CS 188 Summer 2024 Introduction to Artificial Intelligence

Exam Prep 7 Solutions

Q1. HMMs

Consider a process where there are transitions among a finite set of states $s_1, \cdots, s_k$ over time steps $i = 1, \cdots, N$. Let the random variables $X_1, \cdots, X_N$ represent the state of the system at each time step and be generated as follows:

- Sample the initial state s from an initial distribution $P_1(X_1)$, and set $i = 1$
- Repeat the following:
  1. Sample a duration d from a duration distribution P_D over the integers $\{1, \cdots, M\}$, where M is the maximum duration.
  2. Remain in the current state s for the next d time steps, i.e., set

$$x_i = x_{i+1} = \cdots = x_{i+d-1} = s \quad (1)$$

  3. Sample a successor state s' from a transition distribution $P_T(X_t|X_{t-1} = s)$ over the other states $s' \neq s$ (so there are no self transitions)
  4. Assign $i = i + d$ and $s = s'$.

This process continues indefinitely, but we only observe the first N time steps.

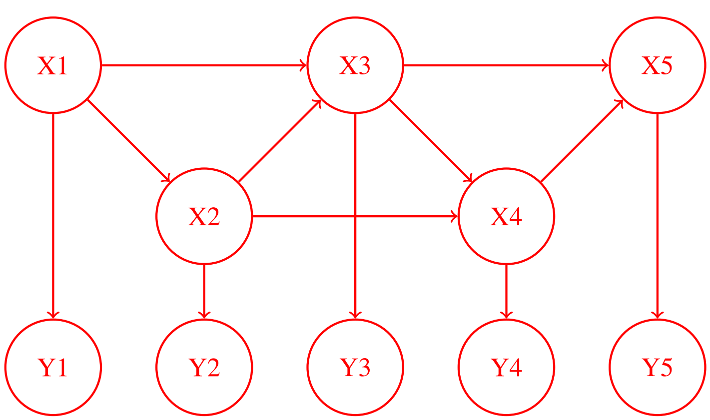
(a) Assuming that all three states s_1, s_2, s_3 are different, what is the probability of the sample sequence $s_1, s_1, s_2, s_2, s_2, s_3, s_3$? Write an algebraic expression. Assume $M \geq 3$.

$$p_1(s_1)p_D(2)p_T(s_2|s_1)p_D(3)p(s_3|s_2)(1 - p_D(1)) \quad (2)$$

At each time step i we observe a noisy version of the state X_i that we denote Y_i and is produced via a conditional distribution $P_E(Y_i|X_i)$.

(b) Only in this subquestion assume that $N > M$. Let $X_1, \cdots, X_N$ and $Y_1, \cdots, Y_N$ random variables defined as above. What is the maximum index $i \leq N - 1$ so that $X_1 \perp\!\!\!\perp X_N | X_i, X_{i+1}, \cdots, X_{N-1}$ is guaranteed?
$i = N - M$

(c) Only in this subquestion, assume the max duration $M = 2$, and P_D uniform over $\{1, 2\}$ and each x_i is in an alphabet $\{a, b\}$. For $(X_1, X_2, X_3, X_4, X_5, Y_1, Y_2, Y_3, Y_4, Y_5)$ draw a Bayes Net over these 10 random variables with the property that removing any of the edges would yield a Bayes net inconsistent with the given distribution.



(d) In this part we will explore how to write the described process as an HMM with an extended state space. Write the states $z = (s,t)$ where s is a state of the original system and t represents the time elapsed in that state. For example, the state sequence $s_1, s_1, s_1, s_2, s_3, s_3$ would be represented as $(s_1, 1), (s_1, 2), (s_1, 3), (s_2, 1), (s_3, 1), (s_3, 2)$.

Answer all of the following in terms of the parameters $P_1(X_1)$, $P_D(d)$, $P_T(X_{j+1}|X_j)$, $P_E(Y_i|X_i)$, k (total number of possible states), N, and M (max duration).

(i) What is $P(Z_1)$?

$$P(x_1, t) = \begin{cases} P_1(x_1) & \text{if } t = 1 \\ 0 & \text{o.w.} \end{cases} \tag{3}$$

(ii) What is $P(Z_{i+1}|Z_i)$? Hint: You will need to break this into cases where the transition function will behave differently.

$$P(X_{i+1}, t_{i+1}|X_i, t_i) = \begin{cases} P_D(d \geq t_i + 1|d \geq t_i) & \text{when } X_{i+1} = X_i \text{ and } t_{i+1} = t_i + 1 \text{ and } t_{i+1} \leq M \\ P_T(X_{i+1}|X_i)P_D(d = t_i|d \geq t_i) & \text{when } X_{i+1} \neq X_i \text{ and } t_{i+1} = 1 \\ 0 & \text{o.w.} \end{cases} \tag{4}$$

Where $P_D(d \geq t_i + 1|d \geq t_i) = P_D(d \geq t_i + 1)/P_D(d \geq t_i)$.
Being in X_i, t_i, we know that d was drawn $d \geq t_i$. Conditioning on this fact, we have two choices, if $d > t_i$ then the next state is $X_{i+1} = X_i$, and if $d = t_i$ then $X_{i+1} \neq X_i$ drawn from the transition distribution and $t_{i+1} = 1$.

