6.825 Midterm Quiz, Fall 2005

October 28, 2005

Name: Dan Ports

The exam is on the long side. Do not write long explanations; be brief. The fact that there is a lot of white space does not mean that we expect you to fill it. If you don't immediately see how to do one question, move to another one and come back later.

1 (15pts)	15
2 (25pts)	~~24~~ 25
3 (20pts)	19
4 (20pts)	17
5 (20pts)	17
Total	93

1 Propositional Logic (15 points)

Recall the first proof from Project 1b:

- All men are mortal
- All mortals are boring
- Hera is not boring.
- Therefore, there exists someone who is not a man

The predicates required for this proof are $Man(x)$, $Mortal(x)$ and $Boring(x)$.

1. (3 pts) Explain briefly how you can use an algorithm for checking propositional satisfiability to carry out this proof.

Reduce the universe to a finite set (e.g. just Hera), replace universal quantifiers with an AND over all elements in the universe and existential quantifiers with an OR. Negate the conclusion and use a sound & complete SAT solver to show that there are no satisfying assignments.

3/3

2. (2 pts) Indicate whether you should use DPLL or WalkSAT for this application. Explain briefly why.

We need to show there are no satisfying assignments so we need a complete SAT solver like DPLL. WalkSAT is not complete, so we don't know that there actually aren't satisfying assignments if it doesn't find one.

2/2

3. (5 pts) Write down in detail the propositional sentence involved in this approach to finding a proof.

$(Man(Hera) \Rightarrow Mortal(Hera)) \wedge$
$(Mortal(Hera) \Rightarrow Boring(Hera)) \wedge$
$(\neg Boring(Hera)) \wedge$
$\neg(\neg Man(Hera))$

5/5

4. (2 pts) What is the size of the complete search space for this proof?

Three predicates (Man, Mortal, Boring) and one instance, so $2^3 = 8$ possibilities.

2/2

5. (3 pts) Would DPLL need to examine every state in the search space for this problem? Explain briefly.

No, it uses logic such as the unit clause rule to prune the search space. But still worst-case exponential.

3/3

得 25

2 First Order Logic (25 points)

Here are two sentences in FOL (note the placement of parentheses):

$$(\forall x.P(x) \Rightarrow (\exists y.Q(y)))$$
$$(\forall x.P(x)) \Rightarrow (\exists y.Q(y))$$

$\forall x\, (\neg P(x)\ \vee (\exists y\ Q(y)))$ (1)
$\exists x\ \neg P(x)\ \vee \exists y\, (Q(y))$ (2)

1. (5 pts) Assume the universe is $U = \{A, B\}$, for each of the interpretations below, indicate which sentence is true and which is false in that interpretation.

P(A)	P(B)	Q(A)	Q(B)	Sentence 1 is T/F?	Sentence 2 is T/F?
F	F	F	F	T	T
T	T	T	F	T	T
T	T	F	F	F	F
F	T	T	F	T	T
F	T	F	F	F	T

5

2. (2 pts) Does the first sentence logically entail the second? Explain briefly.

2

答案 everywhere 1 is true 2 is true

Yes. We can write (1) as $(\forall x\ \neg P(x)) \vee (\exists y\ Q(y))$ and (2) as $(\exists x\ \neg P(x)) \vee (\exists y\ Q(y))$, the first is clearly stronger than the second because the only difference is the quantifier.

3. (2 pts) Does the second sentence logically entail the first? Explain briefly.

2

No. Sentence 1 is not true under the last case above, and Sentence 2 is, so the set of interpretations where (1) holds is smaller than that for (2).

8

4. (8 pts) Using resolution refutation, try to show that **given the first sentence you can prove the second**. Be very clear about what clauses you are using in the proof. Specify which clauses you are using at each step and the unifier, if any. Be sure to say whether you think the proof succeeds or fails.

