MDL Assignment 3, Part 1

Ayush Sharma (2019101004)

Nitin Chandak (2019101024)

Partially Observable Markov Decision Process(POMDP)

A partially observable Markov decision process (POMDP) is a generalization of a Markov decision process (MDP). A POMDP models an agent decision process in which it is assumed that the system dynamics are determined by an MDP, but the agent cannot directly observe the underlying state. Instead, it must maintain a probability distribution over the set of possible states, based on a set of observations and observation probabilities, and the underlying MDP.

First Roll Number = 2019101004
Second Roll Number = 2019101024
Used Roll Number = 2019101004
Value of x = 0.85
Value of y = 1

State space

State	S1	S2	S3	S4	S5	S6
Type	RED	GREEN	RED	GREEN	GREEN	RED

Probability of actions

	Success	Failure
LEFT	0.85	0.15
RIGHT	0.85	0.15

Probability of observed states for a given state

Columns → Actual state
Rows → Observed states
Table → P(Observed state | Actual state)

- P(Observation = Red|State = Red) = 0.95
- P(Observation = Green|State = Green) = 0.8

	Green	Red
Green	0.8	0.05
Red	0.2	0.95

Formula for calculating next Belief state

$b(s)$ → Previous belief state
$b'(s')$ → New belief state

a → Action by agent
α → Normalizing constant
e → Perceived evidence
Equation:

$$b'(s') = \alpha P(e|s')\Sigma_s P(s'|s,a)b(s)$$

Deriving Beliefs for each Actions:

Initial Beliefs:

S1	S2	S3	S4	S5	S6
0.3333	0.0000	0.3333	0.0000	0.0000	0.3333

Action 1 | Agent took the action Right and observed Green

b'(S[0]) = 0.0500 × (0.1500 × 0.3333 + 0.1500 × 0.0000 + 0.0000 × 0.3333 + 0.0000 × 0.0000 + 0.0000 × 0.0000 + 0.0000 × 0.3333) = 0.0025

b'(S[1]) = 0.8000 × (0.8500 × 0.3333 + 0.0000 × 0.0000 + 0.1500 × 0.3333 + 0.0000 × 0.0000 + 0.0000 × 0.0000 + 0.0000 × 0.3333) = 0.2667

b'(S[2]) = 0.0500 × (0.0000 × 0.3333 + 0.8500 × 0.0000 + 0.0000 × 0.3333 + 0.1500 × 0.0000 + 0.0000 × 0.0000 + 0.0000 × 0.3333) = 0.0000

b'(S[3]) = 0.8000 × (0.0000 × 0.3333 + 0.0000 × 0.0000 + 0.8500 × 0.3333 + 0.0000 × 0.0000 + 0.1500 × 0.0000 + 0.0000 × 0.3333) = 0.2267

b'(S[4]) = 0.8000 × (0.0000 × 0.3333 + 0.0000 × 0.0000 + 0.0000 × 0.3333 + 0.8500 × 0.0000 + 0.0000 × 0.0000 + 0.1500 × 0.3333) = 0.0400

b'(S[5]) = 0.0500 × (0.0000 × 0.3333 + 0.0000 × 0.0000 + 0.0000 × 0.3333 + 0.0000 × 0.0000 + 0.8500 × 0.0000 + 0.8500 × 0.3333) = 0.0142

Normalization Factor = 0.5500

After normalizing, the new Beliefs are:
Action: 1 State: 1 New Belief: 0.004545454545454549
Action: 1 State: 2 New Belief: 0.4848484848484848
Action: 1 State: 3 New Belief: 0.0
Action: 1 State: 4 New Belief: 0.4121212121212121
Action: 1 State: 5 New Belief: 0.07272727272727274
Action: 1 State: 6 New Belief: 0.025757575757575778

S1	S2	S3	S4	S5	S6
0.0045	0.4848	0.0000	0.4121	0.0727	0.0257

Action 2 | Agent took the action Left and observed Red.

