Yeo Ying Xuan
1003835

50.007 Machine Learning
Homework 3

Q1
A HMM is defined by a tuple $< \daleth, \sigma, \theta >$, where:
- $\daleth$ is the set of states including the $START$ state and $STOP$ state: $0, 1, \ldots, |\daleth| - 1$. By convention, we always assume 0 is the start state $START$, and $|\daleth| - 1$ is the stop state $STOP$.
    - $\daleth = \{START, X, Y, Z, STOP\}$
- σ is the set of observation symbols
    - $\sigma = \{"a", "b", "c", "d"\}$
- $\theta =< a\,, b\ >$ consists of two sets of parameters:
    - Parameter $a_{u,v} \equiv p(y_{next} = v | y_{curr} = u)\ for\ u = 0,1, \ldots, |\daleth| - 2$ and $v = 1, \ldots, |\daleth| - 1$ is the probability of transitioning from state u to state v: $\sum_{v\prime=1}^{|\daleth|-1} a_{u,v\prime} = 1\ \ for\ any\ u.$
    - Parameter $b_u(o) \equiv p(x = o | y = u) for u = 0,1, \ldots, |\daleth| - 2$ and $o \in \sigma$ is the probability of emitting symbol o from state u: $\sum_{o\, \in\, \sigma} b_u(o) = 1\ for\ any\ u.$

Parameters are transition probability $a_{u,v}$ and emission probability $b_u(o)$

$$count(X) = 6$$
$$count(Y) = 4$$
$$count(Z) = 8$$

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

$u \backslash v$	X	Y	Z	$STOP$
$START$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
X	$\frac{1}{6}$	0	$\frac{2}{3}$	$\frac{1}{6}$
Y	$\frac{1}{4}$	0	0	$\frac{3}{4}$
Z	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{4}$

$$b_u(o) = \frac{count(u \to o)}{count(u)}$$

$u \backslash o$	a	b	c	d
X	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
Y	$\frac{1}{2}$	0	$\frac{1}{4}$	$\frac{1}{4}$
Z	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Q2

Base case

$$\pi(0, v) = \begin{cases} 1, & if\ v = \ START \\ 0, & otherwise \end{cases}$$

Moving forward recursively, for any $k \in 1, \ldots, n$

Given $n \ = \ 2$,

$k = 1$

$$\begin{aligned} \pi(1, X) &= \pi(0, START) \cdot a_{START,X} \cdot b_Y(a) \\ &= 1 \cdot \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{6} \end{aligned}$$

$$\begin{aligned} \pi(1, Y) &= \pi(0, START) \cdot a_{START,Y} \cdot b_Y(a) \\ &= 1 \cdot 0 \cdot \frac{1}{2} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \pi(1, Z) &= \pi(0, START) \cdot a_{START,Z} \cdot b_Z(a) \\ &= 1 \cdot \frac{1}{2} \cdot \frac{1}{8} \\ &= \frac{1}{16} \end{aligned}$$

$k = 2$

$$\begin{aligned} \pi(2, X) &= \max_u \{\pi(1, u) \cdot a_{u,X} \cdot b_X(d)\} \\ &= \max \{\pi(1, X) \cdot a_{X,X} \cdot b_X(d), \pi(1, Y) \cdot a_{Y,X} \cdot b_X(d), \pi(1, Z) \cdot a_{Z,X} \cdot b_X(d)\} \\ &= max\ \{\frac{1}{6} \cdot \frac{1}{6}\ \cdot \frac{1}{6}, 0 \cdot \frac{1}{4} \cdot \frac{1}{6}, \frac{1}{16} \cdot \frac{1}{8} \cdot \frac{1}{6}\} \\ &= \frac{1}{216} \end{aligned}$$

$$\begin{aligned} \pi(2, Y) &= \max_u \{\pi(1, u) \cdot a_{u,Y} \cdot b_Y(d)\} \\ &= \max \{\pi(1, X) \cdot a_{X,Y} \cdot b_Y(d), \pi(1, Y) \cdot a_{Y,Y} \cdot b_Y(d), \pi(1, Z) \cdot a_{Z,Y} \cdot b_Y(d)\} \\ &= max\ \{\frac{1}{6} \cdot 0\ \cdot \frac{1}{4}, 0 \cdot 0 \cdot \frac{1}{4}, \frac{1}{16} \cdot \frac{1}{2} \cdot \frac{1}{4}\} \\ &= \frac{1}{128} \end{aligned}$$

