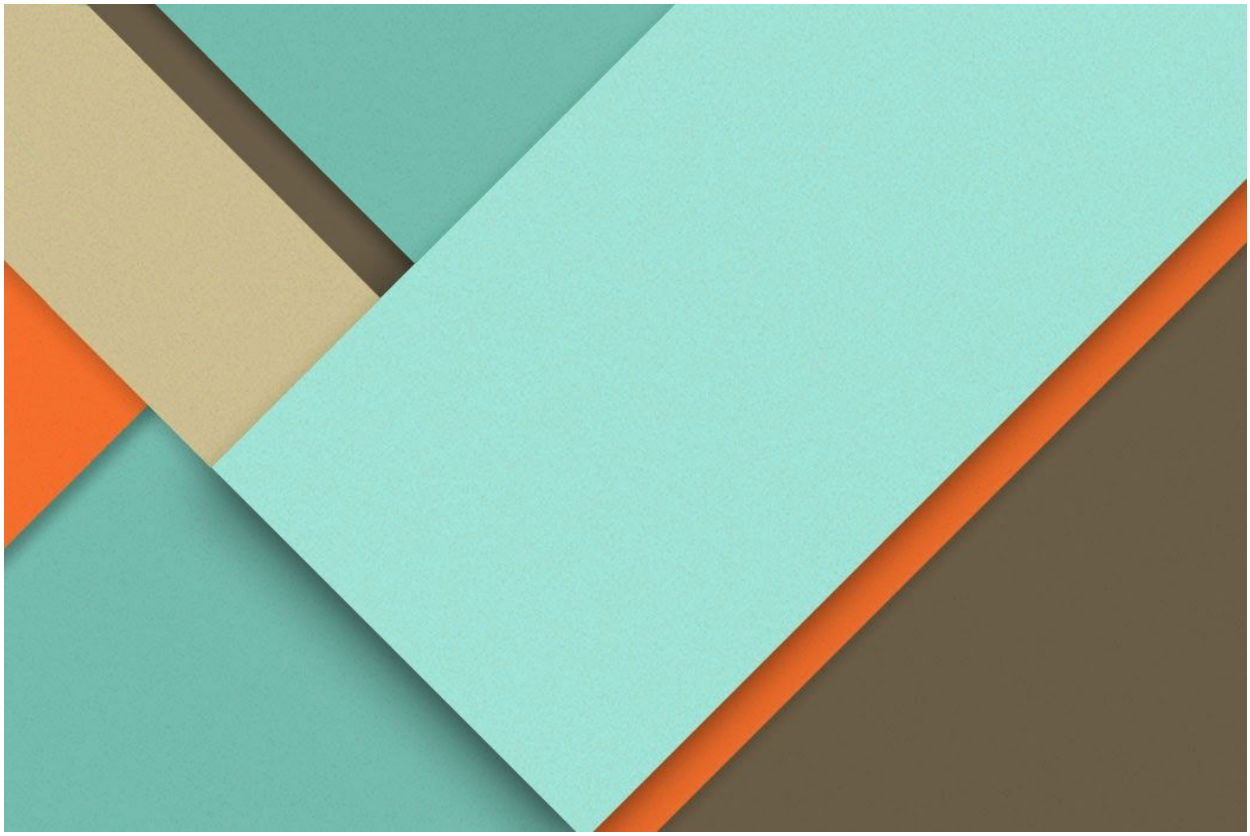


MDL Assignment 4 Part A

29.04.2020

Nishant Sachdeva
2018111040

POMDPs

First let's understand what a POMDP is. It can be regarded as Generalisation of MDP where an agent does not have complete information of the environment rather has beliefs about being in a particular state. Basically, The POMDP or Partially Observable Markov Decision

Process 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, what it does is maintain a probability distribution over the set of possible states, based on a set of observations and observation probabilities, and the underlying MDP.

Formulas Used:

1. A POMDP is modelled as an MDP with continuous belief states = b where it is a distribution of probability over states(S) of an MDP.
2. Action update of a belief state:

$$b'(s) = \sum_{s' \in S} p(s|a, s').b(s')$$

3. For observation update :

$$b'(s) = p(s|o, b) = \frac{p(o|s).b(s)}{\sum_{s' \in S} p(o|s').b(s')}$$

4. From here, we finally get

$$b'(s) = p(s|o, a, b) = \frac{p(s|o, b).p(o|s, a, b)}{p(o|a, b)}$$

Where :

$$b'(s) = p(s|o, a, b)$$

$$p(s|o, a, b) = \sum_{s \in S} p(o|s). \sum_{s \in S} p(s|a, s').b(s')$$

$$p(o|s, a, b) = p(o|s)$$

$$p(s|a, b) = \sum_{s' \in S} p(s|a, s').b(s')$$

The Table to be used in the question

Last three digits of my roll numbers: 040

Thus,

X = 0.99

Y = 1

- Thus the initial table is as follows:

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

Table : 1

The solving of the question

Now, here, I'll upload all the steps that are followed in the solving of the question:

