Assignment 1
AIL722

Assignment 1

1. Will be released by today evening
2. Deadline is roughly 2 week from date of release
3. Needs to be done individually
4. More logistics at the end of PPT

Assignment about

All about HMMs.

Implementing the following from scratch (some starter code will be provided):

1. Sampling
2. Likelihood (Forward Algorithm)
3. Decoder (Viterbi)
4. Learning (Baum-Welch)

HMM: Quick Recap

Problem 1 (Likelihood):

Given an HMM $\lambda = (\mathcal{T}, \mathcal{B})$ and an observation sequence O, determine the likelihood $P(O \mid \lambda)$.

Problem 2 (Decoding):

Given an observation sequence O and an HMM $\lambda = (\mathcal{T}, \mathcal{B})$, discover the best hidden state sequence.

Problem 3 (Learning):

Given an observation sequence O and the set of states in the HMM, learn the HMM parameters $\mathcal{T}$ and $\mathcal{B}$.

$$\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$$

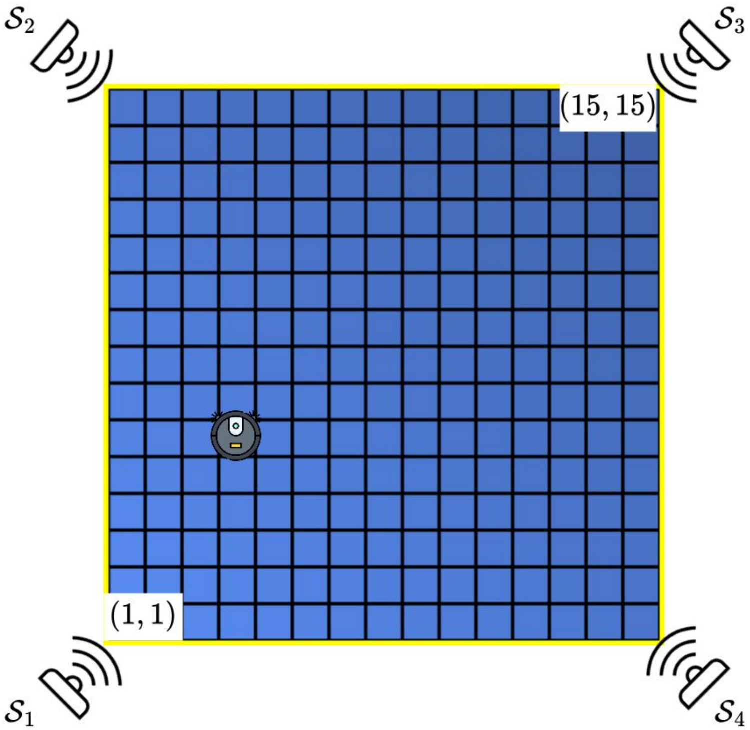
$$T = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}$$

$$\rho = \{p(s_1), p(s_2), \ldots, p(s_N)\}$$

$$\mathcal{O} = \{o_1, o_2, \ldots, o_M\}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1M} \\ b_{21} & b_{22} & \cdots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \cdots & b_{NM} \end{pmatrix}$$