$$
\begin{aligned}
\pi(2,Z) &= \max_u\{\pi(1,u)\cdot a_{u,Z}\cdot b_Z(d)\}\\
&= \max\{\pi(1,X)\cdot a_{X,Z}\cdot b_Z(d), \pi(1,Y)\cdot a_{Y,Z}\cdot b_Z(d), \pi(1,Z)\cdot a_{Z,Z}\cdot b_Z(d)\}\\
&= max\{\frac{1}{6}\cdot\frac{2}{3}\cdot\frac{1}{8}, 0\cdot 0\cdot\frac{1}{8}, \frac{1}{16}\cdot\frac{1}{8}\cdot\frac{1}{6}\}\\
&= \frac{1}{72}
\end{aligned}
$$

Transition from y_n to $STOP$

$$
\begin{aligned}
\pi(3,STOP) &= \max_v\{\pi(2,v)\cdot a_{v,STOP}\}\\
&= max\{\pi(2,X)\cdot a_{X,STOP}, \pi(2,Y)\cdot a_{Y,STOP}, \pi(2,Z)\cdot a_{Z,STOP}\}\\
&= max\{\frac{1}{216}\cdot\frac{1}{6}, \frac{1}{128}\cdot\frac{3}{4}, \frac{1}{72}\cdot\frac{1}{4}\}\\
&= \frac{3}{512}
\end{aligned}
$$

Backtracking

$$
\begin{aligned}
y_2^* &= \arg\max_v\{\pi(2,v)\cdot a_{v,STOP}\}\\
&= \arg\max_v\{\pi(2,X)\cdot a_{X,STOP}, \pi(2,Y)\cdot a_{Y,STOP}, \pi(2,Z)\cdot a_{Z,STOP}\}\\
&= \arg\max_v\{\frac{1}{216}\cdot\frac{1}{6}, \frac{1}{128}\cdot\frac{3}{4}, \frac{1}{72}\cdot\frac{1}{4}\}\\
&= \arg\max_v\{\frac{1}{1296}, \frac{3}{512}, \frac{1}{288}\}\\
&= Y
\end{aligned}
$$

$$
\begin{aligned}
y_1^* &= \arg\max_u\{\pi(1,u)\cdot a_{u,Y}\}\\
&= \arg\max_u\{\pi(1,X)\cdot a_{X,Y}, \pi(1,Y)\cdot a_{Y,Y}, \pi(1,Z)\cdot a_{Z,Y}\}\\
&= \arg\max_u\{\frac{1}{6}\cdot 0, 0\cdot 0, \frac{1}{16}\cdot\frac{1}{2}\}\\
&= \arg\max_u\{0, 0, \frac{1}{32}\}\\
&= Z
\end{aligned}
$$

Optimal state sequence: Z, Y

Q3

$$p(x_1, \dots, x_n, y_1, \dots, y_n) = \prod_{i=1}^{n+1} p(y_i | y_{i-2}, y_{i-1}) \times \prod_{i=1}^{n} p(x_i | y_i)$$

$$y_{-1} = y_0 = START$$
$$y_{n+1} = STOP$$

Looking at a truncated version of the joint probability, focusing on the first k tags for any $k \in \{1, \dots, n\}$:

$$r(y_1, \dots, y_k) = \prod_{i=1}^{k} a_{y_{i-2}, y_{i-1}, y_i} \times \prod_{i=1}^{k} b_{y_i}(x_i)$$
$$where\ k \neq n+1,$$

$$a_{t,u,v} = \frac{count(t,u,v)}{count(t,u)}$$

$$b_v(o) = \frac{count(v \to o)}{count(v)}$$

$t, u, v \in \mathcal{T}$, where $\mathcal{T}$ is the set of states including the $START$ and $STOP$ states.
$\mathcal{O}$ is the set of observation symbols.

