6.034 Midterm Quiz **Solutions**, Spring 2012

March 16, 2012

Name:

1 (25pts)	
2 (25pts)	
3 (25pts)	
4 (25pts)	
Total	

1 Propositional Entailment (25 points)

Circle the numbers of ALL the formulas for which the indicated entailment holds.

Please make sure that we can figure out which ones you mean to circle, if there's any possibility of confusion because of crossing out, write out clearly a list of the ones you meant to circle.

1. $(p \rightarrow q) \wedge (\neg p) \models \neg q$ **False**
2. $(p \vee q) \wedge (\neg q) \models p$ **True**
3. $((p \vee q) \rightarrow r) \wedge (p \rightarrow (\neg q)) \wedge r \models q$ **False**
4. $(p \rightarrow q) \wedge (p \leftrightarrow r) \wedge (r) \models q$ **True**
5. $(p \leftrightarrow q) \wedge (q) \models p$ **True**
6. $(p \rightarrow q) \wedge (\neg q) \models \neg p$ **True**
7. $(p \vee \neg q) \wedge (\neg q) \models \neg p$ **False**
8. $((p \wedge q) \rightarrow r) \wedge (\neg r) \models (\neg p \vee \neg q)$ **True**
9. $(p \rightarrow q) \wedge (q) \models p$ **False**

2 Propositional Proof Methods (25 points)

Here are some simple propositional statements:

- if Man then Mortal
- if Mortal then Boring
- not (Man and Boring)

From these statements we want to prove "not Man".
You can use the following for your formulas, A for Man, O for Mortal, B for Boring.

1. Carry out a resolution refutation proof. Label your clauses with numbers and indicate which clauses are being used at each step.

   1.$\neg A \vee O$
   2.$\neg O \vee B$
   3.$\neg A \vee \neg B$
   4.A – negation of goal
   5.O – resolve 1 and 4
   6.B – resolve 2 and 5
   6.$\neg A$ – resolve 3 and 6
   7.$Contradiction$ – resolve 4 and 7

2. Explain briefly how you can use an algorithm for checking propositional satisfiability as an alternative way of proving "not Man" from the input statements. Write down in detail the propositional sentence involved in this approach to finding a proof.

   Test the conjunction of the first 4 clauses above for satisfiability. If that sentence is unsatisfiable, that constitutes a proof by contradiction.

3. Indicate whether you should use DPLL or WalkSAT for this application. Explain briefly why.

   DPLL, since you need to show it is unsatisfiable. WalkSAT is good for finding satisfying assignments not showing that there are none.

4. What is the size of the complete DPLL search space for this proof? Explain.

   There are 3 binary variables: A, O, B. So, there are 8 states.

5. Would DPLL need to examine every state in the search space for this problem? Explain briefly, based on the operation of the algorithm, why or why not.

   No, A is a unit clause and that fixes that variable to be True.

6. A typical Minesweeper game agent does two "proofs" per square: calling DPLL to find models for:

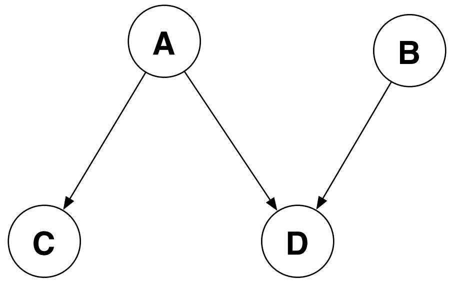
   (a) $Board \wedge M_{i,j}$, and
   (b) $Board \wedge \neg M_{i,j}$

where $Board$ is the formula encoding the board and $M_{i,j}$ represents whether there is a mine at that square. There are four possible outcomes to these two calls: (Model, Model), (Model, None), (None, Model) and (None, None). Explain the meaning of each of these outcomes and what it means for game play.

(Model, Model) means both possibilities are possible; (Model, None) means that we know that $M_{i,j}$ must be true on this board, so we can mark the mine; (None, Model) means that we know that $\neg M_{i,j}$ must be true on this board, so we can probe it; (None, None) should never happen.

3 Bayes Nets (25 points)

Here is a Bayesian Net involving four variables.



A has two values (a_1, a_2), B has three values (b_1, b_2, b_3), C has two values (c_1 and c_2) and D has two values (d_1 and d_2).

1. What is the number of (non-redundant) probability values that need to be specified at each node of this network? What is the total number for the whole network?

   11 $P(A) = 1 + P(B) = 2 + P(C|A) = 2 + P(D|A, B) = 6$

2. Suppose you find out that $D = d_1$, write a formula for the probability distribution over C, given this, that is $P(C|D = d_1)$. The formula should only use expressions whose values can be read off from the CPTs in the network.

   $P(C|D = d_1) = Normalize[\sum_{a_i} P(A = a_i)P(C|A = a_i) \sum_{b_j} P(B = b_j)P(D = d_1|A = a_i, B = b_j)]$

3. If we knew nothing else (ignore previous question), could learning that $C = c_1$ affect the probability of $B = b_2$? Explain.