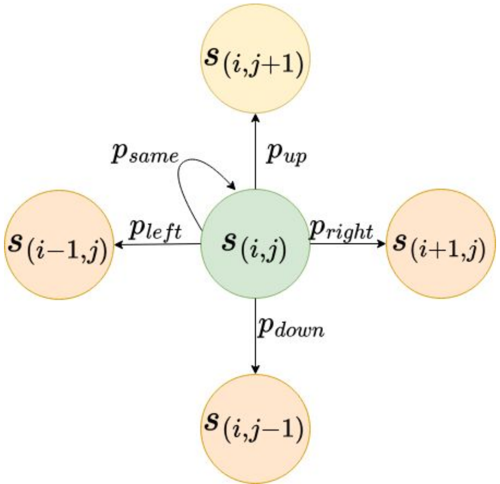
HMM: Assignment

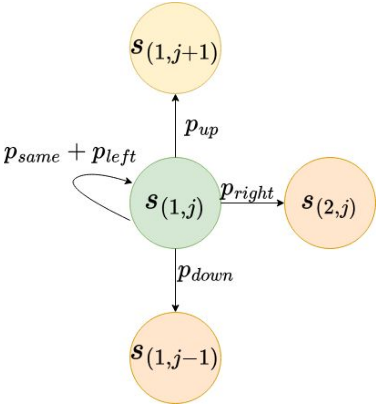


$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \ldots, s_{(15,15)}\}$$

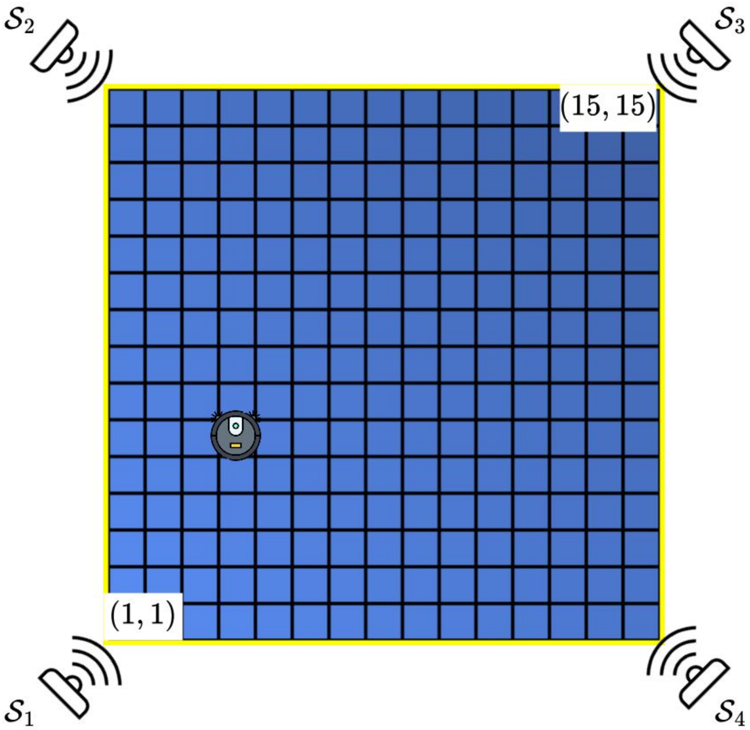
$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \ldots, o_{(1,1,1,1)}\}$$

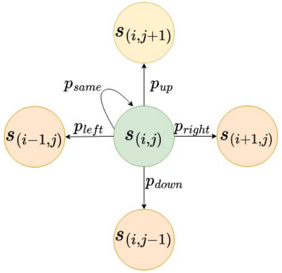
$$\rho = [1, 0, ..]$$





HMM: Assignment





$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \ldots, s_{(15,15)}\}$$

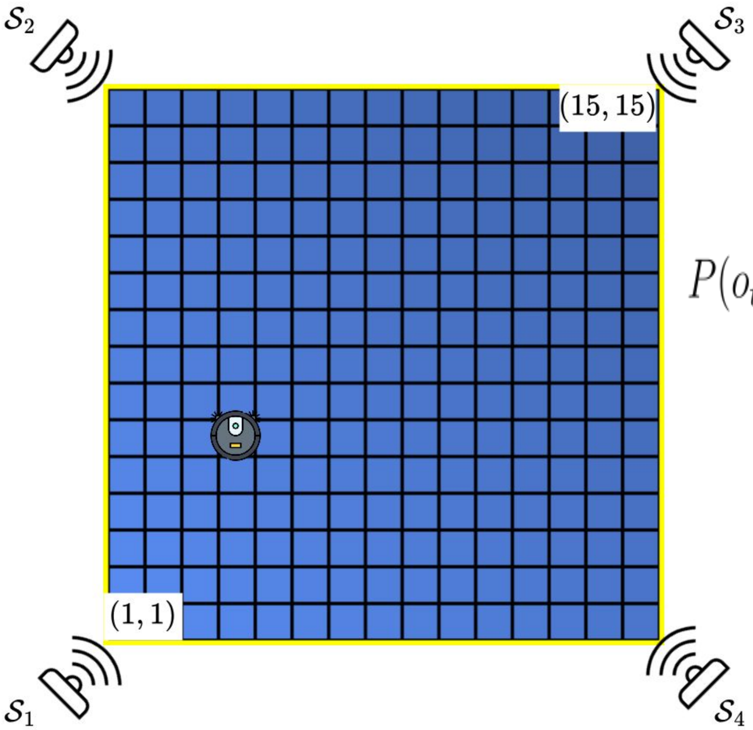
$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \ldots, o_{(1,1,1,1)}\}$$