b'(S[0]) = 0.9500 × (0.8500 × 0.0045 + 0.8500 × 0.4848 + 0.0000 × 0.0000 + 0.0000 × 0.4121 + 0.0000 × 0.0727 + 0.0000 × 0.0258) = 0.3952

b'(S[1]) = 0.2000 × (0.1500 × 0.0045 + 0.0000 × 0.4848 + 0.8500 × 0.0000 + 0.0000 × 0.4121 + 0.0000 × 0.0727 + 0.0000 × 0.0258) = 0.0001

b'(S[2]) = 0.9500 × (0.0000 × 0.0045 + 0.1500 × 0.4848 + 0.0000 × 0.0000 + 0.8500 × 0.4121 + 0.0000 × 0.0727 + 0.0000 × 0.0258) = 0.4019

b'(S[3]) = 0.2000 × (0.0000 × 0.0045 + 0.0000 × 0.4848 + 0.1500 × 0.0000 + 0.0000 × 0.4121 + 0.8500 × 0.0727 + 0.0000 × 0.0258) = 0.0124

b'(S[4]) = 0.2000 × (0.0000 × 0.0045 + 0.0000 × 0.4848 + 0.0000 × 0.0000 + 0.1500 × 0.4121 + 0.0000 × 0.0727 + 0.8500 × 0.0258) = 0.0167

b'(S[5]) = 0.9500 × (0.0000 × 0.0045 + 0.0000 × 0.4848 + 0.0000 × 0.0000 + 0.0000 × 0.4121 + 0.1500 × 0.0727 + 0.1500 × 0.0258) = 0.0140

Normalization Factor = 0.8403

After normalizing, the new Beliefs are:
Action: 2 State: 1 New Belief: 0.4702681992337164
Action: 2 State: 2 New Belief: 0.00016227180527383377
Action: 2 State: 3 New Belief: 0.47823304034257386
Action: 2 State: 4 New Belief: 0.014712643678160919
Action: 2 State: 5 New Belief: 0.01992337164750958
Action: 2 State: 6 New Belief: 0.01670047329276539

S1	S2	S3	S4	S5	S6
0.4702	0.0001	0.4782	0.0147	0.0199	0.0167

Action 3 | Agent took the action Left and observed Green

b'(S[0]) = 0.0500 × (0.8500 × 0.4703 + 0.8500 × 0.0002 + 0.0000 × 0.4782 + 0.0000 × 0.0147 + 0.0000 × 0.0199 + 0.0000 × 0.0167) = 0.0200

b'(S[1]) = 0.8000 × (0.1500 × 0.4703 + 0.0000 × 0.0002 + 0.8500 × 0.4782 + 0.0000 × 0.0147 + 0.0000 × 0.0199 + 0.0000 × 0.0167) = 0.3816

b'(S[2]) = 0.0500 × (0.0000 × 0.4703 + 0.1500 × 0.0002 + 0.0000 × 0.4782 + 0.8500 × 0.0147 + 0.0000 × 0.0199 + 0.0000 × 0.0167) = 0.0006

b'(S[3]) = 0.8000 × (0.0000 × 0.4703 + 0.0000 × 0.0002 + 0.1500 × 0.4782 + 0.0000 × 0.0147 + 0.8500 × 0.0199 + 0.0000 × 0.0167) = 0.0709

b'(S[4]) = 0.8000 × (0.0000 × 0.4703 + 0.0000 × 0.0002 + 0.0000 × 0.4782 + 0.1500 × 0.0147 + 0.0000 × 0.0199 + 0.8500 × 0.0167) = 0.0131

b'(S[5]) = 0.0500 × (0.0000 × 0.4703 + 0.0000 × 0.0002 + 0.0000 × 0.4782 + 0.0000 × 0.0147 + 0.1500 × 0.0199 + 0.1500 × 0.0167) = 0.0003

Normalization Factor = 0.4866

After normalizing, the new Beliefs are:
Action: 3 State: 1 New Belief: 0.04108919168788052
Action: 3 State: 2 New Belief: 0.7843076877471088
Action: 3 State: 3 New Belief: 0.0012875596118145343
Action: 3 State: 4 New Belief: 0.14578372629886133
Action: 3 State: 5 New Belief: 0.026967328834887783
Action: 3 State: 6 New Belief: 0.0005645058194470697

S1	S2	S3	S4	S5	S6
0.0410	0.7843	0.0012	0.1457	0.0269	0.0005

Belief states obtained

	S1	S2	S3	S4	S5	S6
Initial	0.3333	0.0000	0.3333	0.0000	0.0000	0.3333
Action 1	0.0045	0.4848	0.0000	0.4121	0.0727	0.0257
Action 2	0.4702	0.0001	0.4782	0.0147	0.0199	0.0167
Action 3	0.0410	0.7843	0.0012	0.1457	0.0269	0.0005