Group 44:
Ricardo Rei nº 78047
Miguel Carvalho nº 78052

TÉCNICO LISBOA

MSc in Computer Science and Engineering

Learning and Decision Making 2016-2017

Homework 3. Partially Observable Markov Decision Problems

1.

(a)

$\mathbb{X}$ = {A♣, A♦}, the first state means the opponent is holding Ace of Clubs and the second one means he's holding Ace of Diamonds.

$\mathbb{A}$ = {GA♣, GA♦, peek}, the first and second actions are Guess Ace of Clubs and Ace of Diamonds respectively and the third one is peek.

$\mathbb{Z}$ = {OA♣, OA♦}, the possible observations are Observe Ace of Clubs and Observe Ace of Diamonds respectively.

(b)

$$\mathbb{O} = \{ \; O_{Guess\ Ace\ of\ Clubs} = O_{Guess\ Ace\ of\ Diamonds} = \begin{matrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{matrix}, \quad O_{Peek} = \begin{matrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{matrix} \}$$

$$\mathbb{P} = \{ \; P_{Guess\ Ace\ of\ Clubs} = P_{Guess\ Ace\ of\ Diamonds} = \begin{matrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{matrix}, \quad P_{Peek} = \begin{matrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{matrix} \}$$

$$\mathbb{C} = \begin{matrix} 1.0 & 0.0 & 0.5 \\ 0.0 & 1.0 & 0.5 \end{matrix}$$

(c)

The initial belief is that the opponent has an ace of clubs with probability 0.7 which means that the distribution for that belief is [0.7, 0.3].

The agent decides to peek and observes an ace of diamonds, which means that the agent performs the action "peek" and receives the observation OA♦.

Using the forward algorithm:

$\mu_0 = [0.7 \quad 0.3]$.

$Z_{1:1} = \{$ OA♦$\}$

$\alpha_0 = \mu_0^T$

$$\alpha_1 = \text{diag}(O_1) * P^T * \alpha_0 = \begin{matrix} 0.1 & 0.0 \\ 0.0 & 0.9 \end{matrix} * \begin{matrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{matrix} * \begin{matrix} 0.7 \\ 0.3 \end{matrix} = \begin{matrix} 0.07 \\ 0.27 \end{matrix}$$

Normalizing: $\alpha_1 = \begin{matrix} 0.206 \\ 0.794 \end{matrix}$

Which means the the belief of the agent after action peek and observation OA♦ is the distribution:

$$\text{Belief} = [0.206, 0.794]$$