COMP 307/AIML 420 Test 1 (ONLINE VERSION)

6 May 2022

Instructions

- Time allowed: **45 minutes**.
- Try to answer **all** the questions, even if you are not sure. You will get partial marks for showing your understanding!
- There are **40 marks** in total (1 mark per minute plus five minutes spare).
- Please write your answers in a single document and upload at the end as a `.txt` or `.pdf` file. If you draw figures, please take a photo and upload these as well.
- If you think a question is unclear, ask for clarification.
- This test contributes either 15% (COMP 307) or 10% (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 use the calculator on your computer, but **NOT** a web browser.

Questions	**Marks**	
1. Search	[5]	
2. Machine Learning	[15]	
3. Neural Networks	[8]	
4. Evolutionary Computation	[12]	
	TOTAL:	

Question 1. Search [5 marks]

(a) [3 marks] State whether each of the following statements about search algorithms is *True* (T) or *False* (F).

1. Greedy search is always optimal.
2. An optimal heuristic function never underestimates the cost to the goal.
3. Beam search can produce multiple solutions per run.

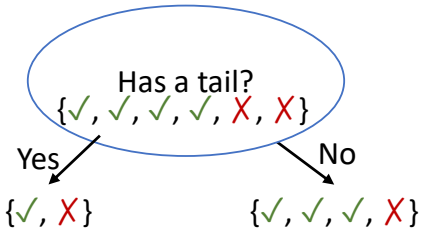
(b) [2 marks] What limitation of *hill climbing* does *simulated annealing* attempt to solve? How does it do this?

Question 2. Machine Learning [15 marks]

(a) [3 marks] Classification and regression are two common supervised learning tasks. State the main difference between classification and regression. Give one example of a real-world classification task, and one of a regression task.

(b) [3 marks] Kerry has an image dataset with 80 photos of cats and dogs. They want to perform k-fold cross-validation to test how well the k-nearest neighbour algorithm can distinguish cats and dogs. Describe the major steps for performing the k-fold cross-validation.

(c) [4 marks] Consider the decision tree below, which classifies animals as either mammal or non-mammal.



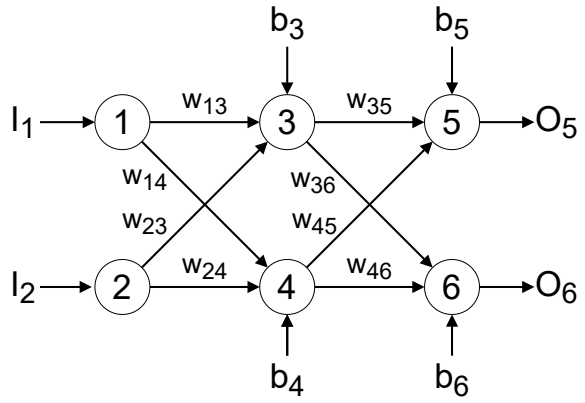
(i) [2 marks] What is the average impurity of the "Has a tail" node? Use the same impurity function from class: `Impurity` $= P(A) \times P(B)$. Show your working. You may leave your answer unsimplified.

(ii) [2 marks] Suppose that there is another feature ("Has a beak") available that instead gives an average impurity of 0.06. Would this be a better or worse feature to choose for this node? Why?

(d) [5 marks] In supervised learning, we want to produce models that are neither *underfitted* nor *overfitted*. Describe your understanding of these two terms and describe an approach you could use to reduce the chance that they occur.

Question 3. Neural Networks [8 marks]

Consider the following feedforward neural network that uses the sigmoid activation function:



$w_{13} = 1.5$	$w_{45} = 0.5$		
$w_{14} = -0.5$	$w_{46} = -1$		
$w_{23} = -2$	$b_3 = 0.5$		
$w_{24} = 1$	$b_4 = -0.5$		
$w_{35} = -1$	$b_5 = 0$		
$w_{36} = 0.5$	$b_6 = 1$		

(a) Given an input vector $(1.0, 1.0)$ (which represents input feature values for I_1 and I_2), what will be the outputs of node 5 (O_5)?

(b) Assume that the training set consists of only a single instance with an input vector $(1.0, 1.0)$, and an output vector $(1.0, 0.0)$ (which represents the target output values for node O_5 and O_6), and that the learning rate η is 0.2.

What will be the new value of the weight w_{13} after one epoch of training using the back propagation algorithm? Show your working.

The appendix at the end of this test paper contains formulae you may find useful.

Question 4. Evolutionary Computation [12 marks]

(a) [6 marks] Briefly describe the general evolutionary process in genetic algorithms. You may wish to draw a figure to help your explanation.