   Here are the two sentences again:

$$(\forall x.P(x) \Rightarrow (\exists y.Q(y)))$$
$$(\forall x.P(x)) \Rightarrow (\exists y.Q(y))$$

Convert (1) to clause form:

$\forall x(\neg P(x) \lor \exists y\ Q(y))$

$\neg P(x) \lor Q(F(x))$ (A1)

Negate (2) and convert to clause form:

$\neg((\forall x\ P(x)) \Rightarrow (\exists y\ Q(y)))$

$\neg(\neg(\forall x\ P(x)) \lor (\exists y\ Q(y)))$

$\forall x\ P(x) \land \neg(\exists y\ (Q(y)))$

$\forall x\ P(x) \land \forall y(\neg Q(y))$

$P(x) \land \neg Q(y)$

$P(x)$ (A2)

$\neg Q(y)$ (A3)

Proof follows:

$Q(F(x_1))$ - resolve (A1) w/ (A2), $\theta = \{x_1/x_2\}$ (4)

F - resolve (A3) w/ (4) $\theta = \{F(x_1)/y\}$ (5)

Proof succeeds

8

5. (8 pts) Using resolution refutation, try to show that **given the second sentence you can prove the first**. Be very clear about what clauses you are using in the proof. Specify which clauses you are using at each step and the unifier, if any. Be sure to say whether you think the proof succeeds or fails.

Here are the two sentences again:

$$(\forall x.P(x) \Rightarrow (\exists y.Q(y)))$$
$$(\forall x.P(x)) \Rightarrow (\exists y.Q(y))$$

Convert (2) to CNF:

$\neg(\forall x\ P(x)) \vee (\exists y (Q(y)))$

$\exists x\ \neg P(x) \vee \exists y\ Q(y)$

$\neg P(foo) \vee Q(bar)$ (A1)

Negate (1) and convert to CNF:

$\neg(\forall x\ P(x) \Rightarrow (\exists y\ Q(y)))$

$\neg(\forall x\ \neg P(x) \vee (\exists y\ Q(y)))$

$\exists x\ P(x) \wedge \forall y \neg Q(y)$

$P(baz)$ (A_2)

$\neg Q(y)$ (A_3)

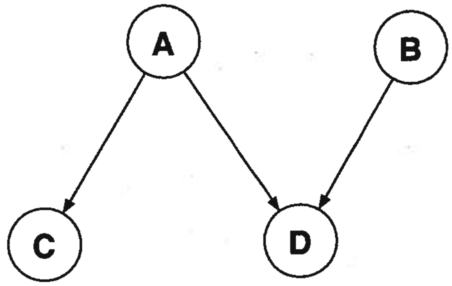
Proof:

$\neg P(foo)$ — resolve (A1) w/ (A_3), $\theta = \{y / bar\}$ (4)

Proof fails. No further resolution steps can be performed

3 Bayesian Nets (20 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. (5 pts) 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?

5

A: 1 value
B: 2 values
C: 2 values

D: 2 · 3 = 6 Values

Total: 11 Values

2. (5 pts) Suppose you find out that $D = d_1$, write a formula for the probability distribution over C, given this. Show the formula so that it involves the minimal amount of computation.

4

$$Pr(C \mid D = d_1) = \alpha \, Pr(C, d_1) =$$

$$= \alpha \sum_A Pr(C \mid A) Pr(A) \sum_B Pr(d_1 \mid A, B) \; Pr(B)$$

3. (5 pts) If we knew nothing else (ignore previous question), could learning that $C = c_1$ affect the probability of $B = b_2$? Explain.

5

No, Information cannot flow to B via the common effect pattern $A \searrow D \swarrow B$ without knowing D.