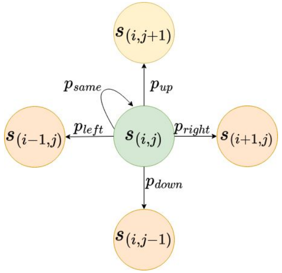
$$T = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix} \qquad \mathcal{T} : |\mathcal{S}| \times |\mathcal{S}|$$

But transition matrix can completely be determined by 5 values

$$\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})$$

HMM: Assignment





$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \dots, s_{(15,15)}\}$$

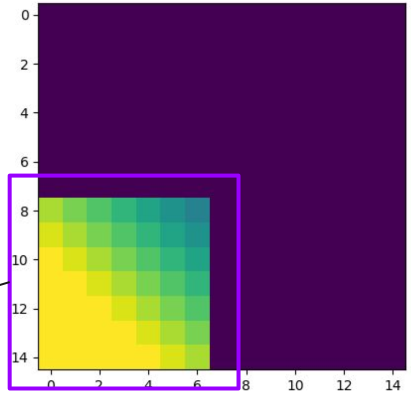
$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \dots, o_{(1,1,1,1)}\}$$

$$\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})$$

$$|\mathcal{B}| : |\mathcal{S}| \times 16$$

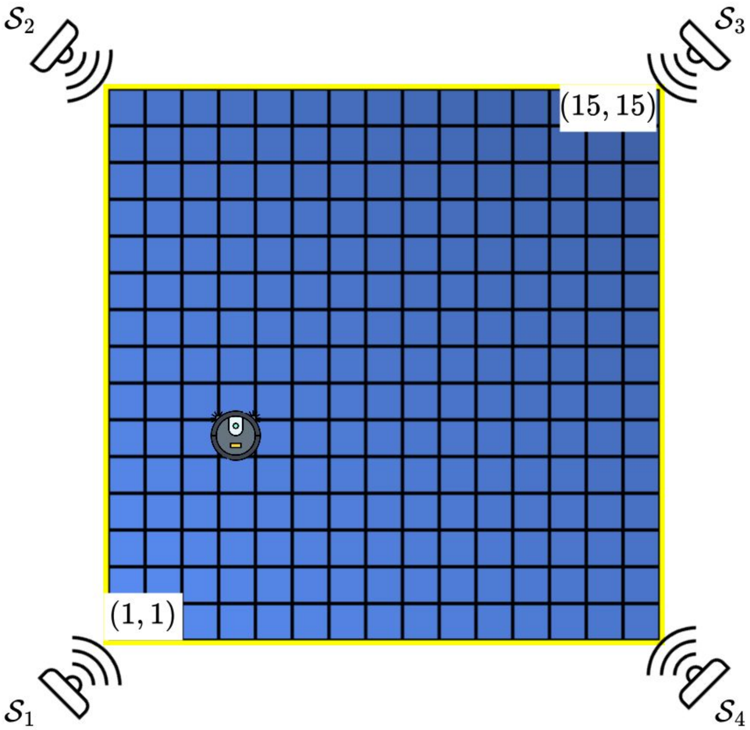
$$P(o_t = (o_t^1, o_t^2, o_t^3, o_t^4) | s_{(i,j)}) = P(o_t^1 | s_{(i,j)}) P(o_t^2 | s_{(i,j)}) P(o_t^3 | s_{(i,j)}) P(o_t^4 | s_{(i,j)})$$

