1	2	3	4	5	6	7	total
20	20	20	30	20	20	25	155

UMBC CMSC 471 01, Final Exam, 21 May 2018

Name:_____

Please write all of your answers on this exam. The exam is closed book and has seven problems that add up to 155 points. You have the two hours to work on this exam. Good luck.

1. True/False (20 points, 1 each) Circle T or F for each statement

T F Every complete inference procedure is also sound.

T F Every sound inference procedure is also complete.

T F A sound but incomplete inference procedure can produce an incorrect answer.

T F An unsound inference procedure can produce an incorrect answer.

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

T F If a propositional sentence is valid, it is satisfied in all possible models.

T F The STRIPS planning algorithm was designed to find the shortest possible plan for a given problem.

T F Regression planners search for plans by starting from a goal state and working back toward the initial state.

T F Decision tree learning can only be used to produce binary classifiers.

T F In a well-formed Bayesian Belief Network, a node is always conditionally independent of its non-descendants given its parents.

T F Information gain is used to determine the network structure in a Bayesian Network.

T F Bayes' rule can be used to relate the probability of a cause given it symptoms to the probability of its symptoms give a cause.

T F An important advantage of support vector machines (SVMs) is that they can directly implement classifiers with a large number of classes.

T F The K-nearest neighbors algorithm is an unsupervised machine learning technique for clustering instances into K clusters where K is a positive, non-zero integer.

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

T F Bagging is a technique used in building Bayesian Belief Networks that helps identify variables that are not statistically independent.

T F Overfitting occurs when a machine learning model was trained with two examples with different labels.

T F The precision metric used in machine learning is defined as the ratio of the number of true positives to the sum of the number of true and false positive.

T F The precision metric used in machine learning is defined as the ratio of the number of true positives to the sum of the number of true and false positive.

T F A learning curve in machine learning shows the trade-off between a system's precision and recall.

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

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

2.1 Bayesian Belief networks …

(a) Have an associated directed, acyclic graphs

(b) Encode conditional independence assertions among random variables

(c) Generally require less storage than the full joint distribution

(d) Make the assumption that all parents of a single child are independent given the child

(e) None of the above

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

(a) one

(b) three

(c) five

(d) seven

(e) eight

(f) None of the above

2.4 Which tasks are appropriate for unsupervised machine learning algorithms like clustering?

(a) predicting if a credit card transaction is fraudulent or legitimate

(b) finding items that are frequently bought together in a convenience store

(c) organizing a person's connections in a social media network into groups

(d) estimating how much a house will cost given its square footage

(e) none of the above

2.5 What are potential advantages of using a random forest classifier over a decision tree classifier?

(a) It tends to give more accurate results

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

(c) It can make decisions based on non-linear combinations of features

(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, 5, 15)

Consider the following propositional variables:
    R - it is raining
    U - you have an umbrella
    W - you get wet
and the following statements, which can be expressed as implications:

    If it is rains, you always have an umbrella (**R** ⇒ **U**)

    If you have an umbrella, you do not get wet (**U** ⇒ ~**W**)

    If it is not raining, you do not get wet (~**R** ⇒ ~**W**)

> 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 (R, U, W) and logical connectives (∧, ∨, ¬, →). Encode each of the three English statements into *one or more* logic sentences in conjunctive normal form (CNF). (10 pts)

#	English statements	CNF clauses (one or more)
0	If it is rains, you always have an umbrella	0.1
		0.2
		0.3
1	If you have an umbrella, you do not get wet	1.1
		1.2
		1.3
2	If it is not raining, you do not get wet	2.1
		2.2
		2.3

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. (5 points)

4.3 Show a resolution refutation proof for *"you will not get wet"*. That is, given your KB, prove that you will not get wet. Start with the negation of what's to be proved, add statements from your KB and then use resolution to derive a contradiction ($\bot$). The table to the right shows an example resolution refutation proof (15 points).

step	action	result
1	assume	$\neg Q$
2	KB	$\neg P \lor Q$
3	KB	P
4	resolve 2,3	Q
5	resolve 1,4	$\bot$

Sample proof of **Q** given **P→Q & P**

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

5. English and logic (20 points, 5, 5, 5, 5)

For each English sentence below, write the **letter or letters** corresponding to **all** of the logic expressions that express its meaning. If none of the logic expressions correspond to the English sentence, write *none* in the box.

