Calculators may be used in this examination provided they are not capable of being used to store alphabetical information other than hexadecimal numbers

UNIVERSITY OF BIRMINGHAM

School of Computer Science

Artificial Intelligence (First Year)

Main Summer Examinations 2019

Time allowed: 2:00

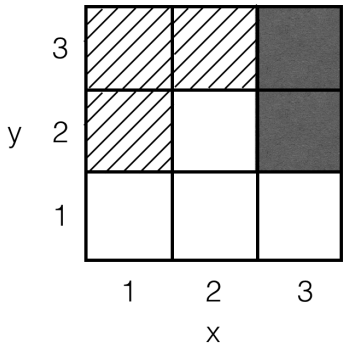
[Answer all questions]

Note

Answer ALL questions. Each question will be marked out of 20. The paper will be marked out of 80, which will be rescaled to a mark out of 100.

Question 1 Search and Optimisation

Assume that you are developing an algorithm to find the lowest cost path to move an army from a starting to a goal position in a strategy game. The field is organised as a grid, where the positions whose coordinates (x, y) are (3,2) and (3,3) have earth terrain; (1,1), (2,1), (3,1) and (2,2) have sand terrain; and (1,2), (1,3) and (2,3) have water terrain:



Each move can take the army one position up, down, left or right on the grid, so long as this does not move the army outside the grid. Independent of the type of terrain of the current position of the army, the cost of moving to an earth, sand and water position is 1, 3 and 10, respectively. For example, the cost of moving left from (3,2) is 3.

(a) Your army is in position (3,1) and you wish to reach the goal position (1,3). Assume that you have decided to use a breadth-first search algorithm to solve this problem, respecting the following rules:

- A state in the state space graph is identified by the (x, y) coordinates of the current position of the army, meaning that there are 9 possible states.
- Do not place children in the frontier if their corresponding state is already in the frontier or list of visited nodes.
- Stop when you place in the frontier a node which contains the goal state.
- When deciding which node to visit, if there is a draw, choose to visit the node with the smallest x-coordinate first. If this still results in a draw, choose to visit the node with the smallest y-coordinate first. For example, if there is a draw between nodes whose states are (3,1) and (2,2), visit (2,2) first. If there is a draw between (3,1) and (3,2), visit (3,1) first.

Question 1 continued over the page

Write down the following information:

- Search tree produced by breadth-first search. Indicate which nodes are visited nodes and which nodes are in the frontier when the algorithm terminates.
- Sequence of nodes *visited* by breadth-first search. Note: you can identify a node through its state.
- Sequence of states that compose the path retrieved as a solution by breadth-first search.

[8 marks]

(b) Is breadth-first search a good algorithm for this problem? Justify your answer.

[6 marks]

(c) Consider that you wish to use Hill Climbing to solve this problem. For that, you need to provide a problem formulation.

- Specify the design variable of your problem formulation.
- Explain your design variable by giving two examples of candidate solutions. If infeasible solutions are possible, give one example of feasible, and one example of infeasible solution, explaining the reason for them to be feasible / infeasible. If no infeasible solutions exist given your problem formulation, give two examples of feasible solutions, and explain why no infeasible solutions exist.

[6 marks]

Question 2 k-Nearest Neighbours and Naïve Bayes

(a) Consider the problem of predicting the best hashtag to be associated to a tweet. To solve this problem, you have access to incoming tweets that can be used as training examples. Each hashtagged tweet is a training example. Every second, on average, around 6,000 tweets are tweeted on Twitter, and most of them use hashtags. Each tweet is described by a set of input attributes and one output attribute. There is one categorical input attribute for each possible word that can appear in a tweet. Each input attribute can assume values *true* or *false*, representing whether or not the corresponding word appears in the tweet. The output attribute is a categorical value corresponding to the first hashtag that appears in the tweet. Other hashtags are ignored.