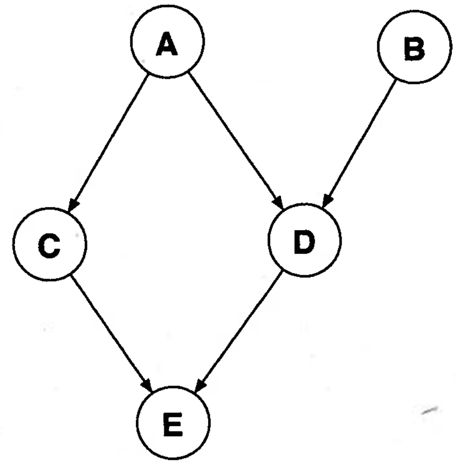
4. (5 pts) If we knew $D = d_1$ (and nothing else), could learning that $C = c_1$ affect the probability of $B = b_2$? Explain.

5

Yes, Information can flow $C \to A \to D$ since A is a common cause of C and D, and A is not known, and it can flow $A \to D \to B$ since D is a common effect, and D is known.

4 Bayesian Net Inference (20 points)

Here is a Bayesian Net involving five variables.



Factors:

~~P(A)~~
P(B)
~~P(C|A)~~
~~P(D|A,B)~~
~~P(E|C,D)~~,

~~P(A)~~
τ(B)
P(C|A)
~~p(D|AB)~~
~~p(E|C,...)~~

Assume each variable X has two values, x_1 and x_2.

1. (5 pts) Show how $P(B)$ is computed via the Variable Elimination algorithm using the variable order: A, E, C, D. Show the intermediate factors created by the algorithm as described in class.

~~D is eliminated, giving the factor $\tau_1(A,B,C,E) = \sum_{A,B,C,E} P(D|A,B)\, P(E|C,D)\, P(B)\, P(A)$~~

5/5

A is eliminated, giving the factor

$\tau_1(B,C,D) = \sum_A P(A)\, P(D|A,B)\, P(C|A)$

E is eliminated, giving the factor

$\tau_2(\cancel{B},C,D) = \sum_E P(E|C,D)\, \cancel{\tau_1(B,C,D)}$

τ doesn't depend on E

C is eliminated, giving

$\tau_3(B,D) = \sum_C \tau_2(\cancel{B},C,D)\, \tau_1(B,C,D)$

D is eliminated, giving

$\tau_4(B) = \sum_D \tau_3(B,D)$

$P(B) = \tau_4(B)\, P(B)$ prior — see solutions

2. (1 pt) What is the largest factor created during the computation and how big is it?

1/1

The largest factor is τ_1 or τ_2, which each have 3 variables. So size $2^3 = 8$.

3. (2 pts) 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.

This will generate $\tau_1(A, B, C, E)$, a factor of 4 variables (hence the largest, since this is all the remaining variables.) The size is $2^4 = 16$.

2/2

4. (2 pts) When would you want to use the clique tree algorithm instead of simple variable elimination? Why?

The clique tree algorithm stores intermediate state in order to make future queries much faster. It takes about twice as long as VE, so should be used if there are > 2 queries.

2/2

5. (4 pts) 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 $(a_2, b_1, c_1, d_2, e_1)$?

1/4

$Pr(C=c_1 | A=a_2) \cdot \cancel{Pr(E=e_1 | C=c_1, d=d_2)}$

$P(A=a_2)$

6. (6 pts) Assume we are using a very large Bayesian Net that represents the connections between hundreds of diseases and thousands of symptoms. The network has intermediate nodes (between diseases and symptoms) that represent the states of internal organs, e.g. kidneys and lungs. A given symptom may ultimately be caused by multiple diseases.

We want to use sampling to estimate the probabilities of the following events; indicate which sampling algorithm you would recommend and (briefly) why.

6/6

(a) $P(symptom_i | disease_j)$

Likelihood weighting. The node for disease$_j$ is likely near the top of the graph (it may not even have any parents), so forcing it to a particular value will probably not generate samples of low weight.

