

SUTD 2021 50.007 Homework 4

James Raphael Tiovalen 1004555

Hidden Markov Model

Question 1

The transition probabilities are estimated as:

$$a_{u,v} = \frac{\text{Count}(u,v)}{\text{Count}(u)}$$

$u \backslash v$	X	Y	Z	STOP
START	2/4	0	2/4	0
X	0	2/5	2/5	1/5
Y	1/5	0	1/5	3/5
Z	2/5	3/5	0	0

The emission probabilities are estimated as:

$$b_u(o) = \frac{\text{Count}(u \rightarrow o)}{\text{Count}(u)}$$

$u \backslash o$	a	b	c
X	2/5	3/5	0
Y	1/5	0	4/5
Z	1/5	3/5	1/5

Question 2

Base case:

$$\pi(0, \text{START}) = 1, \text{ otherwise } \pi(0, v) = 0 \text{ if } v \neq \text{START}$$

Moving forward recursively:

$$k = 1$$

$$\pi(1, X) = \pi(0, \text{START}) \times a_{\text{START}, X} \times b_X(\mathbf{b}) = 2/4 \times 3/5 = 3/10$$

$$\pi(1, Y) = \pi(0, \text{START}) \times a_{\text{START}, Y} \times b_Y(\mathbf{b}) = 0$$

$$\pi(1, Z) = \pi(0, \text{START}) \times a_{\text{START}, Z} \times b_Z(\mathbf{b}) = 2/4 \times 3/5 = 3/10$$

$$k = 2$$

$$\begin{aligned}
\pi(2, X) &= \max_{u \in \mathcal{T}} \{\pi(1, u) \times a_{u,X} \times b_X(\mathbf{c})\} \\
&= \max\{3/10 \times 0 \times 0, 0, 3/10 \times 2/5 \times 0\} \\
&= 0 \\
\pi(2, Y) &= \max_{u \in \mathcal{T}} \{\pi(1, u) \times a_{u,Y} \times b_Y(\mathbf{c})\} \\
&= \max\{3/10 \times 2/5 \times 4/5, 0, 3/10 \times 3/5 \times 4/5\} \\
&= 18/125 \\
\pi(2, Z) &= \max_{u \in \mathcal{T}} \{\pi(1, u) \times a_{u,Z} \times b_Z(\mathbf{c})\} \\
&= \max\{3/10 \times 2/5 \times 1/5, 0, 3/10 \times 0 \times 1/5\} \\
&= 3/125
\end{aligned}$$

$$k = 3$$

$$\begin{aligned}
\pi(3, \text{STOP}) &= \max_{u \in \mathcal{T}} \{\pi(2, u) \times a_{u,\text{STOP}}\} \\
&= \max\{0 \times 1/5, 18/125 \times 3/5, 3/125 \times 0\} \\
&= 54/625
\end{aligned}$$

Backtracking:

$$\begin{aligned}
y_2^* &= \arg \max_{v \in \mathcal{T}} \{\pi(2, v) \times a_{v,\text{STOP}}\} = Y \\
y_1^* &= \arg \max_{v \in \mathcal{T}} \{\pi(1, v) \times a_{v,Y}\} = Z
\end{aligned}$$

Therefore, the optimal sequence is: Z, Y .

Question 3

The set of possible states is $\{\text{START}, X, Y, Z, \text{STOP}\}$. In an HMM, for a sequence of length n , we have

$$\begin{aligned}
&P(x_1, \dots, x_{i-1}, y_i = u, x_i, \dots, x_n) \\
&= P(x_1, \dots, x_{i-1}, y_i = u) \times P(x_i, \dots, x_n | y_i = u) \\
&= \alpha_u(i) \beta_u(i),
\end{aligned}$$

where:

$$\begin{aligned}
\alpha_u(i) &= P(x_1, \dots, x_{i-1}, y_i = u), \\
\beta_u(i) &= P(x_i, \dots, x_n | y_i = u)
\end{aligned}$$

As shown in class:

$$P(s_i = u | o_1 = \mathbf{b}, o_2 = \mathbf{c}) = \frac{\alpha_u(i)\beta_u(i)}{\sum_v \alpha_v(j)\beta_v(j)},$$

with any choice of $j \in \{1, 2\}$.

Hence, to calculate the marginal distribution probability term, we will use the Forward-Backward Algorithm.