(b) [6 marks] Genetic Programming (GP) is considered a good method for binary classification. Consider using the standard tree-based GP to solve a *binary classification* problem with the following training set (X1 – X5 are the features):

X1	X2	X3	X4	X5	Class
1.2	3.1	9.1	11.3	12.0	O
4.1	7.3	14.2	22.1	2.1	X
0.6	5.3	11.1	13.2	9.8	O
1.5	1.3	7.8	14.1	15.6	O
1.2	8.8	5.5	19.1	7.5	X

(i) Choose a good terminal set.

(ii) Choose a good function set.

(iii) Each evolved program returns a single floating point number. Describe a way to translate the single output value of an evolved program into the two class labels.

(iv) Assuming there are three classes in the above problem, give two possible ways to perform this multi-class classification with GP.

Appendix

A Some Formulae You Might Find Useful

$$f(x_i) = \frac{1}{1+e^{-x_i}} \tag{1}$$

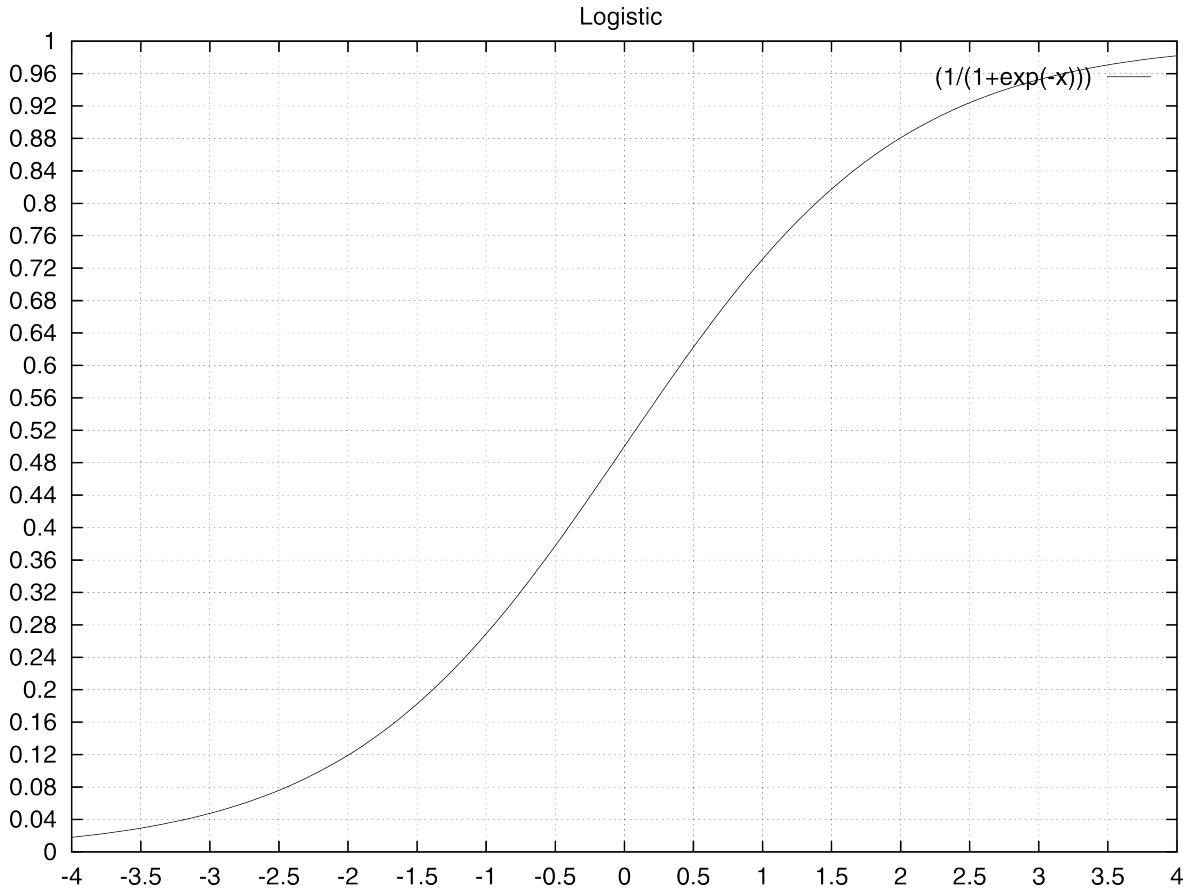
$$O_i = f(I_i) = f(\sum_k w_{k\to i} \cdot o_k + b_i) \tag{2}$$

$$\Delta w_{i\to j} = \eta o_i o_j (1-o_j)\beta_j \tag{3}$$

$$\beta_j = \sum_k w_{j\to k} o_k (1-o_k)\beta_k \text{ (Hidden layers)} \tag{4}$$

$$\beta_z = d_z - o_z \text{ (Output layer)} \tag{5}$$

B Sigmoid/Logistic Function



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