15-780: Graduate AI
Homework Assignment #3

Out: March 23, 2015
Due: April 6, 2015 5 PM

Collaboration Policy: You may discuss the problems with others, but you must write all code and your writeup independently.

Turning In: Please email your assignment by the due date to shayand@cs.cmu.edu and vdperera@cs.cmu.edu. Your solutions should be submitted as a **single** pdf file. If your solutions are handwritten, **scan** them and make sure they are legible and clear. Please submit your code in separate files and provide instructions on how to run it.

1 R-MAX

R-MAX makes the assumption that we know the maximum possible reward. In the real world, this might not always be the case. Let's explore some possible modifications to the algorithm, when we don't know the reward.

(a) Suppose we have an upper bound for the reward (which we are confident is actually an upper bound) and we set R_{max} to be this upper bound. When this bound is very loose, how will the algorithm behave?

(b) Now suppose we initialize $R_{max} = 0$ (which we assume is the minimum possible reward), and every time we see a higher reward ρ, we set $R_{max} = \rho$ (and modify all our unknown states accordingly). Intuitively this might seem like it should work, but in some cases it does not. Give a brief description of a simple MDP where this fails (i.e. where we will never find the optimal policy).

(c) Now suppose we use the same algorithm as in part (b), but instead we replace $R_{max} = \rho+\delta$ for some δ (which may depend on the discount factor γ). Either give a brief explanation of why this won't fail like in part (b), or show that for any value of δ, you can construct an MDP where this fails.

2 Value of Information

Consider a modified version of the example given in class. An oil company wants to buy the rights to drill in one of five blocks of land. It knows that exactly one block contains \$10 million worth of oil and all the other blocks contain \$0 worth of oil. You can take one or both of the following actions:

- Pay a seismologist \$2.5 million to survey one block of land and tell you with certainty whether it is oil rich or not.
- Pay a fortune teller \$1 million to tell you with certainty that the oil-rich block of land is one of two blocks.

Answer the following, justifying your answers with value of information calculations. For simplicity, you may write \$1 to mean \$1 million.

(a) Consider the myopic algorithm where you choose the action whose value of information minis cost is the greatest and repeat until no actions are left or all actions give negative reward. What is the policy if the company uses this algorithm and what is the expected payoff?

(b) What is the optimal policy and expected payoff?

(c) Now suppose the fortune teller rises his cost to \$4.5 million dollars. How do your answers to part (a) and (b) change?

3 Bayes Nets

Let H_i be a random variable taking on values l or r that denotes the handedness of some individual i. A simple hypothesis for handedness is that it is inherited in the following way:

- there is a single gene G_i that effects H_i,
- $H_i = G_i$ with probability $p > 0.5$,
- the gene is inherited from a single parent, and it is equally likely to be inherited from either,
- there is a small non-zero probability m that the gene mutates after inheritance (e.g. if the child inherits from the father and $G_{father} = r$, then with probability m, $G_{child} = l$).

(a) Draw a Bayes' network with nodes G_i and H_i for $i \in \{child, mother, father\}$ that shows this hypothesis.

(b) Answer the following using d-separation arguments.

  (a) is H_{mother} independent of H_{father}?

  (b) is H_{mother} independent of H_{father} given H_{child}?

(c) is H_{child} independent of G_{mother} given G_{child}?

(d) is H_{child} independent of G_{mother} given H_{mother}?

(c) Give the conditional probability table for G_{child}.

(d) Suppose $P(G_{father} = l) = P(G_{mother} = l) = x$. Derive an expression for $P(G_{child} = l)$ in terms of x and m by conditioning on the parent nodes.

(e) Suppose genetic equilibrium holds, i.e. that the distribution of genes in every generation is the same. Calculate x. Do you think the hypothesis for handedness described in this question holds? Explain.

4 Bayesian Knowledge Tracing

The Bayesian Knowledge Tracing (BKT) model (Corbett and Andersen, 1994) is a Bayesian network used to model students' mastery of skills when interacting with an intelligent tutoring system (ITS) or educational game for example. The BKT model is simply an HMM with one binary-valued state (whether the student has mastered the skill or not), and one binary-valued observation (whether the student answers a question correct or not). The student is repeatedly given problems that should help them learn a skill, and the BKT model is used to assign a probability to how likely the student is to have mastered the skill and to predict the probability of the student answering questions correctly in the future. The analogous HMM is shown in Figure 1. K_i is the students internal state after i problems (i.e. it is 1 if the student has mastered the skill and 0 if the student has not), and C_i represents the students answer to the ith problem (i.e. it is 1 if it was correct, and 0 if it was incorrect).

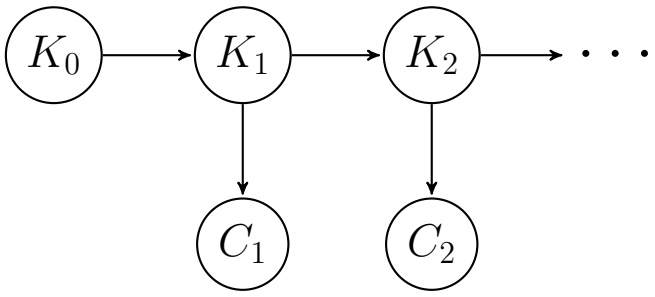


Figure 1: HMM for the BKT model

The standard BKT model assumes a student never forgets a skill, e.g. $P(K_i = 0|K_{i-1} = 1) = 0$. Therefore the four possible parameters of the BKT model are described in Table 1.

