# MDL Assignment-3 Part 2

**Tanishq Goel & Rajat Kumar**

- **Roll Number used: 2019114015**

| | | | |
|---|---|---|---|
| **0** **(0,0)** | **1** **(0,1)** | **2** **(0,2)** | **3** **(0,3)** |
| **4** **(1,0)** | **5** **(1,1)** | **6** **(1,2)** | **7** **(1,3)** |

Number of States: 8*8*2=128 possible states
(cell of agent, cell of target, cell state (0 if off, 1 if on))
(a,t,c)-> a*16 + t*2 + c

## Question 1

Target is in (1,0) ie **4**.
o6 means agent is in 1,2,3,6,7
also call can be 1 or 0.

Hence the possible starting states will be
(1,4,0); (2,4,0); (3,4,0); (6,4,0); (7,4,0);
(1,4,1); (2,4,1); (3,4,1); (6,4,1); (7,4,1)

Initial belief state will have all of these with the same probability 0.1.

Rest all states will have initial belief state 0.

Policy file is attached. i.e.Initial beliefs have been taken into account by mapping above states to single integer representation.
start include:

## Question 2

Agent is in (1,1) ie 5
One neighbourhood means within distance 1.
So, the target is at cells 1,4,5,6.
Given call=0
Initial belief state will have all of these with the same probability ie 1⁄4.
Rest all states will have initial belief state 0.

Initial beliefs have been taken into account by mapping above states to single integer representation.
They are specified by including the line
So, possible states are (5,1,0), (5,4,0), (5,5,0),(5,6,0)

## Question 3

Expectations were calculated by using the following command:
**--simLen 100 --simNum 1000 --policy-file**
flag with **pomdpsim** program, and output file from **pomdpsol**.

Expected value for q1: 3.03813
Expected value for q2: 9.30083
Image of each output is below.

```
Loading the model ...
  input file   : 2019114015.pomdp

Loading the policy ...
  input file   : out.policy

Simulating ...
  action selection :  one-step look ahead

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 100             2.80643
 200             2.75434
 300             2.87231
 400             2.85928
 500             2.85151
 600             2.9122
 700             2.99115
 800             3.03335
 900             3.02411
 1000            3.03813
-----------------------------------

Finishing ...

-------------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------------
 1000            3.03813            (2.80151, 3.27476)
-------------------------------------------------------------------
```

```
Loading the model ...
  input file   : 2019114015_b.pomdp

Loading the policy ...
  input file   : out.policy

Simulating ...
  action selection :  one-step look ahead

-----------------------------------
 #Simulations  | Exp Total Reward
-----------------------------------
 100             9.03479
 200             9.10983
 300             9.14962
 400             9.18448
 500             9.10691
 600             9.1975
 700             9.2329
 800             9.31046
 900             9.29923
 1000            9.30083
-----------------------------------

Finishing ...

-------------------------------------------------------------
 #Simulations  | Exp Total Reward | 95% Confidence Interval
-------------------------------------------------------------
 1000            9.30083            (9.04713, 9.55452)
-------------------------------------------------------------
```

**Question 4**

The target can be at the following locations:
(0,1) -> 1
(0,2) -> 2
(1,1) -> 5
(1,2) -> 6
With equal probability of **0.25.**

If the agent is at location (0,0):

- Target is at location (0,1), then o2 will be observed as the target is in the right cell of the agent. For the remaining 3 positions of the target, observation will be o6 because target is not in the 1 neighbourhood of the agent.
- So the conditional probability for observations will be:
  - 0.25 for o2
  - 0.75 for o6
  - 0 for remaining observations (o1,o3,o4,o5)

If the agent is at location (1,3):

- Target is at location (1,2), then o4 will be observed as the target is in the left cell of the agent. For the remaining 3 positions of the target, observation will be o6 because target is not in the 1 neighbourhood of the agent.
- So the conditional probability for observations will be:
  - 0.25 for o4
  - 0.75 for o6
  - 0 for remaining observations (o1,o2,o3,o5)

Now solving for the final probabilities for the observation using the given values that are 0.4 when agent location is (0,0) and 0.6 when agent location is (1,3):

o1: 0
o2: 0.4 * 0.25 = 0.1
o3: 0
o4: 0.6 * 0.25 = 0.15
o5: 0
o6: (0.4 * 0.75) + (0.6 * 0.75) = 0.75

So, it can be clearly observed that **o6** is most likely to be observed.

## Question 5

On running  pomdp sol

```
-------------------------------------------------------------------------------
 Time   |#Trial |#Backup |LBound    |UBound    |Precision  |#Alphas |#Beliefs
-------------------------------------------------------------------------------
 0.01    22      169      3.16238    3.16324    0.000859307 48       39
-------------------------------------------------------------------------------
```

We will use the #Trial as T value for calculation.

The formula used in calculation is

How many policy trees, if |A| actions, |O| observations, T horizon:

1. Number of nodes in tree:

$$N = \Sigma_{i=1}^{T-1} |O|^i = \frac{(|O|^T - 1)}{|O| - 1}$$

2. Number of trees:

$$|A|^N$$

Here
|A| = 5
|O| = 6
T = 22

Thus,

$$N = \frac{(6^{22} - 1)}{(6 - 1)} = 2.6324341e + 16$$

Now,

$$|A|^N = 5^{(2.6324341e+16)} = 1.5315747e + 16$$

is the approximate number of policy trees obtained.