Assignment 3- Part II/II

Probabilistic Reasoning

Deadline: December 15th, 11:55pm.
Perfect score: 50.

Assignment Instructions:

Teams: Assignments should be completed by teams of up to three students. No additional credit will be given for students that complete an assignment individually. Please inform the TAs as soon as possible about the members of your team so they can update the scoring spreadsheet (find the TAs' contact info under the course's site on Sakai).

Submission Rules: Submit your reports electronically as a PDF document through Sakai (`sakai.rutgers.edu`). For programming questions, you need to also submit a compressed file via Sakai, which contains your code. Do not submit Word documents, raw text, or hardcopies etc. Make sure to generate and submit a PDF instead. Each team of students should submit only a single copy of their solutions and indicate all team members on their submission. Failure to follow these rules will result in lower grade in the assignment.

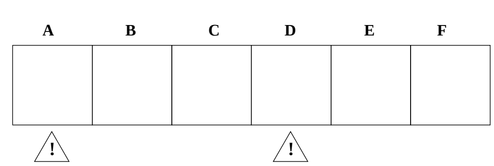
Late Submissions: No late submission is allowed. 0 points for late assignments.

Extra Credit for LaTeX: You will receive 10% extra credit points if you submit your answers as a typeset PDF (using LaTeX, in which case you should also submit electronically your source code). There will be a 5% bonus for electronically prepared answers (e.g., on MS Word, etc.) that are not typeset. If you want to submit a handwritten report, scan it and submit a PDF via Sakai. We will not accept hardcopies. If you choose to submit handwritten answers and we are not able to read them, you will not be awarded any points for the part of the solution that is unreadable.

Precision: Try to be precise. Have in mind that you are trying to convince a very skeptical reader (and computer scientists are the worst kind...) that your answers are correct.

Collusion, Plagiarism, etc.: Each team must prepare its solutions independently from other teams, i.e., without using common notes, code or worksheets with other students or trying to solve problems in collaboration with other teams. You must indicate any external sources you have used in the preparation of your solution. Do not plagiarize online sources and in general make sure you do not violate any of the academic standards of the department or the university. Failure to follow these rules may result in failure in the course.

Problem 1 (*40 points):* You are an interplanetary search and rescue expert who has just received an urgent message: a rover on Mercury has fallen and become trapped in Death Ravine, a deep, narrow gorge on the borders of enemy territory. You zoom over to Mercury to investigate the situation. Death Ravine is a narrow gorge 6 miles long, as shown below. There are volcanic vents at locations A and D, indicated by the triangular symbols at those locations.



The rover was heavily damaged in the fall, and as a result, most of its sensors are broken. The only ones still functioning are its thermometers, which register only two levels: *hot* and *cold*. The rover sends back evidence $E = hot$ when it is at a volcanic vent (A and D), and $E = cold$ otherwise. There is no chance of a mistaken reading. The rover fell into the gorge at position A on day 1, so $X_1 = A$. Let the rover's position on day t be $X_t \in \{A, B, C, D, E, F\}$. The rover is still executing its original programming, trying to move 1 mile east (i.e. right, towards F) every day. However, because of the damage, it only moves east with probability 0.80, and it stays in place with probability 0.20. Your job is to figure out where the rover is, so that you can dispatch your rescue-bot.

1. Filtering: Three days have passed since the rover fell into the ravine. The observations were ($E_1 = hot, E_2 = cold, E_3 = cold$). What is $P(X_3 \mid hot_1, cold_2, cold_3)$, the probability distribution over the rover's position on day 3, given the observations? (This is a probability distribution over the six possible positions).

2. Smoothing: What is $P(X_2 \mid hot_1, cold_2, cold_3)$, the probability distribution over the rover's position on day 2, given the observations? (This is a probability distribution over the six possible positions).

3. Prediction: What is $P(hot_4 \mid hot_1, cold_2, cold_3)$, the probability of observing hot_4 in day 4 given the previous observations in days 1,2, and 3? (This is a single value, not a distribution).

4. Prediction: You decide to attempt to rescue the rover on day 4. However, the transmission of E_4 seems to have been corrupted, and so it is not observed. What is the rover's position distribution for day 4 given the same evidence, $P(X_4 \mid hot_1, cold_2, cold_3)$?

5. Bonus Question for Extra Credit: What is $P(hot_4, hot_5, cold_6 \mid hot_1, cold_2, cold_3)$, the probability of observing hot_4 and hot_5 and $cold_6$ in days 4,5,6 respectively, given the previous observations in days 1,2, and 3? (This is a single value, not a distribution). Note: The answer is a little bit too long, I don't recommend answering this question unless you have already answered it. The max score for HW3, including all extra credits (LaTeX and this question), is 110.

You need to apply the formulas covered in the lecture to get to the answers. Do not just guess the answer using your logic. There are a lot of calculations involved, but most of them result in zeros. Save your time, if a product of probabilities contains a probability that is 0, then just write 0 and do not write down the entire product.

Problem 2 (*10 points*): Consider the Markov Decision Process (MDP) with transition probabilities and reward function as given in the tables below. Assume the discount factor $\gamma = 1$ (i.e., there is no actual discounting).

s	a	s'	$T(s,a,s')$
A	1	A	1
A	1	B	0
A	2	A	0.5
A	2	B	0.5

s	a	$R(s,a)$
A	1	0
A	2	-1

s	a	s'	$T(s,a,s')$
B	1	A	0
B	1	B	1
B	2	A	0
B	2	B	1

s	a	$R(s,a)$
B	1	5
B	2	0

We follow the steps of the Policy Iteration algorithm as explained in the class.

1. Write down the Bellman equation.

2. The initial policy is $\pi(A) = 1$ and $\pi(B) = 1$. That means that action 1 is taken when in state A, and the same action is taken when in state B as well. Calculate the values $V_2^\pi(A)$ and $V_2^\pi(B)$ from **two** iterations of policy evaluation (Bellman equation) after initializing both $V_0^\pi(A)$ and $V_0^\pi(B)$ to 0.

3. Find an improved policy π_{new} based on the calculated values $V_2^\pi(A)$ and $V_2^\pi(B)$.