# CS 440: Introduction to Artificial Intelligence

## Homework #3 Part 2 Probabilistic Reasoning

May 2, 2023

*Professor Abdelsam Boularias*

**Jay Patwardhan** 208001851
**Alan Wu** 208000574
**Neel Shejwalkar** 207004853

# Problem 1

Before we begin to tackle these problems, let's define the transition and observation matrices to make the calculations easier.

$$T = \begin{bmatrix} 0.2 & 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$O_{hot} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$O_{cold} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

**1. Filtering**

In order to find $P(X_3|hot_1, cold_2, cold_3)$, we can use the forward model:

$$\begin{aligned} P(X_3|hot_1, cold_2, cold_3) &= O_{hot} * T^T * P(X_2|hot_1, cold_2) \\ &= O_{hot} * T^T * O_{cold} * T^T * P(X_1|hot_1) \end{aligned}$$

where $P(X_1|hot_1)$ is simply $[1, 0, 0, 0, 0, 0]$ as stated in the problem (because the rover is guaranteed to be on location A on the first day). So then,

$$O_{hot} * T^T * O_{cold} * T^T * [1, 0, 0, 0, 0, 0]^T = [0, 0.16, 0.64, 0, 0, 0]$$

which, after normalization, becomes

$$[0, 0.2, 0.8, 0, 0, 0].$$

**2. Smoothing**

To find $P(X_2|hot_1, cold_2, cold_3)$, we can split up the calculation according into a backward model and a forward model to get the smoothed probability.

$$P(X_2|hot_1, cold_2, cold_3) = b_3 * f_{1:2}$$

where

$$\begin{aligned} b_3 &= T * O_{cold} * b_4 \\ &= T * O_{cold} * [1, 1, 1, 1, 1, 1]^T \\ &= [0.8, 1, 0.2, 0.8, 1, 0] \end{aligned}$$

Problem 1 continued on next page. . .

and

$$\begin{aligned} f_{1:2} &= O_{cold} * T^T * P(X_1|hot_1) \\ &= [0, 0.8, 0, 0, 0, 0] \end{aligned}$$

which makes

$$\begin{aligned} P(X_2|hot_1, cold_2, cold_3) &= b_3 * f_{1:2} \\ &= [0.8, 1, 0.2, 0.8, 1, 0] * [0, 0.8, 0, 0, 0, 0] \\ &= [0, 0.8, 0, 0, 0, 0]. \end{aligned}$$

After normalization, this becomes

$$[0, 1, 0, 0, 0, 0].$$

**4. Prediction**
In order to find the distribution $P(X_4|hot_1, cold_2, cold_3)$, we can simply split this up into:

$$\begin{aligned} P(X_4|hot_1, cold_2, cold_3) &= T^T * P(X_3|hot_1, cold_2, cold_3) \\ &= T^T * [0, 0.2, 0.8, 0, 0, 0]^T \\ &= [0, 0.04, 0.32, 0.64, 0, 0]. \end{aligned}$$

We just used the normalized $P(X_3|hot_1, cold_2, cold_3)$ so as to bypass normalizing this new distribution again.

**3. Prediction**
We can use the distribution from part 4 to calculate $P(hot_4|hot_1, cold_2, cold_3)$.

$$\begin{aligned} P(hot_4|hot_1, cold_2, cold_3) &= P(X_4 = 1 \text{ or } X_4 = 4|hot_1, cold_2, cold_3) \\ &= 0 + 0.64 \\ &= 0.64 \end{aligned}$$

# Problem 2

**1. Bellman Equation**
The Bellman equation for this particular MDP would be:

$$V^\pi(s) = R(s, \pi(s)) + \sum_{s' \in \{A,B\}} T(s, \pi(s), s')V^\pi(s')$$

**2. Policy Evaluation**
Using the initial policy rules $\pi(A) = 1$ and $\pi(B) = 1$ with initial policy values $V_0^\pi(A) = V_0^\pi(A) = 0$, two iterations of policy evaluation would look like this:

1. Iteration 1:

$$\begin{aligned} A : V_1^\pi(A) &= R(A, \pi_0(A)) + [T(A, \pi_0(A), A)V_0^\pi(A) + T(A, \pi_0(A), B)V_0^\pi(B)] \\ &= 0 + [0 + 0] \\ &= 0 \\ B : V_1^\pi(B) &= R(B, \pi_0(B)) + [T(B, \pi_0(B), A)V_0^\pi(A) + T(B, \pi_0(B), B)V_0^\pi(B)] \\ &= 5 + [0 + 0] \\ &= 5 \end{aligned}$$

2. Iteration 2:

$$\begin{aligned}
A: V_2^\pi(A) &= R(A,\pi_0(A)) + [T(A,\pi_0(A),A)V_1^\pi(A) + T(A,\pi_0(A),B)V_1^\pi(B)] \\
&= 0 + [0+0] \\
&= 0 \\
B: V_2^\pi(B) &= R(B,\pi_0(B)) + [T(B,\pi_0(B),A)V_1^\pi(A) + T(B,\pi_0(B),B)V_1^\pi(B)] \\
&= 5 + [0+5] \\
&= 10
\end{aligned}$$

So the updated values are $V^\pi(A) = 0$ and $V^\pi(B) = 10$.

**3. Improved Policy**
We can find $\pi_{new}$ using the policy improvement equation:

$$\begin{aligned}
\pi_{new}(A) &= \underset{a\in\{1,2\}}{\operatorname{argmax}}[R(A,a) + (T(A,a,A)V_2(A) + T(A,a,B)V_2(B))] \\
&= 2 \\
\pi_{new}(B) &= \underset{a\in\{1,2\}}{\operatorname{argmax}}[R(B,a) + (T(B,a,A)V_2(A) + T(B,a,B)V_2(B))] \\
&= 1
\end{aligned}$$

The work is shown below.

- $\pi_{new}(A)$
  - a=1: $0 + 1*0 + 0*10 = 0$
  - a=2: $-1 + 0.5*0 + 0.5*10 = 4$
  - We pick a = 2
- $\pi_{new}(B)$
  - a=1: $5 + 0 + 1*10 = 15$
  - a=2: $0 + 0 + 1*10 = 10$
  - We pick a = 1