(iii) What is $P(Y_i|Z_i)$?
$p(Y_i|X_i, t_i) = P_E(Y_i|X_i)$

(e) In this question we explore how to write an algorithm to compute $P(X_N|y_1, \cdots, y_N)$ using the particular structure of this process.

Write $P(X_t|y_1, \cdots, y_{t-1})$ in terms of other factors. Construct an answer by checking the correct boxes below:

$P(X_t|y_1, \cdots, y_{t-1}) =$ ____**(i)**____ ____**(ii)**____ ____**(iii)**____

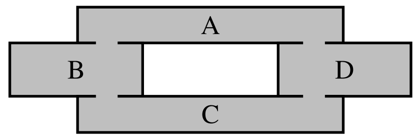
(i)
- ● $\sum_{i=1}^k \sum_{d=1}^M \sum_{d'=1}^M$
- ○ $\sum_{i=1}^k \sum_{d=1}^M$
- ○ $\sum_{i=1}^k$
- ○ $\sum_{d=1}^M$

(ii)
- ○ $P(Z_t = (X_t, d)|Z_{t-1} = (s_i, d))$
- ○ $P(X_t|X_{t-1} = s_i)$
- ○ $P(X_t|X_{t-1} = s_d)$
- ● $P(Z_t = (X_t, d')|Z_{t-1} = (s_i, d))$

(iii)
- ○ $P(Z_{t-1} = (s_d, i)|y_1, \cdots, y_{t-1})$
- ○ $P(X_{t-1} = s_d|y_1, \cdots, y_{t-1})$
- ● $P(Z_{t-1} = (s_i, d)|y_1, \cdots, y_{t-1})$
- ○ $P(X_{t-1} = s_i|y_1, \cdots, y_{t-1})$

Q2. HMMs: Help Your House Help You

Imagine you have a smart house that wants to track your location within itself so it can turn on the lights in the room you are in and make you food in your kitchen. Your house has 4 rooms (A, B, C, D) in the floorplan below (A is connected to B and D, B is connected to A and C, C is connected to B and D, and D is connected to A and C):



At the beginning of the day ($t = 0$), your probabilities of being in each room are p_A, p_B, p_C, and p_D for rooms A, B, C, and D, respectively, and at each time t your position (following a Markovian process) is given by X_t. At each time, your probability of staying in the same room is q_0, your probability of moving clockwise to the next room is q_1, and your probability of moving counterclockwise to the next room is $q_{-1} = 1 - q_0 - q_1$.

(a) Initially, assume your house has no way of sensing where you are. What is the probability that you will be in room D at time $t = 1$?

○ $q_0 p_D$ ○ $q_0 p_D + q_1 p_A + q_{-1} p_C + 2q_1 p_B$ ● $q_0 p_D + q_1 p_A + q_{-1} p_C$

○ $q_0 p_D + q_{-1} p_A + q_1 p_C$ ○ $q_1 p_A + q_1 p_C + q_0 p_D$ ○ None of these

This probability is given by the sum of three probabilities: 1) $q_0 p_D$: You are in room D to start (p_D) and stay there (q_0), 2) $q_1 p_A$: You are in room A to start (p_A) and move clockwise to room D (q_1), and 3) $q_{-1} p_C$: You are in room C to start (p_C) and move counterclockwise to room D (q_{-1}).

Now assume your house contains a sensor M^A that detects motion ($+m$) or no motion ($-m$) in room A. However, the sensor is a bit noisy and can be tricked by movement in adjacent rooms, resulting in the conditional distributions for the sensor given in the table below. The prior distribution for the sensor's output is also given.

M^A	$P(M^A \mid X = A)$	$P(M^A \mid X = B)$	$P(M^A \mid X = C)$	$P(M^A \mid X = D)$
$+m^A$	$1 - 2\gamma$	γ	0.0	γ
$-m^A$	2γ	$1 - \gamma$	1.0	$1 - \gamma$

M^A	$P(M^A)$
$+m^A$	0.5
$-m^A$	0.5

(b) You decide to help your house to track your movements using a particle filter with three particles. At time $t = T$, the particles are at $X^0 = A, X^1 = B, X^2 = D$. What is the probability that the particles will be resampled as $X^0 = X^1 = X^2 = A$ after time elapse? Select **all terms in the product.**

■ q_0 □ q_0^2 □ q_0^3 ■ q_1 □ q_1^2 □ q_1^3 ■ q_{-1} □ q_{-1}^2 □ q_{-1}^3 ○ None of these

The probability that all particles will be resampled as being in room A is $q_0 q_1 q_{-1}$ since particle X^0 stays in A with probability q_0, particle X^1 moves clockwise to A with probability q_1, and particle X^2 moves counterclockwise with probability q_{-1}.

