C1 Artificial Intelligence (25 points)
Machine Learning

Part I. Suppose that $X_1, ..., X_m$ are categorical input attributes and Y is categorical output attribute. Suppose we plan to learn a decision tree without pruning, using the standard algorithm. *For full credit, please justify your answer.*

a. (4 points) (True or False): If the Information Gain $IG(Y|X_i) = 0$ when splitting on X_i at the root node, then X_i will not appear in the decision tree.

b. (4 points) (True or False): The maximum depth of the decision tree must be less than m+1 (where m is the number of categorical input attributes).

c. (4 points) (True or False): Suppose data has R records, the maximum depth of the decision tree must be less than $1 + \log_2 R$

d. (4 points) (True or False): Suppose one of the attributes has R distinct values, and it has a unique value in each record. Then the decision tree will certainly have depth 0 or 1 (i.e. will be a single node, or else a root node directly connected to a set of leaves)

Part II. For this part, you're going to answer a couple questions regarding the dataset shown below. You'll be trying to determine whether Little Fred finds a particular type of food appealing based on the food's temperature, taste, and size. You only need to write out the equations.

Appealing	Temperature	Taste	Size
No	Hot	Salty	Small
No	Cold	Sweet	Large
No	Cold	Sweet	Large
Yes	Cold	Sour	Small
Yes	H	Sour	Small
No	H	Salty	Large
Yes	H	Sour	Large
Yes	Cold	Sweet	Small
Yes	Cold	Sweet	Small
No	H	Salty	Large

a. (3 points) What is the initial entropy of Appealing?

b. (3 points) Assume that Taste is chosen for the root of the decision tree. What is the information gain associated with this attribute?

c. (3 points) Assuming again that Taste is chosen for the root, draw the full decision tree learned for this data (without any pruning). Show your work.

C2 Artificial Intelligence (25 Points)
Search

A Mars rover has to leave the lander, collect rock samples from three places (in any order), and return to the lander.

Assume that it has a navigation module that can take it directly to places of interest. So it has primitive actions *go-to-lander*, *go-to-communications-location*, *go-to-rock-1*, *go-to-rock-2*, and *go-to-rock-3*.

We know the time it takes to traverse between each pair of special locations. Our goal is to find a sequence of actions that will perform this task in the shortest amount of time.

(a) (5 points) Formulate this problem as a problem-solving problem by specifying the state representation, initial state, path-cost function, and goal-test. Assume that the world dynamics are deterministic.

(b) (4 points) Say what search technique would be most appropriate, and why. If your search technique requires a heuristic, give an appropriate one.

(c) (4 points) Now assume that, in addition, on every day that it is not at the lander, the rover needs to be at a special communications location at 3PM. Any plan that misses a communications rendezvouz is unsatisfactory.

How would you modify the search space, path-cost function, and/or goal-test to handle this additional requirement? Assume that there is an additional action, called *communicate*, which can only be executed when the rover is at the communications location, and which waits until 3PM, and then executes a communications sequence that takes an hour.

(d) (5 points) Let's go back to thinking about the original version of the problem (without the communications requirement). What, if any, would be the advantages of treating it as a STRIPS planning problem? What, if any, would be the disadvantages?

(e) (7 points) The **heuristic path algorithm** is a best-first search in which the objective function is `f(n) = (2 - w)*g(n) + w*h(n)`. If `w` is limited to non-negative integer values, for what values of `w` is this algorithm guaranteed to be optimal? (You may assume that h is admissible.) Justify your response.

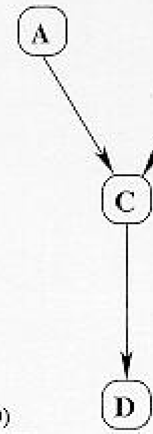
C3: Artificial Intelligence (25 points)

Once we reach a level S_i in a Graphplan planning graph where none of the propositions in our goal are mutex with one another, must a plan exist that can be executed in at most i time-periods (with perhaps some actions being executed in parallel)? If your answer is *Yes*, then prove it. If your answer is *No*, then give a counterexample. (If you give a counterexample, you should provide a goal, a Strips-type description of a set of actions, a planning graph where the propositions in the goal are not mutex at level S_i, and then show that the goal cannot really be achieved within i time periods.)

C4: Artificial Intelligence (25 points)

Consider the following Bayesian network in which C and D are Boolean variables, variable A can have the value 0, 1, or 2, and variable B can have the value 8, 9, or 10.

A	P(A)
0	.4
1	.5
2	.1



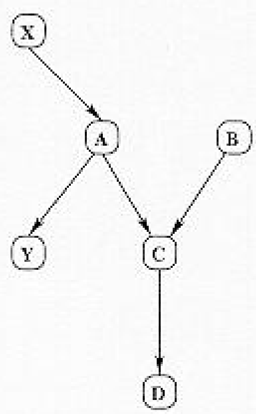
B	P(B)
8	.1
9	.1
10	.8

A	B	P(C)
0	8	.05
0	9	.10
0	10	.30
1	8	.02
1	9	.03
1	10	.15
2	8	.25
2	9	.04
2	10	.16

C	P(D)
t	.3
f	.5

For the following problems, you need not actually do the computations; it will be sufficient to give an equation and fill in the values, such as Z = 3*.998 + 4*.86.

1. (9 points) Compute P(D) — that is the probability that D has the value **true**. You must show the steps that you are taking.
2. (10 points) Compute P(A=0 | D) — that is, the probability that A has the value 0 given that D has the value **true**. You must show the steps that you are taking.
3. (6 points) Suppose that the Bayesian network contains additional causal links from X to A and from A to Y as shown below:



   (a) Suppose that we know the value of D. What nodes have their probabilities altered if we are then told the value of X? Explain why.

   (b) Suppose instead that we know the value of A. What nodes have their probabilities altered