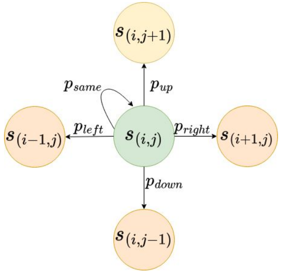
$P(o_t^1 | s_{(i,j)})$



Cover of sensor

HMM: Assignment





$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \ldots, s_{(15,15)}\}$$
$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \ldots, o_{(1,1,1,1)}\}$$
$$\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})$$

$$P(o_t = (o_t^1, o_t^2, o_t^3, o_t^4)|s_{(i,j)}) = P(o_t^1|s_{(i,j)})P(o_t^2|s_{(i,j)})P(o_t^3|s_{(i,j)})P(o_t^4|s_{(i,j)})$$

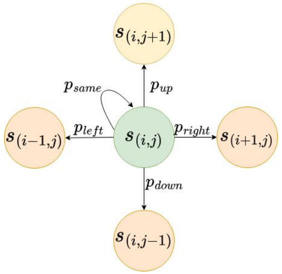
Q1. Generate the samples of states and sensor observation from HMM. Plot the state trajectory (we will provide plot code).

Q2. Compute the likelihood of sensor observations (sampled above) using Forward algorithm

Q3. Decode the sampled sensor observations to find most probable state trajectory using Viterbi algorithm.

HMM: Q4 Baum-Welch

Q4. Implement Baum Welch algorithm to learn the T and B matrices using the sensor observation sequence. Assume Rho to be fixed.



$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \ldots, s_{(15,15)}\}$$

$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \ldots, o_{(1,1,1,1)}\}$$

$$\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})$$

$$P(o_t = (o_t^1, o_t^2, o_t^3, o_t^4)|s_{(i,j)}) = P(o_t^1|s_{(i,j)})P(o_t^2|s_{(i,j)})P(o_t^3|s_{(i,j)})P(o_t^4|s_{(i,j)})$$

For single sequence

$$p(o_{1:T}) = \sum_{i=1}^{N} \alpha_T(i)$$

$$\gamma_t(j) = \frac{\alpha_t(j).\beta_t(j)}{\sum_{i=1}^{N} \alpha_T(i)}$$

$$\xi_t(i,j) = \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{\sum_{i=1}^{N} \alpha_T(i)}$$

$$\hat{p}(o^k \mid s^j) = \frac{\sum_{t=1}^{T} \gamma_t(j) 1(o_t = o_k)}{\sum_{t=1}^{T} \gamma_t(j)}$$

$$\hat{p}(s^j \mid s^i) = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{k=1}^{N} \xi_t(i,k)}$$

For multiple sequence

$$\hat{p}(s^j|s^i) = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T-1} \xi_{t,r}(i,j)}{\sum_{r=1}^{R} \sum_{t=1}^{T-1} \sum_{k=1}^{n} \xi_{t,r}(i,k)}$$

$$\hat{p}(o^k|s^j) = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T-1} \texttt{Ind}(o_{t,r} = o^k)\gamma_{t,r}(j)}{\sum_{r=1}^{R} \sum_{t=1}^{T-1} \gamma_{t,r}(j)}$$

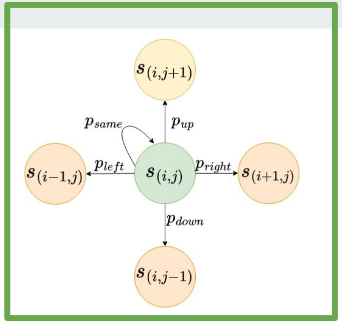
HMM: Q4 Baum-Welch

Q4. Implement Baum Welch algorithm to learn the T and B matrices using the sensor observation sequence. Assume Rho to be fixed. Three subparts (depending on B learning)