- List one advantage and one disadvantage of using k-Nearest Neighbours for this specific problem and explain your answer.
- List one advantage and one disadvantage of using Naïve Bayes for this specific problem and explain your answer.

[6 marks]

(b) Assume a much reduced fictitious version of this problem where you have received only six training examples so far and you have only two input attributes. The frequency tables created for Naïve Bayes **without** Laplace Smoothing are shown below, where x_1 and x_2 are the input attributes and y is the output attribute:

Frequency table for x_1	y=birthday	y=contest	y=health	total
x_1 = true	2	2	0	4
x_1 = false	0	0	2	2
total	2	2	2	6

Frequency table for x_2	y=birthday	y=contest	y=health	total
x_2 = true	2	0	1	3
x_2 = false	0	2	1	3
total	2	2	2	6

Question 2 continued over the page

Recall that Naïve Bayes uses the following equation to make predictions:

$$P(c|f_1, \cdots, f_n) = \alpha P(c) \prod_{i=1}^{n} P(f_i|c)$$

where

- $\alpha = 1/\beta$
- $\beta = \sum_{c \in Y} \left(P(c) \prod_{i=1}^{n} P(f_i|c) \right)$
- c is an output value;
- Y is the set of possible output values;
- f_i is the value of input attribute i; and
- n is the number of input attributes.

Show the step-by-step calculations of the probabilities of a new tweet with input attributes ($x_1 =$ true, $x_2 =$ true) having birthday, contest and health as best hashtags, respectively, **using Laplace Smoothing**. What hashtag would Naïve Bayes predict and why?

[8 marks]

(c) Consider now that you wish to predict more than one possible hashtags that could be associated with a given tweet. This means that you can have more than one output value associated to each tweet. Discuss how you would use Naïve Bayes to predict more than one possible hashtag. Your discussion should include the following items:

- How would you create frequency tables?
- How would you use the Naïve Bayes equation to give predictions?

[6 marks]

Question 3 Linear Regression and Logistic Regression

(a) The cost of a hypothesis function parametrised by z is given by the following equation:

$$z^2 - 12z + 2$$

What is the value of the parameter z at which the cost is minimum?

[4 marks]

(b) The cost function for Logistic Regression is given by the following equation:

$$Cost(w) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} log(h_w(x^{(i)})) + (1 - y^{(i)}) log(1 - h_w(x^{(i)}))\right]$$

where w represents the weights of the hypothesis function h, and $y^{(i)}$ and $x^{(i)}$ are the input and output values of a given example.

Derive this expression from the version which shows the cost for each case $y = 0$ and $y = 1$ separately. Additionally, detail why this is a reasonable cost function. You might find it easier to use the separated version to show this by analysing the values of $-log(x)$ and $-log(1 - x)$.

[8 marks]

(c) The Hypothesis function for Univariate Linear Regression is $y = w_0 + w_1 x$. The cost function associated with this hypothesis function h, parametrised by some w_0 and w_1, is $\sum_i (y^{(i)} - h_w(x^{(i)}))^2$, where $y^{(i)}$ and $x^{(i)}$ represent the output and input values of the i$^{\text{th}}$ training example.

(i) Write out and provide an explanation for the general form of the hypothesis function and cost function of Linear Regression in two variables.

(ii) Similarly, write out and provide an explanation for the general form of the hypothesis function for Univariate Non-linear Regression. Assume that the non-linear hypothesis function is a quadratic.

