write a short note for the botanist on assembling a training dataset for the task, including what data are needed, the training dataset coverage, and how to know when enough data is collected. [4]

**5.** (a) Explain (in words) with examples what it means for two events A and B to be (i) independent of each other; (ii) conditionally independent of each other given a third event C. In Bayesian network notation, how are each of these two cases represented? [7]

(b) Explain what a joint probability distribution is, including an explanation of elementary propositions and atomic events. How can Bayesian networks produce a more compact representation than a full joint probability distribution? [6]

(c) You are moderating a machine learning discussion forum and wish to classify posts as "On-Topic" or "Off-Topic" based on their subject line. Here is a small training set that you have assembled:

| ***Off-Topic:*** | ***On-Topic:*** |
|---|---|
| "funny cats video" | "classify tree type" |
| "cats jump video" | "classify cats dogs" |
| "funny dogs" | "predict dogs type" |
| "tree cats" | "classify nuts tree" |
| | "predict funny" |

Using a Naïve Bayes approach, compute the probability of the following two messages being On-Topic: (i) "cats jump tree"; (ii) "funny cats". Show all steps in your computation and explain any assumptions you make. You do not need to use smoothing. [12]

[PTO]

**6.** (a) A group of botanists have collected data about the growth rate of trees (measured by the total mass of the tree at 10 years old), and related information about weather conditions and soil properties. They would like to use this to predict the growth rate of future trees. They have asked you to advise on whether they should use: (i) linear regression; (ii) logistic regression; or (iii) polynomial regression. In addition, provide an explanation on what kind of task each of these algorithms is suitable for, what kind of data each one needs. [7]

(b) Describe in detail the batch gradient descent algorithm, as applied to multiple linear regression, including how to initialise values, details of how they are updated, and how to check for convergence. As part of your answer, provide a definition of the empirical error cost function used in this algorithm. [10]

(c) Describe in detail an algorithm to build a linear classifier with multiple inputs, including an explanation of the perceptron learning rule. [8]

[END]