   No, they are d-separated since we don't know D and thus that triple is not active.

4. If we knew $A = a_1$ (and nothing else), could learning that $C = c_1$ affect the probability of $B = b_2$? Explain.

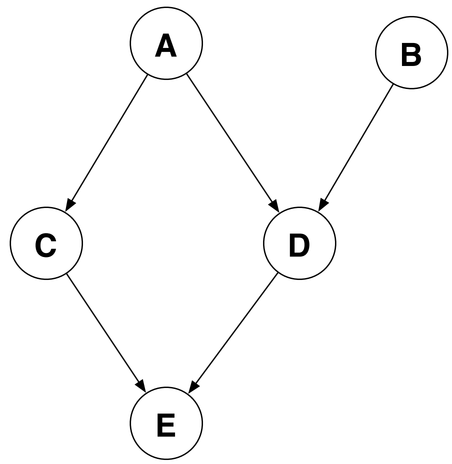
   No, same reason as above.

5. If we knew $D = d_1$ (and nothing else), could learning that $C = c_1$ affect the probability of $B = b_2$? Explain.

   Yes, all the triples are active.

4 Bayes Net Inference (25 points)

Here is a Bayesian Net involving five variables.



Assume each variable X has two values, x_1 and x_2, e.g. a variable A has values a_1 and a_2. In your answers below, specify a factor by writing down the variables involved in the factor, if evidence is involved, use the value of the evidence instead of the variable name, e.g. $f_1(A, E)$ or $f_3(E, C, d_1)$.

1. Show how $P(B)$ is computed via the Variable Elimination algorithm using the variable order: A, E, C, D. Show the sequence of new factors created by the algorithm; show which factors are being multiplied and summed over to create the new one. Identify each new factor created by a unique name, such as $f_5(A, E)$.

```
initial factors: fA(A), fB(B), fC(A,C), fD(A,B,D), fE(C,D,E)
eliminate A: fA(A) * fC(A,C) * fD(A,B,D), sum out A, -> f1(B,C,D)
eliminate E: fE(C,D,E), sum out E, -> f2(C,D)
eliminate C: f1(B,C,D) * f2(C,D), sum out C, -> f3(B,D)
eliminate D: f3(B,D), sum out D, -> f4(B)
fB(B) * f4(B), normalize to get P(B)
```

2. What is the largest factor created during the computation and how big is it (in terms of assignments)? Give the size before the call to `sumOut`.

   The largest factor is fA*fC*fD in the first step (eliminating A), it depends on A,B,C,D, so, $2^4 = 16$ assignments.

3. How big is the biggest factor if we used the variable order D, C, A, E? Hint: You don't need to do the full VE process to answer this.

   The largest factor is fD*fE, which depends on all 5 variables, so $2^5 = 32$ assignments.

4. If you wanted to compute $P(B|C = c_1)$ instead, how would the VE computation you showed in part 1 change? Be specific.

   In fC and fE, we would eliminate all the rows where C is not equal to c_1. Also we would not need to eliminate C.

5. If you are doing likelihood weighting to compute $P(E = e_1|C = c_1, A = a_2)$, what is a formula for the weight that you have to assign to the sample $(a_2, b_1, c_1, d_2, e_1)$? Write the formula in terms of the network CPTs.

   $P(C = c_1, A = a_2) = P(C = c_1|A = a_2)P(A = a_2)$

6. Assume we are using a very large Bayesian Net that represents the connections between hundreds of diseases and hundreds of symptoms. The disease nodes are at the "top" of the network and have no parents. The network has many intermediate nodes (between diseases and symptoms) that represent the states of internal organs, e.g. kidneys and lungs. The symptom nodes are at the bottom of the network. A given symptom may ultimately be caused by many diseases. (Let's take this structure for granted and not worry whether it's the best way to represent the domain.)

   We want to use sampling to estimate conditional probabilities; in particular, likelihood weighting. Indicate the strengths (if any) and weaknesses (if any) you see with that algorithm for the particular queries indicated below.

   (a) $P(symptom_i|disease_j)$

   Likelihood weighting works well here since evidence is at the top of net, so all the samples are generated taking into account the evidence. It doesn't matter if the evidence is unlikely, all the samples would have the same weight, so we could get a decent estimate with relatively few samples.

   (b) $P(disease_j|symptom_i)$

   Likelihood weighting does not work well here; the evidence is at the bottom of the net so all the samples are generated without taking into account the evidence, so most of the samples will have relatively low weight and you have to wait around to get enough samples that are compatible with the evidence.

   (c) When (if ever) would you use rejection sampling instead of likelihood weighting? Explain.

   There really is not much reason to use rejection sampling except for the fact that it might be slightly faster to run but that's only an advantage if it doesn't have to generate many more samples. The only time rejection sampling is competitive is when the evidence is very likely.