CPSC 422

Practice Final Exam: Solutions

April 2005

Remember that you can bring in 1 letter-sized piece of paper with anything written on it.

Some important points, based on the midterm:

- Read and answer the question. You will not get marks for writing things (whether they are true or not) that are not relevant to the question.
- Use proper English in full sentences. You will not get marks if we cannot work out what you are saying.
- If a question asks about a particular instance of a problem, make sure your answer refers to that instance. Writing a general formula that you may have copied from the sheet you can bring in, is not worth any marks. (The questions are usually asking to apply that formula to a particular case, to make sure you understand it).

1. Suppose our Q-learning agent, with fixed alpha, and discount gamma, was in state 34 did action 7, received reward 3 and ended up in state 65. What value(s) get updated? Give an expression for the new value. (You need to be as specific as possible)

   Solution: $q[34,7] = q[34,7] + alpha * (3 + gamma * max_a q[65,a] - q[34,7])$

2. In temporal difference learning (e.q. Q-learning), to get the average of a sequence of k values, we let alpha = 1/k. Explain why it may be advantageous to keep alpha fixed in the context of reinforcement learning.

Solution: The initial values are not as good estimates as newer values, and so you may not want to weight them as much. It is simpler to ignore the counts (and so keep alpha fixed). With a fixed alpha it is able to adjust when the environment changes.

3. Explain what happens in reinforcement learning if the agent always chooses the action that maximizes the Q-value. Suggest two ways that can force the agent to explore.

   Solution: It gets stuck in non-optimal policies because it does not explore enough to find the best action from each state. To explore, it can pick random actions occasionally. You could also set the initial values high, so that unexplored regions look good.

4. In MDPs and reinforcement learning explain why we often use discounting of future rewards.

   Solution: With no discounting the sum of the rewards is often infinite. Discounting means that more recent rewards are more valuable than rewards far in the future.

5. What is the main difference between asynchronous value iteration and standard value iteration? Why does asynchronous value iteration often work better than standard value iteration?

   Solution: In standard value iteration all of the values are updated from the previous values in a sweep through the values. In asynchronous value iteration, the values are updated from the current value and can be done in any order (you don't need to sweep through all of the values). It often works better because the latest values are always used and it can concentrate on updating values where they make the most difference (as it doesn't need to sweep through all of the values each time).

6. In learning under uncertainty, when the the EM algorithm used? What is the E-step? What is the M-step?

   Solution: EM algorithm is used for learning probabilities when the value for some variable is not observed in the data. (E.g., the class variable may not be observed). In the E-step, the data is filled in based on the probabilistic model (we get the expected number in the data). In the M-step the probabilities are updated based on the augmented data (we get the maximum likelihood probabilities).

7. Why don't we use the empirical frequencies when learning probabilities from data? That is, if we observe n occurrences of A out of m cases, why shouldn't we just use n/m as our probability estimate?

   Solution: This gives not very good estimates if m is small or if $n = 0$ or $n = m$. Just because you have not observed something does not mean that it should have probability zero (which means it is impossible).

8. Suppose that in a decision network, the decision variable run has parents $look$ and see. Suppose that we are using variable elimination to find the optimal policy. Suppose that after eliminating all of the other variables, we have the factor

look	see	run	value
true	true	yes	23
true	true	no	8
true	false	yes	37
true	false	no	56
false	true	yes	28
false	true	no	12
false	false	yes	18
false	false	no	22

   (a) What is the resulting factor after eliminating run? (Hint: you do not sum out run as it is a decision variable).

   (b) What is the optimal decision function for run?

Solution:

(a) After eliminating run by maximizing, you have the following factor on $look$ and see:

look	see	value
true	true	23
true	false	56
false	true	28
false	false	22

(b) The optimal decision function for run is:

look	see	run
true	true	yes
true	false	no
false	true	yes
false	false	no

That is, if the agent sees, it should run.

