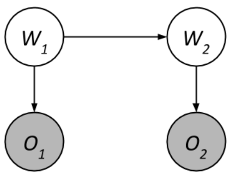
CS188 Fall 2018 Section 8: HMMs + Particle Filtering

1 HMMs

Consider the following Hidden Markov Model.



W_1	$P(W_1)$
0	0.3
1	0.7

W_t	W_{t+1}	$P(W_{t+1}\|W_t)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

W_t	O_t	$P(O_t\|W_t)$
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

Suppose that we observe $O_1 = A$ and $O_2 = B$.
Using the forward algorithm, compute the probability distribution $P(W_2|O_1 = A, O_2 = B)$ one step at a time.

1. Compute $P(W_1, O_1 = A)$.

   $P(W_1, O_1 = A) = P(W_1)P(O_1 = A|W_1)$
   $P(W_1 = 0, O_1 = A) = (0.3)(0.9) = 0.27$
   $P(W_1 = 1, O_1 = A) = (0.7)(0.5) = 0.35$

2. Using the previous calculation, compute $P(W_2, O_1 = A)$.

   $P(W_2, O_1 = A) = \sum_{x_1} P(x_1, O_1 = A)P(W_2|x_1)$
   $P(W_2 = 0, O_1 = A) = (0.27)(0.4) + (0.35)(0.8) = 0.388$
   $P(W_2 = 1, O_1 = A) = (0.27)(0.6) + (0.35)(0.2) = 0.232$

3. Using the previous calculation, compute $P(W_2, O_1 = A, O_2 = B)$.

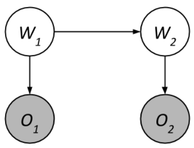
   $P(W_2, O_1 = A, O_2 = B) = P(W_2, O_1 = A)P(O_2 = B|W_2)$
   $P(W_2 = 0, O_1 = A, O_2 = B) = (0.388)(0.1) = 0.0388$
   $P(W_2 = 1, O_1 = A, O_2 = B) = (0.232)(0.5) = 0.116$

4. Finally, compute $P(W_2|O_1 = A, O_2 = B)$.

   Renormalizing the distribution above, we have
   $P(W_2 = 0|O_1 = A, O_2 = B) = 0.0388/(0.0388 + 0.116) \approx 0.25$
   $P(W_2 = 1|O_1 = A, O_2 = B) = 0.116/(0.0388 + 0.116) \approx 0.75$

2 Particle Filtering

Let's use Particle Filtering to estimate the distribution of $P(W_2|O_1 = A, O_2 = B)$. Here's the HMM again:



W_1	$P(W_1)$
0	0.3
1	0.7

W_t	W_{t+1}	$P(W_{t+1}|W_t)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

W_t	O_t	$P(O_t|W_t)$
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

We start with two particles representing our distribution for W_1.
$P_1 : W_1 = 0$
$P_2 : W_1 = 1$
Use the following random numbers to run particle filtering:

$$[0.22, 0.05, 0.33, 0.20, 0.84, 0.54, 0.79, 0.66, 0.14, 0.96]$$

1. **Observe**: Compute the weight of the two particles after evidence $O_1 = A$.

   $w(P_1) = P(O_t = A|W_t = 0) = 0.9$
   $w(P_2) = P(O_t = A|W_t = 1) = 0.5$

2. **Resample**: Using the random numbers, resample P_1 and P_2 based on the weights.

   We now sample from the weighted distribution we found above. After normalizing the weights, we find that P_1 maps to range [0, 0.643), and P_2 maps to range [0.643, 1). Using the first two random samples, we find:
   $P_1 = sample(weights, 0.22) = 0$
   $P_2 = sample(weights, 0.05) = 0$

3. **Elapse Time**: Now let's compute the elapse time particle update. Sample P_1 and P_2 from applying the time update.

   $P_1 = sample(P(W_{t+1}|W_t = 0), 0.33) = 0$
   $P_2 = sample(P(W_{t+1}|W_t = 0), 0.20) = 0$

4. **Observe**: Compute the weight of the two particles after evidence $O_2 = B$.

   $w(P_1) = P(O_t = B|W_t = 0) = 0.1$
   $w(P_2) = P(O_t = B|W_t = 0) = 0.1$

5. **Resample**: Using the random numbers, resample P_1 and P_2 based on the weights.

   Because both of our particles have $X = 0$, resampling will still leave us with two particles with $X = 0$.
   $P_1 = 0$
   $P_2 = 0$

6. What is our estimated distribution for $P(W_2|O_1 = A, O_2 = B)$?

$P(W_2 = 0|O_1 = A, O_2 = B) = 2/2 = 1$
$P(W_2 = 1|O_1 = A, O_2 = B) = 0/2 = 0$

3 HMMs (Optional)

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
  $$x_i = x_{i+1} = \cdots = x_{i+d-1} = s \tag{1}$$
  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.

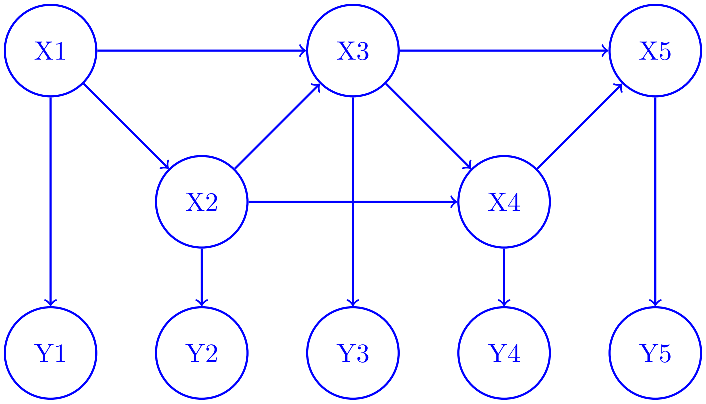
(a) Assuming that all three states s_1, s_2, s_3 are different, what is the probability of the sample sequence $s_1, s_1, s_2, s_2, s_2, s_3, s_3$? Write an algebraic expression. Assume $M \geq 3$.

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

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).

- What is $P(Z_1)$?

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

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

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

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 $t_{i+1} = 1$. (4)

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