CSCI 8920 Decision Making Under Uncertainty Assignment 2: Decision Making in Single-Agent Settings

General Information

Deadline: as shown on eLC
Worth: 100 + 10 (bonus) pts

The Assignment

The purpose of this assignment is to understand and become familiar with advanced concepts and methods in decision making. Your grade will be based on the correctness of the solutions. Please be as specific as possible while writing the answers. Note: This assignment is not a group project and everybody should work on it individually.

In order to complete this assignment successfully, you must first carefully read Chapters 15, 16 and 17 of the AI textbook (third edition).

Problems

Part I (70 + 10 (bonus) points) Paper and Pencil

(a) (10 points) Let continuous variables $X_1, \ldots, X_k$ be independently distributed according to the same probability density function $f(x)$. Prove that the density function for $\max\{X_1, \ldots, X_k\}$ is given by $kf(x)(F(x))^{k-1}$, where F is the cumulative distribution for f.

(b) Consider a student who has the choice to buy or not buy a textbook for a course. We will model this as a decision problem with one Boolean decision node, B, indicating whether the agent chooses to buy the book, and two Boolean chance nodes, M, indicating whether the student has

mastered the material in the book, and P, indicating whether the student passes the course. Of course, there is also a utility node, U. A certain student, Sam, has an additive utility function: 0 for not buying the book and -$100 for buying it; and $2,000 for passing the course and 0 for not passing. Sam's conditional probability estimates are as follows:

$P(p|b,m) = 0.9 \quad P(m|b) = 0.9$
$P(p|b,\neg m) = 0.5 \quad P(m|\neg b) = 0.7$
$P(p|\neg b,m) = 0.8 \quad P(p|\neg b,\neg m) = 0.3$

You might think that P would be independent of B given M, but this course has an open-book final so having the book helps.

i. (10 points) Draw the decision network for this problem.

ii. (10 points) Compute the expected utility of buying the book and of not buying it.

iii. (5 points) What should Sam do?

(c) (10 points) Consider a hypothetical manufacturing operation that produces a finished product once an hour at the end of each hour. This machine consists of two identical internal components, each of which must operate once on the product before it is finished. Unfortunately, each of the component can fail spontaneously, and if a component has failed, there is some probability that the product will be defective. Suppose that there are several actions available to us during each one-hour production interval, and one of them is to examine the quality of the product as it rolls off the production line. We may model the state of this problem as zero, one or two internal components having failed, and our observation is whether the products rolling off the line are defective or not. Let the transition and observation matrices for examining be,

$$\begin{bmatrix} 0.81 & 0.18 & 0.01 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1.0 \end{bmatrix} \quad \begin{bmatrix} 1.0 & 0.0 \\ 0.5 & 0.5 \\ 0.25 & 0.75 \end{bmatrix}$$

You will model this machine maintenance problem as a partially observable Markov decision process (POMDP). Let the initial belief state of the agent be a uniform distribution $\langle 0.33, 0.33, 0.34 \rangle$ over the three possible states. Calculate the exact belief states after the agent *examines* and observes the product to be defective, followed by another *examine* and the product is not defective.

(d) Consider a setting where we have a faulty device. Assume that the failure can be caused by a failure in one of n components, exactly one of which is faulty. The probability that repairing component c_i will repair the device is p_i. By the single-fault hypothesis, we have that $\sum_{i=1}^{n} p_i = 1$. Further assume that each component c_i can be examined with cost C_i^o and then repaired (if faulty) with cost C_i^r. Finally, assume that the costs of observing and repairing any component do not depend on any previous actions taken.

i. (10 points) Derive the expected cost until the device is repaired if we observe and repair components in the order $c_1, \ldots, c_n$.

ii. (5 points) Use that to show that the optimal sequence of actions is the one in which we repair components in order of their $\frac{p_i}{C_i^0}$ ratio.

(e) (10 points) Consider the illustrative Tiger problem along with the associated transition and observation probabilities, which you've studied in class. Let an agent's prior belief over the two possible tiger locations be $\langle 0.85, 0.15 \rangle$. The agent *listens* and hears a *growl from the left*. What is the agent's updated posterior belief? Show your steps.

(f) (**Bonus**: 10 points) Let $\delta = ||V^{t+1} - V^t||$ where V is the value function of an MDP and $||V|| = \max_s V(s)$ is the max norm. Here, δ is often known as the temporal difference error. Let ϵ be a given small value and V^* be the converged value function. Given the well-known property that the Bellman update is a contraction by a factor of γ where γ is the discount factor of the MDP, prove that if

$$||V^{t+1} - V^t|| < \frac{\epsilon(1-\gamma)}{\gamma} \text{ then } ||V^* - V^{t+1}|| < \epsilon.$$

Show all steps of your proof.

Part II (30 points) Programming: Policy iteration for MDP

(a) (10 points) Consider a *UAV* performing reconnaissance in a 4×4 grid of sectors as depicted in the figure above. The UAV has the ability to fly north, south, west and east with each action moving it by one sector. Each action is successful in its intended direction by a probability of 0.85. Remaining probability is divided equally between the two directions perpendicular to its intended action. The UAV prefers the sectors

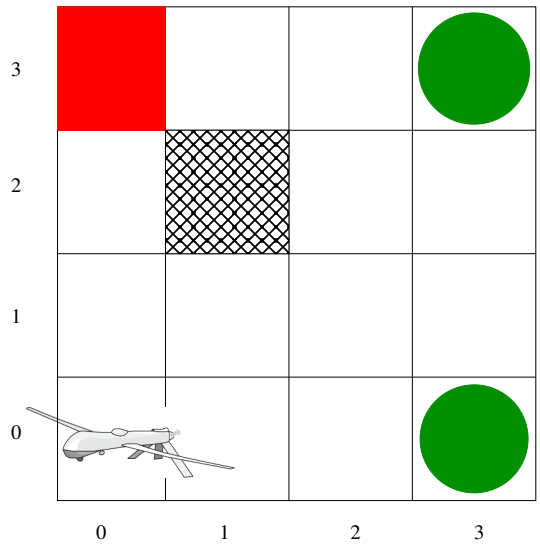

Figure 1: UAV reconnaissance problem for the programming assignment

with a green circle and would like to avoid the red sector. The patterned sector is out of bounds. Write a program in C, C++ or Java that models this problem as a **MDP** consisting of a tuple of states, actions, transition and reward functions. Assign a reward of +1 to the sectors with a green circle and a cost of 1 to the red sector. All other sectors have a cost of 0.05.

(b) (20 points) In the program, implement **policy iteration** for MDPs whose algorithm is provided in Fig. 17.7 of the textbook, for the optimality criterion of discounted infinite horizon with a discount factor of $\gamma = 0.99$. Display the converged policy of the UAV as output.

Use the policy to generate a trajectory from the start state (0,0), and determine if it leads to any of the green sectors. Show this trajectory in the README file.

What and how to hand it in

You'll hand in the *typed or handwritten* answers to Part I in class to the instructor by its deadline. Please be sure to indicate the question numbers alongside the answers. The document should include your name, student id, and all the answers.

You'll submit Part II of your assignment in a *single zipped file* by its deadline via eLC. In the zip file, please include valid and well-commented source code that compiles without any errors. In a separate README file, describe how to compile your program and mention the converged value function output by your program.

Assignments that are **late** but within a day of the deadline will be penalized 33% of the total number of points. Assignments submitted later than one day will not be accepted.