5.1 Every person is either male or female

A. ∀x person(x) ⇒ male(x) ∧ female (x)

B. ~ ∃x Person(x) ∧ ~ (male(x) ∨ female(x))

C. ∀x (Person(x) ⇒ male(x)) ∧ (Person(x) ⇒ female(x))

D. ∃x person(x) ⇒ male(x) ∧ female (x)

5.2 No person can be both male and female

A. ∀x person(x) ⇒ male(x) ∨ female (x)

B. ~ ∃x person(x) ∧ ~ (male(x) ∨ female(x))

C. ∀x (person(x) ⇒ male(x)) ∨ (person(x) ⇒ female(x))

D. ~ ∃x person(x) ∧ male(x) ∧ female (x)

5.3 Every person has a mother

A. ∀x ∃y person(x) ∧ hasMother(x, y)

B. ∃y ∀x person(x) ⇒ hasMother(x, y)

C. ∀x ∃y person(x) ⇒ hasMother(x, y)

D. ∃x ∃y person(x) ⇒ hasMother(x, y)

5.4 Good food is not cheap and cheap food is not good

A. ∀x food(x) ⇒ (good(x) ∧ ~cheap(x) v (~ good(x) ∧ cheap(x))

B. ∀x food(x) ⇒ (good(x) ⟺ ~ cheap(x))

C. ∀x food(x) ⇒ (~ good(x) ⟺ cheap(x))

D. food(x) => (good(x) ⇒ ~ cheap(x)) ∧ (cheap(x) ⇒ ~ good(x))

6. Probabilistic Reasoning (20 pts, 2 each)

The table at the right gives the full joint probability distribution for Boolean variables A, B and C. Calculate the following probabilities. Show your work for partial credit.

A	B	C	p(A,B,C)
t	t	t	0.03
t	t	f	0.12
t	f	t	0.17
t	f	f	0.18
f	t	t	0.03
f	t	f	0.12
f	f	t	0.24
f	f	f	0.11

6.1 p(A=f) =

6.2 p(B=t) =

6.3 p(B=t, C=t) =

6.4 p(A=f, C=t) =

6.5 p(A=t | B=t) =

Recall that

$$P(a, b) = P(a|b) * P(b)$$
$$= P(b|a) * P(a)$$

so

$$P(a|b) = P(b|a) * P(a)/P(b)$$

6.6 (C=f | B=t) =

6.7 Are A and B independent of each other? (Y=Yes, N=No):

6.8 Are B and C independent of each other? (Y=Yes, N=No):

6.9 Are B and C conditionally independent given A? (Y=Yes, N=No):

6.10 Are A and C conditionally independent given B? (Y=Yes, N=No):

7. Decision tree classification (25: 5, 5, 5, 10)

NASA has secretly collected aliens and identified then as coming from either Mars (M) or Venus (V). You are charged with building a decision tree classifier to predict the source of new aliens (M or V) as they are found using four predictive features: height, color, number of legs and stench.

predictive features				target
height	**color**	**# legs**	**stench?**	**source**
short	green	3	yes	M
tall	green	3	no	M
short	green	3	no	M
tall	green	3	no	M
tall	red	2	yes	M
tall	green	2	yes	V
short	red	2	no	V
tall	red	3	no	V
short	red	3	no	V
tall	red	3	no	V

7.1 What is the initial entropy of the target variable source, i.e. E(source), which is defined as the sum of the probability of each value times the negative of the log base 2 of that probability. (5)