(*) Initial belief states

$[1/3, 1/3, 0, 0, 1/3]$

First Action: Right
Observation: Red

(*) State S1

Sum = $0.01 \times 1/3 + 0.01 \times 1/3 + 0 \times 0 + 0 \times 0 + 0 \times 1/3$

= $0.00333 + 0.00333$

= 0.006667

∴ New belief state = $0.8 \times 0.006667 = 0.005333$

(*) State S2

Sum = $0.99 \times 1/3 + 0 \times 1/3 + 0.01 \times 0 + 0 \times 0 + 0 \times 1/3$

= 0.33

New belief state = $0.8 \times 0.33 = 0.264$

(c) State S3

Sum = $0 \times \frac{1}{3} + 0.99 \times \frac{1}{3} + 0 \times 0 + 0.01 \times 0 + 0 \times \frac{1}{3}$

$= 0.33$

∴ New belief state = $0.05 \times 0.33 = 0.0165$

(d) State S4

Sum = $0.0 \times \frac{1}{3} + 0 \times \frac{1}{3} + 0.99 \times 0 + 0 \times 0 + 0.01 \times \frac{1}{3}$

$= 0 + 0 + 0 + 0 + 0.0033$

$= 0.00\overline{33}$

∴ New belief state = $0.05 \times 0.00\overline{33} = 0.000167$ (rounded off)

(e) State S5

Sum = $0 \times \frac{1}{3} + 0 \times \frac{1}{3} + 0 \times 0 + 0.99 \times 0 + 0.99 \times \frac{1}{3}$

$= 0 + 0 + 0 + 0 + 0.33$

$= 0.33$

∴ New belief state = $0.33 \times 0.8 = 0.264$

∴ Before Normalising, our belief state are

[0.00533, 0.264, 0.0165, 0.000167, 0.264]

Now, we divide by 0.55,

After normalising, the belief state are:

[0.009697, 0.48, 0.03, 0.00303, 0.48]

Sum = 0.999.

NEXT

ACTION : left

OBSERVATION : Green

(*) State S1

Sum = 0.99 × 0.009697 + 0.99 × 0.48
+ 0 × 0.03 + 0 × 0.00303 + 0.0 × 0.48

= 0.0096 + 0.4752
= 0.4848

New belief state = 0.2 × 0.4848 = 0.09696

PTO →

(*) State S2

Sum = 0.01 x 0.009697 + 0.0 x 0.48 + 0.99 x 0.03
+ 0.0 x 0.000303 + 0.0 x 0.48

= 0.00009697 + 0.0297

= 0.029797

New belief state = 0.2 x 0.029797
= 0.005959

(*) State S3

Sum = 0.0 x 0.009697 + 0.01 x 0.48
+ 0.0 x 0.03 + 0.99 x 0.000303
+ 0.0 x 0.48

= 0.0 + 0.0048 + 0.0 + 0.0003 + 0.0

= 0.0051

New belief state = 0.95 * 0.0051 = 0.004845

(*) State S4

Sum = 0.0 x 0.009697 + 0.0 x 0.48 + 0.01 x 0.03
+ 0.0 x 0.000303 + 0.99 x 0.48

= 0 + 0 + 0.0003 + 0 + 0.4752
= 0.4755

∴ New belief state = 0.95 x 0.4755 = 0.451725

(*) State S5

Sum = 0 x 0.00967 + 0 x 0.48 + 0 x 0.03
+ 0.01 x 0.000303 + 0.01 x 0.48

= 0 + 0 + 0 + 0.000003 + 0.0048
= 0.004803

∴ New belief state = 0.2 x 0.004803 = 0.000961

Before normalising :

[0.09696, 0.005959, 0.004845, 0.451725, 0.000961]

Now we divide by 0.56045

After normalising, we get

[0.173004, 0.010633, 0.008645, 0.806004, 0.001714]

Sum = 0.9999

ACTION : LEFT
OBSERVATION : GREEN

(*) State S1

Sum = 0.99 × 0.173004 + 0.99 × 0.010633
+ 0 × 0.008645 + 0 × 0.806004
+ 0 × 0.001714
= 0.171274 + 0.010527 + 0 + 0 + 0
= 0.181801

∴ New belief state = 0.2 × 0.181801 = 0.036360

(*) State S2

Sum = 0.01 × 0.173004 + 0.0 × 0.10633
+ 0.99 × 0.008645 + 0.0 × 0.806004
+ 0.0 × 0.001714
= 0.00173 + 0.0 + 0.008558 + 0.0 + 0.0
= 0.010288

∴ New belief state = 0.2 × 0.010288 = 0.002058

(*) State S3

Sum = 0.0 x 0.173004 + 0.01 x 0.010633
+ 0.04 0.008645 + 0.99 x 0.806004
+ 0.0 x 0.001714