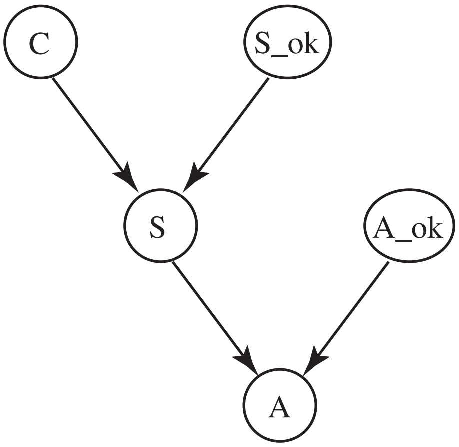
9. Suppose that, in a decision network, there were arcs from random variables “contaminated specimen” and “positive test” to the decision variable “discard sample”. Sally solved the decision network and discovered that there was a unique optimal policy:

contaminated specimen	positive test	discard sample
true	true	yes
true	false	no
false	true	yes
false	false	no

What can you say about the value of information in this case? [You will only get full marks for a precise statements contained in full sentences.]

Solution: The value of “positive test” for the decision “discard sample” is positive (greater than zero). The value of “contaminated specimen” for the decision “discard sample” is zero.

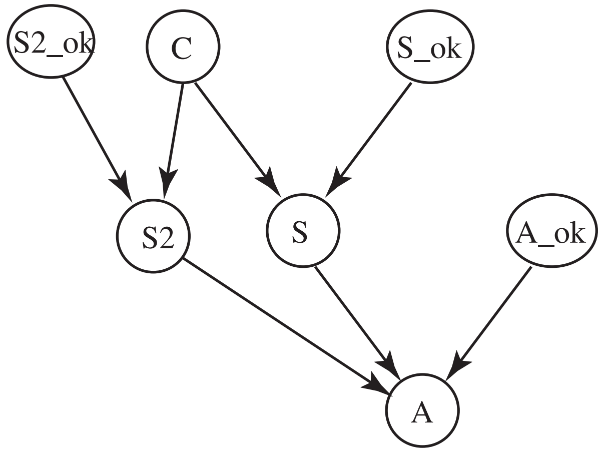
In a nuclear research submarine, a sensor measures the temperature of the reactor core. An alarm is triggered (A=true) if the sensor reading is abnormally high (S=true), indicating an overheating of the core (C=true). The alarm and/or the sensor can be defective (S_ok=false, A_ok=false) which can cause them to malfunction. The alarm system can be modelled by the following belief network (all variables are Boolean):



10.

(a) Suppose we add a second, identical sensor to the system and trigger the alarm when either of the sensors reads a high temperature. The two sensors break and fail independently. Give the corresponding, extended belief network. (Draw the graph and specify any new conditional probabilities).

Solution:

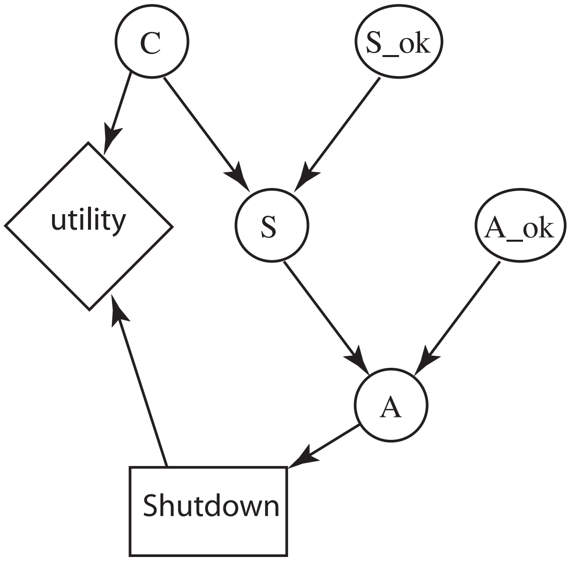


We need to specify

- $P(S2\_ok)$, which would have the same distribution as $P(S\_ok)$
- $P(S2|C, S2\_ok)$ which would have the same distribution as $P(S|C, S\_ok)$.
- $P(A|S, S2, A\_ok)$ which would be true if it is OK and either S or $S2$ is true.

(b) When an alarm is observed, a decision is made whether to shut down the reactor. Shutting down the reactor has a cost c_s associated with it (independent of whether the core was overheating), while not shutting down an overheated core incurs a cost c_m much higher than c_s. Draw the decision network modelling this decision problem for the original system (i.e., only one sensor). Specify any new tables that need to be defined (you should use the parameters c_s and c_m where appropriate in the tables). You can assume that the $utility$ is the negative of $cost$.