Chris and his friend John both come up with models of how students learn how to solve value of information problems, and they get into an argument over which of their models is better. Chris thinks that students can never guess on a question if they haven't mastered a skill. John recalls "guessing" his way through college and still getting good grades. You look at their models, and you think they're both being too extreme, so you propose a model

Parameter	Formula	Interpretation
L_0	$P(K_0 = 1)$	The initial probability that a student has mastered the skill
T	$P(K_i = 1 \mid K_{i-1} = 0)$	The probability that the student transitions from not knowing the skill to knowing the skill
G	$P(C_i = 1 \mid K_i = 0)$	The probability of **guessing**, answering a question correctly when the skill is not mastered
S	$P(C_i = 0 \mid K_i = 1)$	The probability of **slipping**, answering a question incorrectly, despite the skill being mastered

Table 1: BKT Model Parameter Summary

where students guess but not as much. The three models are given in Table 2. Assume for this problem that there is a single true BKT model.

	Chris'	John's	Yours
L_0	2/3	0.25	0.5
T	0.1	0.1	0.1
G	0	0.5	0.3
S	0.1	0.1	0.1

Table 2: Models provided by Chris, John, and You

No observations. First assume the student is given 7 problems but you have no observations of how the student does.

(a) Compute $P(K_i = 1)$ for $i = 0, \ldots, 12$ for each model and plot $P(K_i = 1)$ vs. i putting all three models on the same plot.

(b) Compute $P(C_i = 1)$ for $i = 1, \ldots, 12$ for each model and plot $P(C_i = 1)$ vs. i putting all three models on the same plot.

(c) Describe in 1-2 sentences the similarities and differences among the models in terms of their prediction of the student responses just computed.

Adding observations Now suppose you collect some data for a single student and you see the following trajectory of observations $1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1$.

(d) Using the trajectory above, compute $P(K_i = 1 \mid C_1, \ldots, C_{i-1})$ for $i = 0, \ldots, 12$ for each model and plot $P(K_i = 1 \mid C_1, \ldots, C_{i-1})$ vs. i putting all three models on the same plot.

(e) Using the trajectory above, compute $P(C_i = 1 \mid C_1, \ldots, C_{i-1})$ for $i = 1, \ldots, 12$ for each model and plot $P(C_i = 1 \mid C_1, \ldots, C_{i-1})$ vs. i putting all three models on the same plot.

(f) Which model fits the data better in terms of log likelihood of the observations?

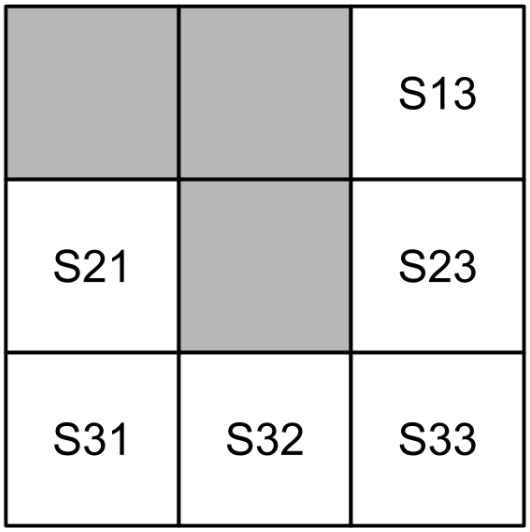
(g) **Extra Credit** (2 pts): In (Beck and Chang, 2007), the authors use Table 1 and Figure 2 to claim that BKT models can be non-identifiable. Based on the results you saw in the previous claim, do you agree with this claim? Why or why not?

References

(Beck and Chang, 2007) Beck, J. E., & Chang, K. M. (2007). Identifiability: A fundamental problem of student modeling. In User Modeling 2007 (pp. 137-146). Springer Berlin Heidelberg.
(Corbett and Anderson, 1994) Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4(4), 253-278.

5 Navigation with POMDP

In this problem, we will explore using POMDPs to plan for navigating in discrete grid worlds, specifically as illustrated below.



The map of the world is known, but the robot's position and orientation is unknown. The goal for the robot is to get to square S31 and announce that it has reached the goal. If it announces and is actually at the goal location, it receives a reward of +100; if it announces and is not at the goal, the reward is -1000. After announcing, the robot is magically transported to a sink state (not illustrated) where all actions leave it in that state and it gets no further reward. The robot can also move forward, turn left, and turn right. Turning left or right is deterministic. Moving forward takes it to the square in front of it, unless that square is blocked by a wall, in which case the robot stays where it is. Walls are the outside boundary of the grid and the boundaries of the gray squares, and are represented as solid black lines. The robot can also observe what is in front of it at a cost of 10. The observation tells whether there is a wall in front with 90% accuracy.

(a) Write this domain as a POMDP. Include descriptions of the states, actions, transition

function, observations, observation function, and rewards. Note that all actions and observation functions must be defined for all states. Use the 3×3 grid above as the environment that the robot needs to explore. For cases where the transition or observation functions are repetitive, you can just enumerate the functions for one set of states and describe the pattern for the other states.

(b) Assume that the robots initial belief is uniformly distributed amongst all the white cells in the environment, but it knows it is facing north (up, in the diagram). Assume that the robot executes the sequence $\langle$*forward, right, observe, left, forward, observe*$\rangle$. Assume the first observation returns wall and the second returns open. Show the belief states of the robot after each action.