

Chapter 5

Hidden Markov Model

5.1.a The transition probability matrix A has a left-right structure, thus the word model is a left-right HMM. It is not ergodic since previous states of the Markov chain affect long-term behavior—if you for instance are in state 3 at time t , you can never return to state 1 or 2 at any future time $t + n$, regardless of how large $n > 0$ is.

5.1.b When applying the Forward Algorithm, many terms will be zero due to the left-right structure of the HMM, and the fact that certain observations are impossible for some states, leading to zeroes in B . For brevity, we will only show the computations for the nonzero elements.

For $t = 1$ we perform an initialization step following the equations

$$\alpha_{j,1}^{temp} = q_j b_j(z_1); \quad c_1 = \sum_{k=1}^N \alpha_{k,1}^{temp}; \quad \hat{\alpha}_{j,1} = \frac{\alpha_{j,1}^{temp}}{c_1}$$

This gives

$$\alpha_{1,1}^{temp} = q_1 b_1(1) = 1 \cdot 1 = 1; \quad \alpha_{2,1}^{temp} = \alpha_{3,1}^{temp} = 0$$

$$c_1 = \sum_{k=1}^3 \alpha_{k,1}^{temp} = 1 + 0 + 0 = 1$$

so

$$\hat{\alpha}_{1,1} = \frac{1}{1} = 1; \quad \hat{\alpha}_{2,1} = \hat{\alpha}_{3,1} = 0$$

For $t = 2, 3, \dots$ we perform forward steps according to the formulas

$$\alpha_{j,t}^{temp} = b_j(z_t) \sum_{i=1}^N \hat{\alpha}_{i,t-1} a_{ij}; \quad c_t = \sum_{k=1}^N \alpha_{k,t}^{temp}; \quad \hat{\alpha}_{j,t} = \frac{\alpha_{j,t}^{temp}}{c_t}$$

For $t = 2$ we get

$$\begin{aligned}\alpha_{2,2}^{temp} &= b_2(2) \sum_{i=1}^3 \hat{\alpha}_{i,1} a_{i2} \\ &= 0.5(1 \cdot 0.7 + 0 \cdot 0.5 + 0 \cdot 0) \\ &= 0.35\end{aligned}$$

while

$$\alpha_{1,2}^{temp} = 0; \quad \alpha_{3,2}^{temp} = 0$$

so

$$c_2 = 0.35; \quad \hat{\alpha}_{2,2} = 1; \quad \hat{\alpha}_{1,2} = \hat{\alpha}_{3,2} = 0$$

For $t = 3$

$$\begin{aligned}\alpha_{2,3}^{temp} &= 0.1(0 \cdot 0.7 + 1 \cdot 0.5 + 0 \cdot 0) = 0.05 \\ \alpha_{3,3}^{temp} &= 0.6(0 \cdot 0 + 1 \cdot 0.5 + 0 \cdot 1) = 0.3 \\ c_3 &= 0.35; \quad \hat{\alpha}_{1,3} = 0; \quad \hat{\alpha}_{2,3} = \frac{1}{7}; \quad \hat{\alpha}_{3,3} = \frac{6}{7}\end{aligned}$$

For $t = 4$

$$\begin{aligned}\alpha_{2,4}^{temp} &= 0.1(0 + \frac{1}{7}0.5 + \frac{6}{7}0) = \frac{1}{140} \\ \alpha_{3,4}^{temp} &= 0.6(0 + \frac{1}{7}0.5 + \frac{6}{7}1) = \frac{78}{140} \\ c_4 &= \frac{79}{140}; \quad \hat{\alpha}_{1,4} = 0; \quad \hat{\alpha}_{2,4} = \frac{1}{79}; \quad \hat{\alpha}_{3,4} = \frac{78}{79}\end{aligned}$$

Finally, for $t = 5$ we get

$$\begin{aligned}\alpha_{3,5}^{temp} &= 0.1(0 + \frac{1}{79}0.5 + \frac{78}{79}1) \approx 0.0994 \\ c_5 &= 0.0994; \quad \hat{\alpha}_{1,5} = \hat{\alpha}_{2,5} = 0; \quad \hat{\alpha}_{3,5} = 1\end{aligned}$$

The complete table becomes

t	1	2	3	4	5
$\hat{\alpha}_{1,t}$	1	0	0	0	0
$\hat{\alpha}_{2,t}$	0	1	1/7	1/79	0
$\hat{\alpha}_{3,t}$	0	0	6/7	78/79	1
c_t	1	0.35	0.35	79/140	0.0994

and the requested probability can be calculated as

$$\begin{aligned}P(\underline{Z} = \underline{z}|\lambda) &= P(z_1 z_2 z_3 z_4 z_5 | \lambda) \\ &= P(\{1, 2, 4, 4, 1\} | \lambda) \\ &= P(1|\lambda)P(2|1, \lambda)P(4|2, 1, \lambda)P(4|4, 2, 1, \lambda)P(1|4, 4, 2, 1, \lambda) \\ &= c_1 c_2 c_3 c_4 c_5 \\ &= 1 \times 0.35 \times 0.35 \times 0.5641 \dots \times 0.0994 \dots \\ &\approx 0.0069\end{aligned}$$

5.1.c This problem can be solved by counting all the possible state sequences. Since we have a three-state HMM, we always start in state 1 at $t = 1$, and can only stay in the current state i or jump to the next state $i + 1 \leq 3$, there can be at the most two transitions in any of the possible sequences.