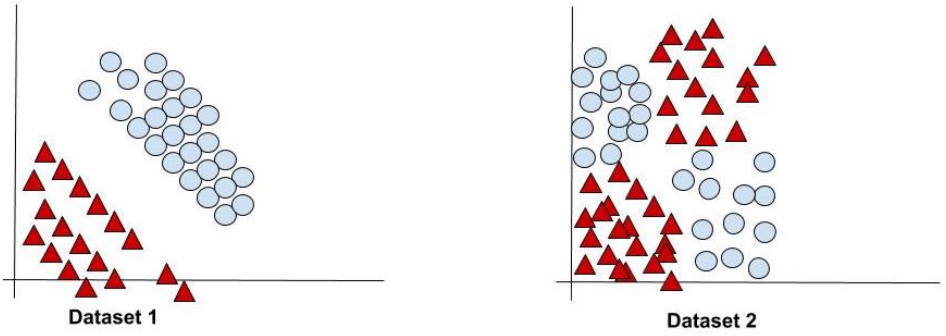
[8 marks]

Question 4 Neural Networks

(a) Describe the meaning of the terms "overfitting" and "underfitting" in the context of neural networks and additionally detail the impact of increasing the number of layers, increasing the regularization and increasing the dropout on these phenomenon.

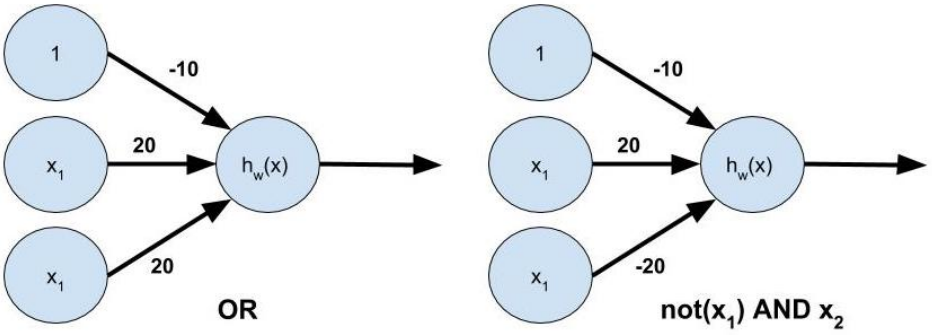
[4 marks]

(b) Consider the graphical illustrations of two datasets below. You are required to train a neural network to differentiate the two different classes. For each dataset, describe the architecture of and provide an illustration of the neural network you will use along with how you will go about training and testing your neural network.



[8 marks]

(c) Consider the following Neural Networks each of which perform the associated logical operation shown in the Truth Table.



Question 4 continued over the page

x_1	x_2	x_1 OR x_2	not(x_1) AND x_2	$x_1 XOR x_2$
0	0	0	0	0
0	1	1	1	1
1	0	1	0	1
1	1	1	0	0

The activation function of the above Neural Network is the sigmoid function given by the equation $\frac{1}{1+e^{-x}}$. You may assume that:

$$sigmoid(x) \approx 0; \text{for } x \leq -4$$
$$sigmoid(x) \approx 1; \text{for } x \geq +4$$

Create a neural network to perform the non-linear separable operation XOR whose Truth Table is given above. Draw a table to illustrate the input and output for each of your nodes that clearly shows why your networks performs this operation. **[8 marks]**

This page intentionally left blank.

Artificial Intelligence (First Year)

Do not complete the attendance slip, fill in the front of the answer book or turn over the question paper until you are told to do so

Important Reminders

- Coats/outwear should be placed in the designated area.
- Unauthorised materials (e.g. notes or Tippex) must be placed in the designated area.
- Check that you do not have any unauthorised materials with you (e.g. in your pockets, pencil case).
- Mobile phones and smart watches **must** be switched off and placed in the designated area or under your desk. They must not be left on your person or in your pockets.
- You are not permitted to use a mobile phone as a clock. If you have difficulty seeing a clock, please alert an Invigilator.
- You are not permitted to have writing on your hand, arm or other body part.
- Check that you do not have writing on your hand, arm or other body part – if you do, you must inform an Invigilator immediately
- Alert an Invigilator immediately if you find any unauthorised item upon you during the examination.

Any students found with non-permitted items upon their person during the examination, or who fail to comply with Examination rules may be subject to Student Conduct procedures.