1	2	3	4	5	6	total
20	*30*	*10*	*10*	*20*	*10*	*100*

Name:____

UMBC CMSC 471 Final Exam, 16 May 2016

Please write all of your answers on this exam. The exam is closed book and has seven problems that add up to 100 points. You have the two hours to work on this exam. Good luck.

1. True/False (20 points)

T F The resolution inference rule is very general and can used for abductive reasoning. False

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

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

T F Every sound reasoner is complete and every complete reasoner is sound. False

T F One problem with the STRIPS planner is that it can undo the completion of one sub-goal in the course of achieving a subsequent sub-goal. True

T F Every plan found by a partial order planner can be converted into at least one linear plan. True

T F Two random Boolean variables are independent if and only if the probability that both are true is equal to the product of the probabilities that each is true. True

T F A well-formed Bayesian belief network can contain at most one cycle. False

T F A Naïve Bayes classifier assumes that all features are independent. True

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. True

T F A node in a Bayesian belief network can have two parents, but only one child. False

T F Knowing the value of a variable in a Bayesian network tells you something about the likelihood of the value of every variable that is not independent of it. True

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. False

T F Supervised learning can only be done with a set of training examples for which the "right answer" is known. True

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

T F Clustering algorithms are examples of unsupervised learning. True

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

T F An advantage of a decision tree model is that we can produce a rule-based classifier from it. True

T F Regression is a machine learning technique that can learn to predict a variable whose values are real numbers. True

T F K-means clustering works with data instances that are vectors of length K. False

2. Resolution Proof in Propositional Logic (30 points)

Consider a game where we toss a coin and the outcome is determined by the rule "Heads I win, tails you lose". Model this with four Boolean variables where two represent the outcome of the coin toss: ***Heads*** (the coin was heads up) and ***Tails*** (the coin was tails up) and two describe the result: ***IWin*** (I won the game) and ***YouLose*** (you lose the game).

(a) Construct a KB of propositional sentences using the four variables and logical connectives (**∧, ∨, ¬, →**). Encode each in CNF. We've done one for you. (10 pts)

English	Conjunctive normal form
Either heads or tails is true	**Heads ∨ Tails**
Either heads or tails is False	¬Heads ∨ ¬Tails
I win if heads is true	¬Heads ∨ IWin
you lose if tails is true	¬Tails ∨ YouLose
if you lose, I win	¬YouLose ∨ IWin
If I win, you lose	¬IWin ∨ YouLose

Recall that a propositional sentence is in CNF (conjunctive normal form) if it is a set of one or more expressions where each is a disjunction of variables or negated variables.

(b) How many models are there for this KB, where a model is an assignment of true and false to each variable such that the KB is consistent? (5 points)

2

(c) Show a resolution refutation proof for **IWin** given your KB. Start with the negation of what's to be proved, add statements from your KB and use resolution to derive a contradiction (⊥). The table to the right shows an example proof (15 points).

step	action	result
1	assume	¬IWin
2	given	¬Heads ∨ IWin
3	Resolve 2, 3	¬Heads
4	given	Heads ∨ Tails
5	Resolve 3, 4	Tails
6	given	¬Tails ∨ YouLose
7	Resolve 5, 6	YouLose
8	given	¬YouLose ∨ IWin
9	Resolve 7, 8	IWin
10	Resolve 1, 9	⊥
11		
12		
13		

step	action	result
1	assume	¬Q
2	given	¬P∨Q
3	given	P
4	resolve 2,3	Q
5	resolve 1,4	⊥

Sample proof of Q given P→Q and P

3. Planning (10 points)

3.1 Briefly describe the differences and similarities between problem solving as search (e.g., A*) and planning. (5 pts)

Both problem solver and planner are concerned with getting from a start state to a goal using a set of defined operations or actions, typically in a deterministic, discrete, observable environment.

In planning, however, we open up the representation of states, goals, and plans, which allows for a wider variety of algorithms that decompose the search space, search forwards or backwards, and use automated generation of heuristic functions.