= 0.0 + 0.000106 + 0 + 0.797944 + 0

= 0.79805

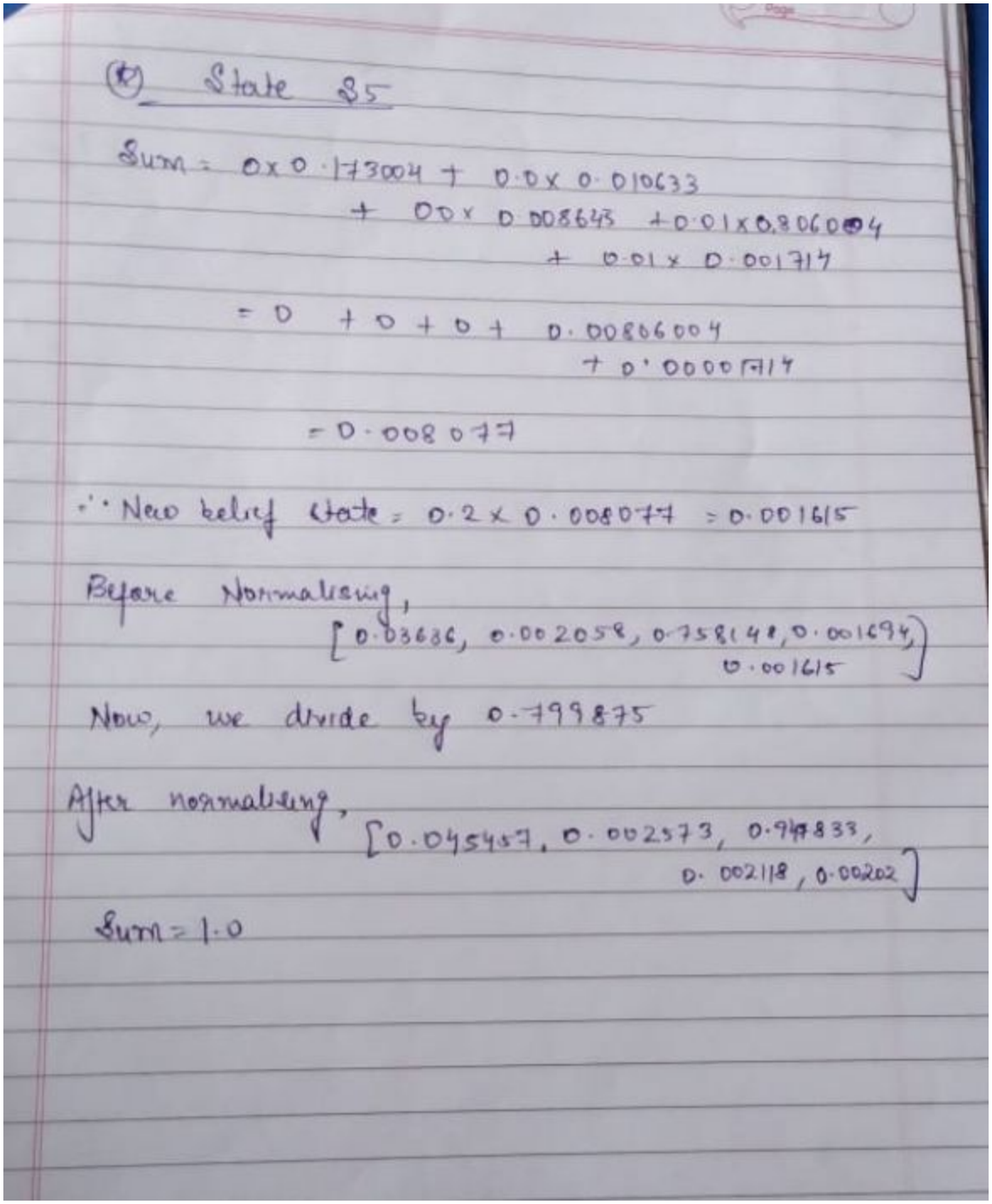
New belief state = 0.95 x 0.79805 = 0.758148

(*) State S4

Sum = (0 x 0.173004) + (0.0 x 0.010633)
+ (0.01 x 0.008645) + (~~0.0 x 0.806004~~)
(0.0 x 0.806004)
+ (0.99 x 0.001714)

= 0 + 0 + 0.00008645 + 0 + 0.001697
= 0.001783

∴ New belief state = 0.95 x 0.001783
= 0.001694



(*) State S5

Sum = 0 x 0.173004 + 0.0 x 0.010633
+ 0.0 x 0.008643 + 0.01 x 0.806004
+ 0.01 x 0.001714

= 0 + 0 + 0 + 0.00806004
+ 0.00001714

= 0.008077

∴ New belief state = 0.2 x 0.008077 = 0.001615

Before Normalising,

[0.03636, 0.002058, 0.758148, 0.001694,
0.001615]

Now, we divide by 0.799875

After normalising,

[0.045457, 0.002573, 0.948833,
0.002118, 0.00202]

Sum = 1.0

Final Answer

So, from the above images, we can see that we arrive at the final answer as follows :

I. The Final Submission that we make is as follows:

```
2018111040
0.99 1
0.009697 0.48 0.03 0.000303 0.48
0.173004 0.010633 0.008645 0.806004 0.001714
0.045457 0.002573 0.947833 0.002118 0.00202
```