Family Name: .....

Other Names: .....

Signature: .....

ID Number: .....

COMP 307/AIML 420 Test 1

22 April 2021

Instructions

- Time allowed: **50 minutes**.
- Answer **all** the questions. There are 45 marks in total (1 mark per minute plus five minutes spare).
- Write down your answers on the blank pages/your own white paper and hand in all sheets.
- Please write your student ID at the top of all pages you hand in.
- If you think a question is unclear, ask for clarification.
- This test contributes either 25% (COMP 307) or 20% (AIML 420) to your final grade.
- You may use paper translation dictionaries, and non-programmable calculators, or programmable calculators with their memories cleared.
- You may write notes and working on this paper, but make sure your answers are clear.

Questions	**Marks**	
1. Search	[5]	
2. Machine Learning	[15]	
3. Neural Networks	[10]	
4. Evolutionary Computation	[15]	
	TOTAL:	

SPARE PAGE FOR EXTRA ANSWERS

Cross out rough working that you do not want marked.
Specify the question number for work that you do want marked.

Question 1. Search [5 marks]

(a) [2 marks] Explain what it means for a heuristic to be *admissible*, and give an example of an admissible heuristic.

(b) [3 marks] Identify one limitation of local search. Explain how this limitation can be avoided by using another search method.

Question 2. Machine Learning [15 marks]

(a) [4 marks] We often characterise machine learning techniques into sub-categories, such as *supervised* and *unsupervised* learning.

(i) State the main difference between supervised and unsupervised learning.

(ii) Give one example of a real-world application of supervised learning, and one of unsupervised learning.

(b) [5 marks] *Overfitting* is a common problem in machine learning. Explain what overfitting is, and draw a **figure** to help show how it can occur.

(c) [3 marks] Give two reasons for using k-fold cross-validation instead of a single training/test split.

(d) [3 marks] Why do we use **weighted** average impurity in decision tree learning instead of just average impurity? You may find it useful to explain using an example.

Question 3. Neural Networks [10 marks]

(a) [3 marks] Explain the main difference between *online* and *offline* learning in neural networks, and discuss an advantage of each approach.

(b) [7 marks] Taylor wants to design a neural network to predict the sale price of used cars on Trade Me. They have a dataset with six features (e.g. brand, colour, engine size) for a variety of cars.

(i) Explain how many input and output nodes their network should have, and why.

(ii) Describe a method they could use to determine how many hidden layers and nodes to include in their network.

(iii) Taylor trained their network for 10,000 iterations with 20 training instances. Their test performance on 40 test instances was very bad. Suggest three ways that Taylor could improve the test performance of their neural network.

Question 4. Evolutionary Computation [15 marks]

(a) [5 marks] Briefly describe the general evolutionary process in genetic algorithms. You may wish to draw a figure to help your explanation.

(b) [10 marks] Genetic Programming (GP) can be used for a variety of machine learning tasks. Describe how you could use GP to perform binary classification, e.g. classifying animals as either mammal or non-mammal. You should:

(i) Define your terminal set (you may assume your data has five features).

(ii) Define your function set, and describe your reasoning.

(iii) Describe what fitness function you could use, and why.

(iv) List three parameters you would need to set, and what parameter values you would choose.

(v) Describe how the output of the GP tree can be used to determine the class label.

* * * * * * * * * * * * * * *