Thus the search can proceeds through a space of possible plans rather than possible states. This allows subgoals to be planned independently, reducing the complexity of the planning problem

3.2 Describe the differences between a simple linear planner like STRIPS and non-linear partial-order planning. Mentions advantages of each approach. (5 pts)

A linear planner (e.g., STRIPS) solves a problem with a list of goals by finding a sub-plan to achieve each goal goal sequentially and concatenates the sub-plans to produce a single linear sequence of actions. Advantages are that the approach is simple and can be very efficient if the sub-plans for achieving the goals do not interact with one another. Disadvantages include the possibility that the plan found will be suboptimal if the sub-plans interfere with one another and that, while a linear planner is sound (it only finds valid plans), it is incomplete (there are some planning problems that have solutions that will not be found).

A non-linear, partial order planner treats a goal as a set of conditions and can explore the space of plans by interleaving work on each goal in the set. The planner can thus work forward from the initial state and also backward from the desired final state. Moreover, the plan can be represented as a partial ordering of actions. Advantages are that partial order planning can be complete, it allows for algorithms that can produce an optimal (i.e., shorted length) plan, and that the plan representations can be more flexible since they need not be reduced to a total ordering. Disadvantages are the larger size of the search space and the additional complexity of the algorithms. For simple problem domains, a straightforward linear planner may be preferable.

4. Using Bayes' rule (10 points)

Based on our records, the probability of acing the 471 final is 0.1, the probability of getting an A in 471 is 0.4, and the probability of a student getting an A in 471 given she aces the final is 0.96. What is the probability of a student acing the 471 final given she gets an A in the class? Show your work.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

Bayes's Rule

P(ace) = 0.1
P(A) = 0.4
P(A|ace) = 0.96

P(ace|A) = 0.24

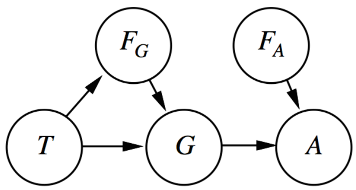
p(A | ace) * p(ace) = p(ace | A) * p(A) = p(ace) * p(A)

p(ace | A) = p(A | ace) * p(ace) / p(A)
= 0.96 * 0.1 / 0.4
= 0.24

5. Bayesian Belief Networks (20 points)

UMBC's nuclear power plant has a gauge that measures the temperature of the core and an alarm that sounds when the gauge reading is too high. Consider Boolean variables A (alarm sounds), Fa (alarm is faulty), and Fg (gauge is faulty) and the multivalued nodes G (gauge reading: normal or high) and T (actual core temperature: normal or high). The gage is nore likely to be faulty when the core temperature is high.

5.1 Draw a Bayesian network for this domain with the five variables as node and a directed edge between two nodes where the first has a direct influence on the second.



5.2 Is your network a polytree, i.e., a network in which there is at most one undirected path between any two nodes. Why or why not?

No. The network is not a polytree since the sub-network {T,Fg,G} there are multiple paths between T and G.

5.3 Assume that the probability that the gauge gives the correct reading (i.e., normal or high) is *0.9* when it is working (Fg=0) and *0.1* when it is not (Fg=1). Give the conditional probability table for G's reading (i.e., normal or high).

	T=normal		T=high	
	Fg=1	Fg=0	Fg=1	Fg=0
G=normal	**0.1**	**0.9**	**0.9**	**0.1**
G=high	**0.9**	**0.1**	**0.1**	**0.9**

5.4 Assume the alarm works correctly unless it is faulty, in which case it never sounds. Give the conditional probability table associated with A.

	G=normal		G=high	
	Fa=1	Fa=0	Fa=1	Fa=0
A=1	**0**	**0**	**0**	**1**
A=0	**1**	**1**	**1**	**0**

6. Decision trees (10 points)

Consider the following data set with three binary input attributes (A_1, A_2, and A_3) and one binary output, y.

instance	A_1	A_2	A_3	y
1	1	0	0	0
2	1	0	1	0
3	0	1	0	0
4	1	1	1	1
5	1	1	0	1

Construct a decision tree to predict y given the inputs from this data using the ID3 algorithm that selects the variable at each level that maximizes the information gained.

7.1 What is the variable at the root of the tree. (5 pts)

A_2

7.2 Show the entire decision tree. (5 pts)