1	2	3	4	5	6	7	8	total
40	20	20	30	25	35	15	15	210

UMBC CMSC 471 01, Final Exam, 17 May 2019

Name:_____

Write all of your answers on this closed-book exam which has eight problems that add up to 210 points. You have the two hours to work on this exam. There are two blank pages at the end that you can use. Good luck.

1. True/False (40 points) Circle T or F for each statement

T F An optimal solution path for a search problem with only positive costs will never have repeated states.

T F A simple breadth-first search always finds a shortest solution if one exists that is of finite length.

T F Hill climbing search algorithms only work for search spaces that are two-dimensional or have solution-preserving projections onto two-dimensions.

T F An inference procedure for logic that is not sound cannot be complete.

T F There are eight models for the sentence (a ∨ b ∨ c) among all of the interpretations over the three Boolean variables?

T F The Ockham's Razor heuristic prefers the simplest consistent explanation.

T F Every model of (a ^ b) is also a model of (a ∨ b).

T F A propositional sentence is valid, if and only if it is satisfied in all possible models.

T F Solving constraint satisfaction problems with forward-checking and arc consistence algorithms means that backtracking search is not required.

T F A partial order planner is one that always produces a shortest possible plan.

T F When overfitting occurs, a machine learning model describes random error or noise instead of the underlying relationships.

T F The ID3 decision tree induction algorithm uses information gain and is guaranteed to find the optimal decision tree consistent with a given training set.

T F A SVM with a soft margin will allow some positive training examples to be on the negative side of the margin and also some negative training examples to be on the positive side of the margin.

T F In a zero-sum, two player game there is necessarily always a winner and a loser.

T F One drawback of the K-means clustering algorithm is that one needs to specify how many clusters the algorithms should find.

T F The cells in a NumPy array can only hold numbers or strings.

T F Overfitting occurs when a machine learning model over-generalizes from the its training data.

T F The recall metric used in a machine learning binary classification task is defined as the ratio of the number of true positives to the sum of the number of true positives and false negatives.

T F A learning curve in machine learning is a way to evaluate how a system's accuracy varies with the amount of training data.

T F Information gain is used to determine the network structure in Recurrent Neural Network.

2. Multiple choice (20 points: 5, 5, 5, 5)

Circle the letters of **all** of the correct answers for each question.

2.1 What are advantages of using the alpha-beta algorithm over simple mini-max for games?

(a) Alpha-beta can handle games with uncertainty, whereas mini-max cannot.

(b) Alpha-beta does not require a static evaluation function, whereas mini-max always does.

(c) Alpha-beta is more accurate than minimax.

(d) Alpha-beta has a lower worst-case time complexity.

(e) none of the above

2.2 How many models does the propositional sentence $(A \wedge (B \Rightarrow C))$ have?

(a) one

(b) three

(c) five

(d) seven

(e) eight

(f) none of the above

2.3 What are some advantages of a partial order planner (POP) over a state-space planner (SSP)?

(a) A POP can find plans that a SSP will not.

(b) A POP plan can represent sub-plans that can be done in any order

(c) A POP plan is always optimal

(d) A POP plan can have loops where as a SSP cannot.

(e) none of the above

2.4 What are possible advantages using a random forest classifier over a simple decision tree one?

(a) It tends to require less training data

(b) It can handle a much larger number of features than a simple decision tree

(c) It can be more accurate

(d) It is less prone to overfitting the training data

(e) none of the above

3. Short Answers (20 points: 10, 10)

3.1 In a few sentences, describe the approach used in the classic STRIPS planning algorithm.

3.2 Describe one problem that the simple STRIPS planning algorithm has that can cause it to produce a plan that is longer than necessary.

4. Resolution Proof in Propositional Logic (30: 10, 10, 10)

Consider a propositional KB with three variables (P, Q, R) and just two sentences:

- $P \vee Q \Rightarrow R$
- $\sim P \Rightarrow Q$

A propositional sentence is in CNF if it is a set of one or more expressions where each is a disjunction of variables or negated variables

4.1 Construct a KB of propositional sentences using the three propositional variables (P, Q, R) and logical connectives ($\vee$, $\sim$). Encode each of sentences into *one or more* logic sentences in conjunctive normal form (CNF). (10 pts)

#	English statements	CNF clauses (one or more)
0	$P \vee Q \Rightarrow R$	
1	$\sim P \Rightarrow Q$	

4.2 How many models are there for this KB? Recall that a model is an assignment of true and false to each variable such that the KB is consistent. (10 pts)

4.3 Show a resolution refutation proof that R is true given these two sentences. Start with the negation of what's to be proved, add statements from your KB and then use resolution to derive a contradiction ($\perp$). The table to the right shows an example resolution refutation proof (10 pts).

step	action	result
1	assume	$\sim$Q
2	KB	$\sim$P$\vee$Q
3	KB	P
4	resolve 2,3	Q
5	resolve 1,4	$\perp$

Sample proof of **Q** given **P**$\rightarrow$**Q** , **P**

step	action	result
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		

5. English and logic (25 points, 5, 5, 5, 5, 5)

We want to represent data about polygons and some of their sub-types, including triangles, squares and simplePolygons, which includes triangles and squares. For each of the following English statements, write **two different** ways to express it in first order logic. (Hint: one can use ∀ and the other ∃) Use the unary predicates **triangle(x)**, **square(x), simplePolygon(x)** and **polygon(x).** Use the standard notation for first order logic as shown in the table on the right and use parentheses as needed to ensure that the precedence intended is used.

∀	forall
∃	exists
∧	and
∨	or
⇒	implies
⇔	iff
~	not

5.1 All simplePolygons are polygons

5.2 No triangle is a square

5.3 All triangles and squares are simple polygons

5.4 Every simple polygon is either a triangle or a square

5.5 Given these definitions, does it follow that every triangle is a polygon?

6. Decision tree reasoning (35: 10, 15, 10)

The table on the right shows 12 examples of decisions about a credit application (approved or not) based on three variables: credit history, income level and current debt. You plan to use this data to train an ID3-based decision tree to predict if an application will be approved given the values of credit, income and debt.

Predictive features			*target*
credit	**income**	**debt**	**approve**
Good	Low	Low	Yes
Good	Medium	Low	Yes
OK	High	High	Yes
OK	High	High	Yes
OK	Medium	Low	Yes
OK	High	Low	Yes
Bad	High	Low	No
Bad	Medium	Low	No
OK	Low	High	No
OK	Low	High	No
OK	Low	Low	No
OK	Low	Low	No

6.1 What is the initial entropy of the target variable, i.e., E(approve), which is defined as the sum of the probability of each value times the negative of the log base 2 of that probability. Recall that $\log_2$ of ½ is -1.

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

6.2 The ID3 algorithm selects the variable at each level that maximizes the information gained. Which attribute would be chosen as the root of the decision tree? (5 pts)

6.3 Show the entire decision tree that would be constructed by ID3, i.e., by recursively applying information gain to select the roots of sub-trees after the initial decision. If the training data does not determine the predicted value for a case, use a "?" for the decision.(10 pts)

6.3 If we add an additional column to the predictive features which is an applicant's unique application number and then use that in addition to the other predictive features, what impact will it have on the decision tree that is learned by the ID3 algorithm?

7. Perceptron vs. feed forward neural network (15)

Explain in a few sentences the difference between a simple perceptron and a feed forward neural network

8. Batches, Epochs and backpropagation (15)

Two important parameters when training a neural network are **batch size** and **epochs**. Explain what each means and how they affect the backpropagation algorithm when a neural network is trained.

scratch page

scratch page