Solution:



You need to specify $utility(C, Shutdown)$:

C	Shutdown	Utility
true	true	$-c_s$
true	false	$-c_m$
false	true	$-c_s$
false	false	0

(c) For the decision network in part (c), suppose you need to compute the optimal policy, and the value of the optimal policy using variable elimination. Show, for one legal elimination ordering, what variables are eliminated and what factors are created. [Just give the variables in the factors, not the tables of numbers.]

Solution:

- sum out S_ok, giving factor $f_1(S, C)$
- sum out A_ok, giving factor $f_2(S, A)$.
- sum out S, giving factor $f_3(A, C)$.
- sum out C, giving a factor $f_4(A, Shutdown, utility)$
- maximize over $Shutdown$ giving factor $f_5(A, utility)$
- sum out A giving $f_6(utility)$.

(d) How can the decision function(s) and its expected value be extracted from the elimination of part (d).

Solution: When $Shutdown$ is maximized, the optimal policy is given by the value of shut down that maximizes the utility. The value $f_6(utility)$, which is just a number, gives the expected utility.

11. Suppose you have a job at a company that is building online teaching tools. As you have taken more than one AI course, and have done a number of assignments using different techniques for the same problem, your boss wants to know your opinion on various options they are considering.

    They are planning on building an intelligent tutoring system for teaching elementary physics (e.g., mechanics and electro-magnetism). One of the things that the system will need to do is to diagnose errors that a student may be making.

    For each of the following, answer the explicit questions and use proper English. Answering parts not asked or giving more than one answer when only one is asked for will annoy the boss and result in reduced marks in the exam. The boss also doesn't like jargon, so please use straightforward English.

    The boss has heard of consistency-based diagnosis, abductive diagnosis and belief networks, but wants to know what they involve *in the context of building an intelligent tutoring system for teaching elementary physics.*

(a) Explain what knowledge (about physics and about students) consistency-based diagnosis requires.

Solution: This requires knowledge of the correct answer and what assumptions about the students the correct answer depends on (what skills they need will be assumable).

(b) Explain what knowledge (about physics and about students) abductive diagnosis requires.

Solution: Abductive diagnosis requires what consistency-based diagnosis does as well as models of what errors the students make, and what answers would follow from these errors.

(c) What is the main advantage of using abductive diagnosis over consistency-based diagnosis in this domain?

Solution: It gives a more precise account of what is going on. It lets the system hypothesize particular errors the students may be making.

(d) What is the main advantage of consistency-based diagnosis over abductive diagnosis in this domain?

Solution: It requires less knowledge. It only requires knowledge of the correct answers. It does not require models of mistakes the students could make.

(e) Explain what knowledge (about physics and about students) a belief network model requires.

Solution: A belief network (for a particular problem) would require knowledge of correct behaviors as well as knowledge of mistakes the students can make and probabilities of these.

(f) What is the main advantage of using belief networks (over using abductive diagnosis or consistency-based diagnosis) in this domain?

Solution: It can rank the hypotheses from most likely to least likely. It can be combined with a utility model to make decisions.

(g) What is the main advantage of using abductive diagnosis or consistency-based diagnosis compared to using belief networks in this domain?

Solution: They do not require the probabilities. They can be more expressive in that they can use logic programs with logical variables which can make them applicable to many problems (as opposed to building a different belief network for each problem).

12. Suppose you are given the following scenario. There are four possible diseases a particular patient may have: p, q, r and s. p causes spots. q causes spots. Fever could be caused by one (or more) of q, r or s. The patient has spots and fever.

    (a) Show how to represent this theory using Horn clauses.

    Solution:

    - $spots \leftarrow p.$
    - $spots \leftarrow q.$
    - $fever \leftarrow q.$
    - $fever \leftarrow r.$
    - $fever \leftarrow s.$

    (b) Show how to diagnose this patient using abduction. Show clearly the query and the resulting answer(s).

    Solution: The query is to explain $spots\&fever$. The minimal explanations are $\{q\}$, $\{p, r\}$, $\{p, s\}$.

    (c) Suppose also that p and s cannot occur together. Show how that changes your theory from part (a). Show how to diagnose the patient using abduction with the new theory. Show clearly the query and the resulting answer(s).

    Solution: You add to the Horn clauses the integrity constraint:

    - $false \leftarrow p\&s.$

    With the same query, $spots\&fever$, there are two minimal explanations: $\{q\}$, $\{p, r\}$.