For multiple sequence

$$\hat{p}(s^j|s^i) = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T-1}\xi_{t,r}(i,j)}{\sum_{r=1}^{R}\sum_{t=1}^{T-1}\sum_{k=1}^{n}\xi_{t,r}(i,k)}$$

$$\hat{p}(o^k|s^j) = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T-1}\mathtt{Ind}(o_{t,r}=o^k)\gamma_{t,r}(j)}{\sum_{r=1}^{R}\sum_{t=1}^{T-1}\gamma_{t,r}(j)}$$



$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \ldots, s_{(15,15)}\}$$

$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \ldots, o_{(1,1,1,1)}\}$$

$$\boxed{\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})}$$

$$P(o_t = (o_t^1, o_t^2, o_t^3, o_t^4)|s_{(i,j)}) = P(o_t^1|s_{(i,j)})P(o_t^2|s_{(i,j)})P(o_t^3|s_{(i,j)})P(o_t^4|s_{(i,j)})$$

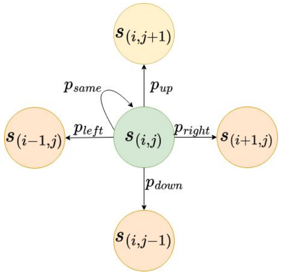
$$\mathcal{N}(x,y) = \{(x,y),(x+1,y),(x-1,y),(x,y+1),(x,y-1)\}$$

$$\hat{p}_{right} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T-1}\sum_{(x,y)=(2,2)}^{(13,13)}\xi_{t,r}((x,y),(x+1,y))}{\sum_{r=1}^{R}\sum_{t=1}^{T-1}\sum_{(x,y)=(2,2)}^{(13,13)}\sum_{(x',y')\in\mathcal{N}(x,y)}\xi_{t,r}((x,y),(x',y'))}$$

$$\hat{\mathcal{T}} = f(\hat{p}_{right}, \hat{p}_{left}, \hat{p}_{up}, \hat{p}_{down}, \hat{p}_{same})$$

HMM: Q4 Baum-Welch

Q4. Implement Baum Welch algorithm to learn the T and B matrices using the sensor observation sequence. Assume Rho to be fixed.
Three subparts (depending on B learning)



$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \dots, s_{(15,15)}\}$$

$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \dots, o_{(1,1,1,1)}\}$$

$$\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})$$

$$P(o_t = (o_t^1, o_t^2, o_t^3, o_t^4)|s_{(i,j)}) = P(o_t^1|s_{(i,j)})P(o_t^2|s_{(i,j)})P(o_t^3|s_{(i,j)})P(o_t^4|s_{(i,j)})$$

$$\hat{\mathcal{T}} = f(\hat{p}_{right}, \hat{p}_{left}, \hat{p}_{up}, \hat{p}_{down}, \hat{p}_{same})$$

Subpart on B learning

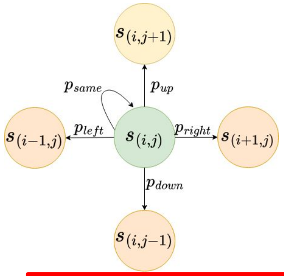
1. Keep B fix (use the ground truth)
2. Learn the B

$$\hat{p}(o^k|s^j) = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T-1} \mathtt{Ind}(o_{t,r} = o^k)\gamma_{t,r}(j)}{\sum_{r=1}^{R}\sum_{t=1}^{T-1} \gamma_{t,r}(j)}$$

3. Exploit the facts about independence of sensor and cover of sensor

HMM: Q4 Baum-Welch

Q4. Implement Baum Welch algorithm to learn the T and B matrices using the sensor observation sequence. Assume Rho to be fixed. Three subparts (depending on B learning)



