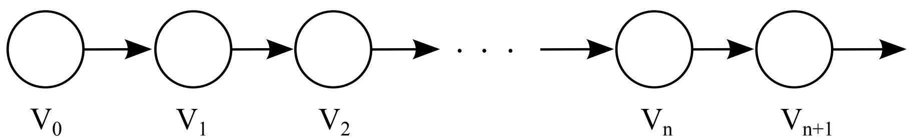
CS 188
Spring 2022

# Regular Discussion 9

## Q1. HMMs




**(a)** Consider a Markov model like the one above. For the first three parts of this problem, assume the domain of our variables is {a, b}. Fill in the table below with the probability of being in each state after a large number of transitions, when $P(V_n) = P(V_{n+1})$. If the values never reach a point when $P(V_n) = P(V_{n+1})$, write 'None'.

In the left part of the table, assume that we start with a uniform distribution ($P(V_0 = \text{a}) = P(V_0 = \text{b}) = 0.5$). In the right part of the table, assume that we start with the distribution that has $P(V_0 = \text{a}) = 1.0$.

<table>
<tr><th rowspan="2">Transition Probabilities</th><th colspan="2">$P(V_n)$ given that $P(V_0)$ is uniform</th><th colspan="2">$P(V_n)$ given that $P(V_0 = \text{a}) = 1.0$</th></tr>
<tr><th>$V_n = \text{a}$</th><th>$V_n = \text{b}$</th><th>$V_n = \text{a}$</th><th>$V_n = \text{b}$</th></tr>
<tr><td><table><tr><th></th><th colspan="2">$P(V_i \mid V_{i-1})$</th></tr><tr><th>$V_{i-1}$</th><th>$V_i = \text{a}$</th><th>$V_i = \text{b}$</th></tr><tr><td>a</td><td>0.5</td><td>0.5</td></tr><tr><td>b</td><td>0.5</td><td>0.5</td></tr></table></td><td></td><td></td><td></td><td></td></tr>
<tr><td><table><tr><th></th><th colspan="2">$P(V_i \mid V_{i-1})$</th></tr><tr><th>$V_{i-1}$</th><th>$V_i = \text{a}$</th><th>$V_i = \text{b}$</th></tr><tr><td>a</td><td>0.9</td><td>0.1</td></tr><tr><td>b</td><td>0.3</td><td>0.7</td></tr></table></td><td></td><td></td><td></td><td></td></tr>
<tr><td><table><tr><th></th><th colspan="2">$P(V_i \mid V_{i-1})$</th></tr><tr><th>$V_{i-1}$</th><th>$V_i = \text{a}$</th><th>$V_i = \text{b}$</th></tr><tr><td>a</td><td>0.0</td><td>1.0</td></tr><tr><td>b</td><td>1.0</td><td>0.0</td></tr></table></td><td></td><td></td><td></td><td></td></tr>
</table>

For this part our variables have the domain {a, b, c}. Fill in the table at the bottom with the probability of being in each state after a large number of transitions, when $P(V_n) = P(V_{n+1})$. In the left part of the table, assume that we start with a uniform distribution ($P(V_0 = \text{a}) = P(V_0 = \text{b}) = P(V_0 = \text{c}) = \frac{1}{3}$). In the right part of the table, assume that we start with the distribution that has $P(V_0 = \text{a}) = 1.0$.

| | $P(V_i \mid V_{i-1})$ | | |
|---|---|---|---|
| $V_{i-1}$ | $V_i = \text{a}$ | $V_i = \text{b}$ | $V_i = \text{c}$ |
| a | 0.5 | 0.5 | 0.0 |
| b | 0.5 | 0.5 | 0.0 |
| c | 0.0 | 0.0 | 1.0 |

| $P(V_n)$ given that $P(V_0)$ is uniform | | | $P(V_n)$ given that $P(V_0 = \text{a}) = 1.0$ | | |
|---|---|---|---|---|---|
| a | b | c | a | b | c |
| | | | | | |

Now we will consider a Hidden Markov Model, and look at properties of the Viterbi Algorithm. The Viterbi algorithm finds the most probable sequence of hidden states $X_{1:S}$ given a sequence of observations $y_{1:S}$. Recall that for the canonical HMM structure, the Viterbi algorithm performs the following update at each time step:

$$m_t[x_t] = P(y_t|x_t) \max_{x_{t-1}}[P(x_t|x_{t-1})m_{t-1}[x_{t-1}]]$$

Assume we have an HMM where:

- The hidden variable $X$ can take on $H$ values
- The (observed) emission variable $Y$ can take on $E$ values
- Our sequence has $S$ steps

**(b)** **(i)** What is the run time of the Viterbi algorithm?

○ $O(SEH)$ ○ $O(SEH^2)$ ○ $O(SH^2)$
○ $O(SH)$ ○ $O(EH)$ ○ $O(EH^2)$
○ $O(SH^2 + SEH)$

Ignoring the storage of the emission probabilities, $P(Y_t|X_t)$, and the transition probabilities, $P(X_t|X_{t-1})$, what are the storage requirements of the Viterbi algorithm?

○ $O(S)$ ○ $O(E)$ ○ $O(H)$
○ $O(SH)$ ○ $O(SE)$ ○ $O(EH)$
○ $O(S+H)$ ○ $O(S+E)$ ○ $O(E+H)$
○ $O(SEH)$

Now, assume that most of the transitions in our HMM have probability zero. In particular, suppose that for any given hidden state value, there are only $K$ possible next state values for which the transition probability is non-zero. To exploit this sparsity, we change the Viterbi Algorithm to only consider the non-zero transition edges during each max computation inside each update. You can think of this as the Viterbi algorithm ignoring edges that correspond to zero probability transitions in the transition lattice diagram.

**(ii)** What is the run time of this modified algorithm?

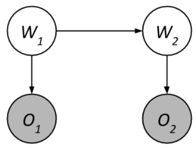
○ $O(SEH)$ ○ $O(SEH^2)$ ○ $O(SH^2)$
○ $O(SH)$ ○ $O(EH)$ ○ $O(EH^2)$
○ $O(SH^2 + SEH)$
○ $O(SEK)$ ○ $O(SEHK)$ ○ $O(SHK)$
○ $O(SK)$ ○ $O(EK)$ ○ $O(EHK)$
○ $O(SK + SEK)$ ○ $O(SHK + SEK)$

Ignoring the storage of the emission probabilities, $P(Y_t|X_t)$, and the transition probabilities, $P(X_t|X_{t-1})$, what are the storage requirements of this modified Viterbi algorithm?

○ $O(S)$ ○ $O(E)$ ○ $O(H)$
○ $O(SH)$ ○ $O(SE)$ ○ $O(EH)$
○ $O(S+H)$ ○ $O(S+E)$ ○ $O(E+H)$
○ $O(SEH)$
○ $O(K)$ ○ $O(SK)$ ○ $O(EK)$
○ $O(HK)$ ○ $O(S+K)$ ○ $O(E+K)$
○ $O(H+K)$ ○ $O(SEK)$

# 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:

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

**(a) Observe**: Compute the weight of the two particles after evidence $O_1 = a$.

**(b) Resample**: Using the random numbers, resample $P_1$ and $P_2$ based on the weights.

**(c) Predict**: Sample $P_1$ and $P_2$ from applying the time update.

**(d) Update**: Compute the weight of the two particles after evidence $O_2 = b$.

**(e) Resample**: Using the random numbers, resample $P_1$ and $P_2$ based on the weights.

**(f)** What is our estimated distribution for $P(W_2|O_1 = a, O_2 = b)$?