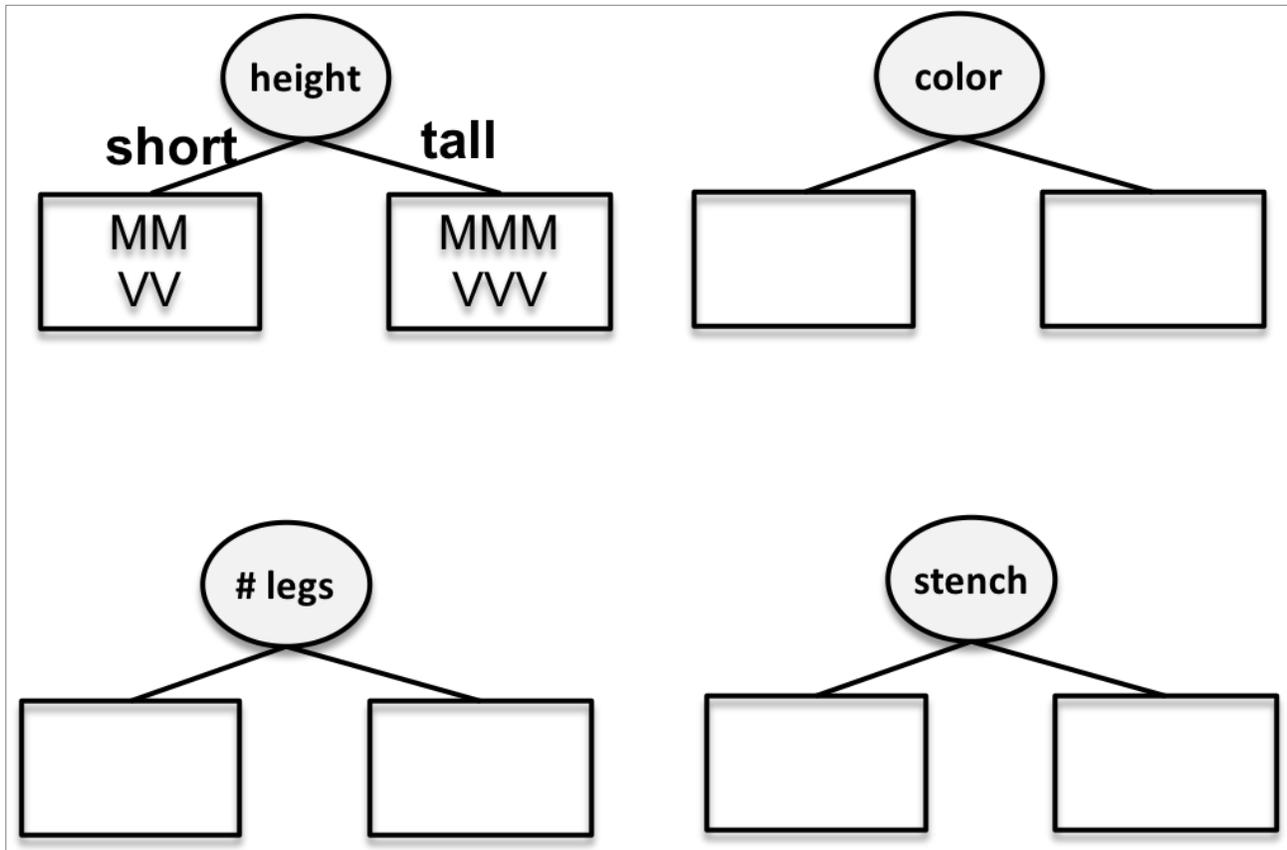$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

p	log$_2$(p)
0.5	-1.0

7.2 Fill in the diagrams on the next page that show the four possibilities of an initial split on each of the four variables. For each one, (1) label the arcs from the root node to the boxes and (2) indicate the source values that go in each box. We've done the first one for you. (5)

7.3 Circle the variable that an ID3-type decision tree algorithm will choose for the initial split. (5)

Height   **Color**   **# Legs**   **Stench**

7.4 Using your answer in 7.3, draw **the entire** decision tree on the next page that an ID3-type algorithm that uses information gain to choose splits will produce. (10)

```
        ( height )                              ( color )
  short        tall
┌─────────┐  ┌─────────┐     ┌─────────┐    ┌─────────┐
│   MM    │  │   MMM   │     │         │    │         │
│   VV    │  │   VVV   │     │         │    │         │
└─────────┘  └─────────┘     └─────────┘    └─────────┘


        ( # legs )                             ( stench )

┌─────────┐  ┌─────────┐     ┌─────────┐    ┌─────────┐
│         │  │         │     │         │    │         │
│         │  │         │     │         │    │         │
└─────────┘  └─────────┘     └─────────┘    └─────────┘
```

Complete the diagram above for question 7.3 and draw your complete decision tree below for 7.4