(c) Assume that the particles are actually resampled after time elapse as $X^0 = D, X^1 = B, X^2 = C$, and the sensor observes $M^A = -m^A$. What are the particle weights given the observation?

Particle	Weight						
$X^0 = \text{D}$	○ γ	● $1 - \gamma$	○ $1 - 2\gamma$	○ 0.0	○ 1.0	○ 2γ	○ None of these
$X^1 = \text{B}$	○ γ	● $1 - \gamma$	○ $1 - 2\gamma$	○ 0.0	○ 1.0	○ 2γ	○ None of these
$X^2 = \text{C}$	○ γ	○ $1 - \gamma$	○ $1 - 2\gamma$	○ 0.0	● 1.0	○ 2γ	○ None of these

We can read these weights off of the tables given above. The weight for X^0 is given by $P(M^A = -m^A | X = D) = 1-\gamma$, the weight for X^1 is given by $P(M^A = -m^A | X = B) = 1-\gamma$, and the weight for X^2 is given by $P(M^A = -m^A | X = C) = 1$.

Now, assume your house also contains sensors M^B and M^D in rooms B and D, respectively, with the conditional distributions of the sensors given below and the prior equivalent to that of sensor M^A.

M^B	$P(M^B \mid X = A)$	$P(M^B \mid X = B)$	$P(M^B \mid X = C)$	$P(M^B \mid X = D)$
$+m^B$	γ	$1-2\gamma$	γ	0.0
$-m^B$	$1-\gamma$	2γ	$1-\gamma$	1.0

M^D	$P(M^D \mid X = A)$	$P(M^D \mid X = B)$	$P(M^D \mid X = C)$	$P(M^D \mid X = D)$
$+m^D$	γ	0.0	γ	$1-2\gamma$
$-m^D$	$1-\gamma$	1.0	$1-\gamma$	2γ

(d) Again, assume that the particles are actually resampled after time elapse as $X^0 = D, X^1 = B, X^2 = C$. The sensor readings are now $M^A = -m^A, M^B = -m^B, M^D = +m^D$. What are the particle weights given the observations?

Particle	Weight
$X^0 = \text{D}$	○ $\gamma^2 - 2\gamma^3$ ○ $3 - 2\gamma$ ○ 0.0 ○ $\gamma - \gamma^2 + \gamma^3$ ● $1 - 3\gamma + 2\gamma^2$ ○ $2 - \gamma$ ○ $1 - 2\gamma + \gamma^2$ ○ None of these
$X^1 = \text{B}$	○ $\gamma^2 - 2\gamma^3$ ○ $3 - 2\gamma$ ● 0.0 ○ $\gamma - \gamma^2 + \gamma^3$ ○ $1 - 3\gamma + 2\gamma^2$ ○ $2 - \gamma$ ○ $1 - 2\gamma + \gamma^2$ ○ None of these
$X^2 = \text{C}$	○ $\gamma^2 - 2\gamma^3$ ○ $3 - 2\gamma$ ○ 0.0 ○ $\gamma - \gamma^2 + \gamma^3$ ○ $1 - 3\gamma + 2\gamma^2$ ○ $2 - \gamma$ ○ $1 - 2\gamma + \gamma^2$ ● None of these

The weight for X^0 is given by $P(M^A = -m^A|X = D)P(M^B = -m^B|X = D)P(m^D = +m^D|X = D) = (1-\gamma)(1.0)(1-2\gamma) = 1 - 3\gamma + 2\gamma^2$, the weight for X^1 is given by $P(M^A = -m^A|X = B)P(M^B = -m^B|X = B)P(M^D = +m^D|X = B) = (1-\gamma)(2\gamma)(0.0) = 0.0$, and the weight for X^2 is given by $P(M^A = -m^A|X = C)P(M^B = -m^B|X = C)P(M^A = +m^D|X = C) = (1.0)(1-\gamma)(\gamma) = \gamma - \gamma^2$.

The sequence of observations from each sensor are expressed as the following: $m^A_{0:t}$ are all measurements $m^A_0, m^A_1, \dots, m^A_t$ from sensor M^A, $m^B_{0:t}$ are all measurements $m^B_0, m^B_1, \dots, m^B_t$ from sensor M^B, and $m^D_{0:t}$ are all measurements $m^D_0, m^D_1, \dots, m^D_t$ from sensor M^D. Your house can get an accurate estimate of where you are at a given time t using the forward algorithm. The forward algorithm update step is shown here:

$$P(X_t \mid m^A_{0:t}, m^B_{0:t}, m^D_{0:t}) \propto P(X_t, m^A_{0:t}, m^B_{0:t}, m^D_{0:t}) \tag{5}$$

$$= \sum_{x_{t-1}} P(X_t, x_{t-1}, m^A_t, m^B_t, m^D_t, m^A_{0:t-1}, m^B_{0:t-1}, m^D_{0:t-1}) \tag{6}$$

$$= \sum_{x_{t-1}} \underline{\hspace{3cm}} P(X_t \mid x_{t-1}) P(x_{t-1}, m^A_{0:t-1}, m^B_{0:t-1}, m^D_{0:t-1}) \tag{7}$$