$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \ldots, s_{(15,15)}\}$$

$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \ldots, o_{(1,1,1,1)}\}$$

$$\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})$$

$$\boxed{P(o_t = (o_t^1, o_t^2, o_t^3, o_t^4)|s_{(i,j)}) = P(o_t^1|s_{(i,j)})P(o_t^2|s_{(i,j)})P(o_t^3|s_{(i,j)})P(o_t^4|s_{(i,j)})}$$

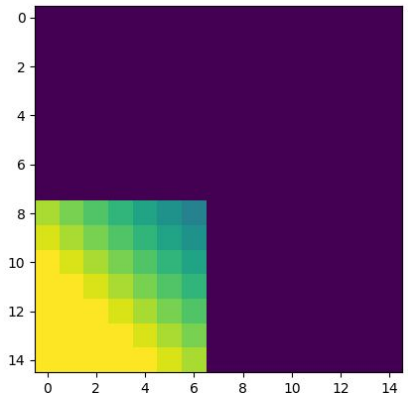
$$\hat{\mathcal{T}} = f(\hat{p}_{right}, \hat{p}_{left}, \hat{p}_{up}, \hat{p}_{down}, \hat{p}_{same})$$

Subpart on B learning

3. Exploit the facts about independence of sensor and cover of sensor

$$\tilde{p}(o_t^1 = 1|s_{(i,j)}) = \frac{\sum_{r=1}^{n}\sum_{t=1}^{T}\text{Ind}(o_t^1 \in o_{t,r})\gamma_{r,t}(s_{i,j})}{\sum_{r=1}^{n}\sum_{t=1}^{T}\gamma_{r,t}(s_{i,j})}$$

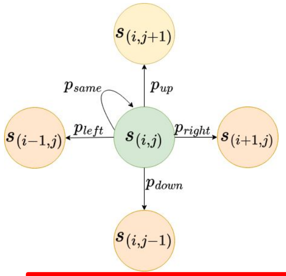
$$\hat{p}(o_t^1 = 1|s_{(i,j)}) = \text{mask}_1[i,j] * \tilde{p}(o_t^1 = 1|s_{(i,j)})$$



HMM: Q4 Baum-Welch

Q4. Implement Baum Welch algorithm to learn the T and B matrices using the sensor observation sequence. Assume Rho to be fixed. Three subparts (depending on B learning)

Food for thought:
We reduced the problem of learning 16*225 + 225*225 to 5 + 4*k*k
where k is cover of sensor and k = 0.6*15



$$\mathcal{S} = \{s_{(1,1)}, s_{(1,2)}, \ldots, s_{(15,15)}\}$$

$$\mathcal{O} = \{o_{(0,0,0,0)}, o_{(0,0,0,1)}, \ldots, o_{(1,1,1,1)}\}$$

$$\mathcal{T} = f(p_{right}, p_{left}, p_{up}, p_{down}, p_{same})$$

$$\boxed{P(o_t = (o_t^1, o_t^2, o_t^3, o_t^4) | s_{(i,j)}) = P(o_t^1 | s_{(i,j)}) P(o_t^2 | s_{(i,j)}) P(o_t^3 | s_{(i,j)}) P(o_t^4 | s_{(i,j)})}$$

What are we providing?

1. Starter code of algorithms
2. Path plotting code

Expected Runtimes and compute need?

1. Part 1,2,3 will take <5 mins to run each if implemented correctly and need <1GB ram
2. Each sub part of part 3 will take 15-20 mins each if implemented correctly and efficiently. Will need <= 5 GB ram

Other Logistics

1. Deadline:
2. Individual assignment
3. Submission: Moodle (Code + Report)
4. Late submission: 5 Buffer days in total (after that 10% each day)
5. Collaboration policy:
    a. Use of Code Generated from LLMs: Code generated from Language Models (LLMs) will be considered cheating. Any such code found in your submission will be treated as a violation of the honor code.
    b. Copying Code from Online Resources: Copying code from any online resource is also considered copying. Ensure that all code submitted is your original work.
    c. You are free to discuss the problems with other students in the class. However, the final solution/code that you produce should come through your individual efforts.

Other Logistics

6. Doubts -> piazza discussion preferred, class and office hours