There is only one sequence with no transitions; the Markov chain then stays in state 1 all the time. There is also only one sequence where the transition to state 2 occurs at $t = 5$; there is no time for additional transitions. There are two sequences where the transition to state 2 occurs at $t = 4$; the Markov chain can either stay in state 2 or move to state 3 at $t = 5$. Similarly, there are three sequences where the transition to state 2 occurs at $t = 3$, and four sequences where the transition to state 2 occurs at $t = 2$. Summing up, there are $1 + 1 + 2 + 3 + 4 = 11$ possible state sequences.

5.1.d An efficient way to determine the most likely hidden state sequence given Markov chain parameters $\lambda = \{q, A, B\}$ and a set of observations $\underline{z} = (1, 2, 4, 4, 1)$ is to use the Viterbi algorithm, which has many similarities to the Forward Algorithm. For brevity we shall only show computations for the nonzero terms that are calculated using this procedure.

For $t = 1$ we initialize the Viterbi partial-sequence vector elements as

$$\chi_{j,1} = q_j b_j(z_1)$$

giving

$$\chi_{1,1} = q_1 b_1(1) = 1 \cdot 1 = 1; \quad \chi_{2,1} = \chi_{3,1} = 0$$

For $t = 2, 3, \dots$ we apply Viterbi Forward steps

$$\chi_{j,t} = b_j(z_t) \max_i \chi_{i,t-1} a_{ij}; \quad \zeta_{j,t} = \operatorname{argmax}_i \chi_{i,t-1} a_{ij}$$

For $t = 2$, in particular, we get

$$\chi_{2,2} = 0.5 \max_i \{1 \cdot 0.7, 0, 0\} = 0.35; \quad \chi_{1,2} = \chi_{3,2} = 0$$

and (since $\chi_{1,1}$ is the only nonzero term from the previous time step)

$$\zeta_{1,2} = \zeta_{2,2} = 1$$

The calculation of $\zeta_{3,2}$ is ambiguous since all terms in the argmax_i are zero. However, this simply means that we can choose any of the legal i -values.

For $t = 3$ we get

$$\chi_{2,3} = 0.1 \max_i \{0, 0.35 \cdot 0.5, 0\} = 0.0175$$

$$\chi_{3,3} = 0.6 \max_i \{0, 0.35 \cdot 0.5, 0\} = 0.105$$

and (since $\chi_{2,2}$ is the only nonzero term from the previous time step)

$$\zeta_{2,3} = \zeta_{3,3} = 2$$

For $t = 4$ we get

$$\chi_{2,4} = 0.1 \max_i \{0, 0.0175 \cdot 0.5, 0.105 \cdot 0\} = 8.75 \cdot 10^{-4}$$

$$\chi_{3,4} = 0.6 \max_i \{0, 0.0175 \cdot 0.5, 0.105 \cdot 1\} = 0.6 \cdot 0.105 \cdot 1 = 0.063$$

and the corresponding indices for the maxima

$$\zeta_{2,4} = 2; \quad \zeta_{3,4} = 3$$

Finally, at $t = 5$ we get $\chi_{2,5} = 0$ since $b_2(1) = 0$, while

$$\chi_{3,5} = 0.1 \max_i \{0, 8.75 \cdot 10^{-4} \cdot 0.5, 0.063 \cdot 1\} = 0.0063$$

The backpointer variables are similar to before,

$$\zeta_{2,4} = 2; \quad \zeta_{3,4} = 3$$

Like with the Forward Algorithm, we can collect the results in a table

t	1	2	3	4	5
$\chi_{1,t}$	1	0	0	0	0
$\chi_{2,t}$	0	0.35	0.0175	$8.75 \cdot 10^{-4}$	0
$\chi_{3,t}$	0	0	0.105	0.063	0.0063
$\zeta_{1,t}$	N/A	1			
$\zeta_{2,t}$	N/A	1	2	2	2
$\zeta_{3,t}$	N/A		2	3	3

Since the $\chi_{3,5}$ is the greatest partial-sequence probability vector element at $t = 5$, we follow the path given by the backpointers using $s_t^* = \zeta_{s_t^*+1, t+1}^*$ starting from $\zeta_{3,5}$ (bolded in the table). We then recover the most likely hidden state sequence, $\underline{S}^* = \{1, 2, 3, 3, 3\}$.

5.3.a

$$P(S_{19} = j | x_1 \dots x_{19}) = \frac{P(S_{19} = j, X_{19} = x_{19} | x_1 \dots x_{18})}{P(X_{19} = x_{19} | x_1 \dots x_{18})}$$

$$P(X_{19} = x_{19} | x_1 \dots x_{18}) = \sum_j P(S_{19} = j, X_{19} = x_{19} | x_1 \dots x_{18})$$

$$\begin{aligned}
P(S_{19} = j, X_{19} = x_{19} | x_1 \dots x_{18}) &= P(X_{19} = x_{19} | S_{19} = j, x_1 \dots x_{18}) P(S_{19} = j | x_1 \dots x_{18}) \\
&= P(X_{19} = x_{19} | S_{19} = j) P(S_{19} = j | x_1 \dots x_{18}) \\
&= b_j(x_{19}) P(S_{19} = j | x_1 \dots x_{18}) \\
&= b_{j3} P(S_{19} = j | x_1 \dots x_{18})
\end{aligned}$$