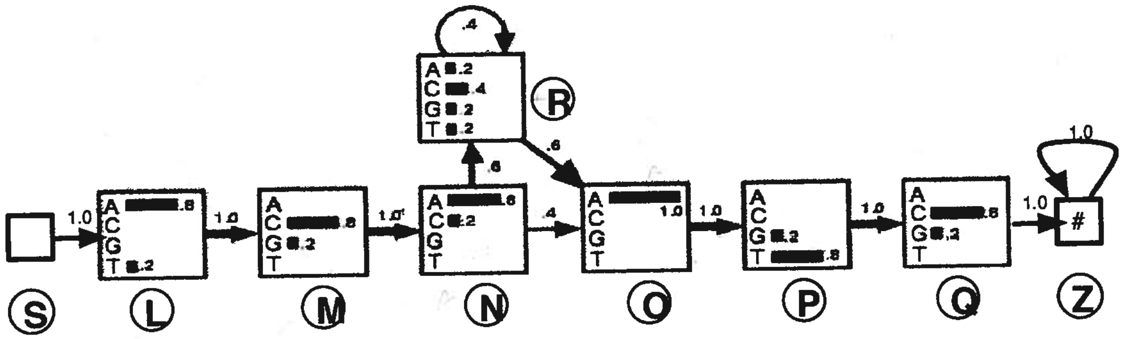
(b) $P(disease_j | symptom_i)$

Gibbs sampling. The node for symptom$_i$ is located at or near the leaves of the tree, so likelihood weighting will choose random diseases that have low probability of giving the appropriate symptom, giving samples of small weight

5 Hidden Markov Models (20 points)

A C A A A A T C
1 2 3 4 5 6 7 8

The following picture specifies a Hidden Markov Model for a DNA sequence "pattern", of length at least 6. There are two special states, S which is the start state and Z which is the end state. We always start in state S, that is, the state distribution at time 0 has all the probability on state S, there are no observations in state S. Once we transition into state Z, all the observations are of an "end of string" symbol (which is not relevant in this problem). The other states show the distribution over the observations possible in that state (one of the four bases of DNA - A,C,G,T). The transition probabilities between the states are given on the arcs between the states.



1. (10 pts) What is the distribution over the states after seeing "ACAAA", i.e. $P(X_5|E_{1:5} = ACAAA)$. Compute the probabilities and write them in the 5th table column. Write the probabilities as sums and products of entries in the diagram, do **not** multiply or add the numbers. So, an entry in the table could be $0.3 \times 0.4 \times 0.2 + 0.5 \times 0.2 \times 0.2$. The table is big enough to keep track of intermediate results, but you do not need to fill in the whole table. If you do, you might want to label important entries with letters, e.g. a and b, so you can reuse them. In this range of time, S and Z have 0 probability, so they're not shown.

State	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
L	1	0	0	0	0
M	0	1	0	0	0
N	0	0	1	0	0
O	0	0	0	1.0 x 0.4 x 1.0 $\boxed{a}$	1.0 x 0.6 x b
P	0	0	0	0	0 x 1.0 x a
Q	0	0	0	0	0
R	0	0	0	0.2 x 0.6 x 1.0 $\boxed{b}$	0.2 x 0.4 x b

10

Note that columns need to be normalized so that they sum to 1

2. (1 pt) Which algorithm should be used for computing the distribution above?

1 Filtering

3. (4 pts) How would you expect the distribution you computed above (for X_5) to change after seeing "ACAAAATC", that is, $P(X_5|E_{1:8} = ACAAAATC)$? You do not need to compute the new distribution numerically; indicate the qualitative changes to the state probabilities.

3 $Pr(X_5 = R)$ would increase, causing $Pr(X_5 = D)$ and $Pr(X_5 = O)$ to decrease.

4. (1 pt) Which algorithm should be used for computing this new distribution?

1 Smoothing

5. (4 pts) Explain how this algorithm would arrive at the expected change in the distribution.

It would factor in $Pr(e_{6:8} | X_5)$, which is computed by a backwards recursive process given

2 by the recurrence

$$Pr(e_{k+1:t} | X_k) = \sum_{X_{k+1}} Pr(e_{k+1} | X_{k+1}) \, Pr(e_{k+2} | X_{k+1}) \, Pr(X_{k+1} | X_k)$$