Forward

- Base case:

$$\begin{aligned}\alpha_X(1) &= a_{\text{START},X} = 2/4 \\ \alpha_Y(1) &= a_{\text{START},Y} = 0 \\ \alpha_Z(1) &= a_{\text{START},Z} = 2/4\end{aligned}$$

- Moving forward recursively:

$$\begin{aligned}\alpha_X(2) &= \sum_v \alpha_v(1) \times a_{v,X} \times b_v(\mathbf{b}) \\ &= (2/4 \times 0 \times 3/5) + 0 + (2/4 \times 2/5 \times 3/5) = 3/25 \\ \alpha_Y(2) &= \sum_v \alpha_v(1) \times a_{v,Y} \times b_v(\mathbf{b}) \\ &= (2/4 \times 2/5 \times 3/5) + 0 + (2/4 \times 3/5 \times 3/5) = 3/10 \\ \alpha_Z(2) &= \sum_v \alpha_v(1) \times a_{v,Z} \times b_v(\mathbf{b}) \\ &= (2/4 \times 2/5 \times 3/5) + 0 + (2/4 \times 0 \times 3/5) = 3/25\end{aligned}$$

Backward

- Base case:

$$\begin{aligned}\beta_X(2) &= a_{X,\text{STOP}} \times b_X(\mathbf{c}) = 1/5 \times 0 = 0 \\ \beta_Y(2) &= a_{Y,\text{STOP}} \times b_Y(\mathbf{c}) = 3/5 \times 4/5 = 12/25 \\ \beta_Z(2) &= a_{Z,\text{STOP}} \times b_Z(\mathbf{c}) = 0 \times 1/5 = 0\end{aligned}$$

- Moving backward recursively:

$$\begin{aligned}\beta_X(1) &= \sum_v a_{X,v} \times b_X(\mathbf{b}) \times \beta_v(2) \\ &= 0 + (2/5 \times 3/5 \times 12/25) + 0 = 72/625 \\ \beta_Y(1) &= \sum_v a_{Y,v} \times b_Y(\mathbf{b}) \times \beta_v(2) \\ &= 0 + (0 \times 0 \times 12/25) + 0 = 0 \\ \beta_Z(1) &= \sum_v a_{Z,v} \times b_Z(\mathbf{b}) \times \beta_v(2) \\ &= 0 + (3/5 \times 3/5 \times 12/25) + 0 = 108/625\end{aligned}$$

Thus:

$$\sum_v \alpha_v(j) \beta_v(j) = (2/4 \times 72/625) + (2/4 \times 108/625) = (3/10 \times 12/25) = 18/125$$

Hence, the marginal distribution table would be:

$u \backslash i$	1	2
X	$0.0576/0.144 = 2/5$	0
Y	0	$0.144/0.144 = 1$
Z	$0.0864/0.144 = 3/5$	0

Taking the maximum marginal probability value for each column i from the table:

$$\begin{aligned} s_1^* &= \arg \max_{s_1} P(s_1 | o_1 = \mathbf{b}, o_2 = \mathbf{c}) = \arg \max_{s_1} P(s_1, o_1 = \mathbf{b}, o_2 = \mathbf{c}) = Z \\ s_2^* &= \arg \max_{s_2} P(s_2 | o_1 = \mathbf{b}, o_2 = \mathbf{c}) = \arg \max_{s_2} P(s_2, o_1 = \mathbf{b}, o_2 = \mathbf{c}) = Y \end{aligned}$$

Therefore, we would also consistently obtain and arrive at the optimal sequence of: Z, Y .

Question 4

Assume we have a set of possible states $\{0, 1, \dots, N-1, N\}$, where $0 = \text{START}$ and $N = \text{STOP}$.

$$\begin{aligned} &P(x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}, z_i = u, x_i, \dots, x_n, y_i, \dots, y_n; \theta) \\ &= P(x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}, z_i = u; \theta) \times P(x_i, \dots, x_n, y_i, \dots, y_n | z_i = u; \theta) \\ &= \alpha_u(i) \beta_u(i), \end{aligned}$$

where

$$\begin{aligned} \alpha_u(i) &= P(x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}, z_i = u; \theta), \\ \beta_u(i) &= P(x_i, \dots, x_n, y_i, \dots, y_n | z_i = u; \theta) \end{aligned}$$

Forward

- Base case:

$$\alpha_u(1) = a_{\text{START},u}, \quad \forall u \in \{1, \dots, N-1\}$$

- Moving forward recursively, for $i = 2, \dots, n$:

$$\alpha_u(i+1) = \sum_v \alpha_v(i) \times a_{v,u} \times b_v(x_i) \times c_{x_i}(y_i),$$

where

$$\begin{aligned} b_v(x) &= P(x|v), \\ c_x(y) &= P(y|x) \end{aligned}$$

Backward

- Base case:

$$\beta_u(n) = a_{u,\text{STOP}} \times b_u(x_n) \times c_{x_n}(y_n), \forall u \in \{1, \dots, N-1\}$$

- Moving backward recursively, for $i = n-1, \dots, 1$:

$$\beta_u(i) = \sum_v a_{u,v} \times b_u(x_i) \times c_{x_i}(y_i) \times \beta_v(i+1)$$

At each time step/position, there are N forward (α) and N backward (β) terms to compute. Hence, to compute each term, there are $O(N)$ operations. Thus, at each time step/position, there are $O(N^2)$ operations. Therefore, assuming that the length of the sentence is n , which is the number of different time steps/positions, the total time complexity for this algorithm is $O(nN^2)$.