$a_{t,u,v}$ is the probability of transitioning from state t to state u to state v
$b_v(o)$ is the probability of emitting observation symbol o from state v

Hence,

$$\begin{aligned} & p(x_1, \dots, x_n, y_0, y_1, \dots, y_{n+1}) \\ & = r(y_1, \dots, y_k) \cdot a_{y_{n-1}, y_n, y_{n+1}} \\ & = r(y_1, \dots, y_n) \cdot a_{y_{n-1}, y_n, STOP} \end{aligned}$$

For any $k \in \{1, \dots, n\}$, let $S(k, u, v)$ be the set of tag sequences $y_1, \dots, y_k$ such that $y_{k-1} = u$, $y_k = v$. In other words, $S(k, u, v)$ is a set of all sequences of length k whose last two tags are in the order of u, v.

Forward recursive algorithm:

$$\pi(k, u, v) = \max_{(y_1, \dots, y_k) \in S(k,u,v)} r(y_1, \dots, y_k)$$

In other words, $\pi(k, u, v)$ can be thought as solving the maximization problem partially, over all the tags $y_1, \dots, y_{k-2}$ with the constraint that tag u is used for y_{k-1} and tag v for y_k. If we have $\pi(k, u, v)$, then $max_v\ \pi(k, u, v)$ evaluates $max_{y_1, \dots, y_k} r(y_1, \dots, y_k)$. We leave u and v in the definition of $\pi(k, u, v)$ so that we can extend the maximization one step further as we unravel the model in the forward direction.

Base case:

$$\pi(0, u, v) = \begin{cases} 1, & if\ u\ =\ START\ and\ v =\ START \\ 0, & otherwise \end{cases}$$

Moving forward recursively:

For any $k \in \{1, \dots, n\}$, for any $u, v\ \in \daleth$,

$$\pi(k, u, v) = \max_{t\, \in\, \daleth} \{\pi(k-1, t, u) \cdot a_{t,u,v} \cdot b_v(x_k)\}$$

Transition from y_n to $STOP$:

$$\max_{y_1,\dots,y_n} p(x_1, \dots, x_n, y_0 = START, y_1, \dots, y_n, y_{n+1} = STOP)$$
$$= \max_{u,v\, \in \daleth} \{\pi(n, u, v) \cdot a_{u,v,STOP}\}$$

Backtracking

$$y^*_{n-1}, y^*_n\ =\ \arg\max_{u,v}\ \{\pi(n, u, v) \cdot a_{u,v,STOP}\}$$

For $\mathrm{k}\ =\ \mathrm{n} - 2\ to\ k\ =\ 1$:

$$y^*_k = \arg\max_t \{\pi(k+1, t, y^*_{k+1}) \cdot a_{t,y^*_{k+1},y^*_{k+2}} \cdot b_{y^*_{k+2}}(x_{k+2})\}$$
$$= \arg\max_t \{\pi(k+1, t, y^*_{k+1}) \cdot a_{t,y^*_{k+1},y^*_{k+2}}\}$$

Q4

Assume a predefined set of possible states $\{0, 1, \ldots, N-1, N\}$, where $0 = START$ and $N = STOP$.

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

$\alpha_u(i) = p(x_1, \ldots, x_{i-1}, y_1, \ldots, y_{i-1}, z_i = u; \theta)$ is the forward probability
$\beta_u(i) = p(x_i, \ldots, x_n, y_i, \ldots, y_n \,| z_i = u; \theta)$ is the backward probability

Forward probability algorithm

Base case:

$$\alpha_u(1) = a_{START,u} \qquad \forall u \in \{1, \ldots, N-1\}$$

Recursive case:

$For\ i = 1, \ldots, n-1:$

$$
\begin{aligned}
\alpha_u(i+1) &= p(x_1, \ldots, x_i, y_1, \ldots, y_i, z_{i+1} = u; \theta) \\
&= \sum_v p(x_1, \ldots x_i, y_1, \ldots, y_i, z_i = v, z_{i+1} = u) \\
&= \sum_v p(x_1, \ldots x_{i-1}, y_1, \ldots, y_{i-1}, z_i = v, x_i, y_i, z_{i+1} = u) \\
&= \sum_v p(x_1, \ldots x_{i-1}, y_1, \ldots, y_{i-1}, z_i = v) \\
&\qquad \cdot p(x_i | z_i = v) \cdot p(y_i | z_i = v, x_i) \cdot a_{v,u} \\
&= \sum_v \alpha_v(i) \cdot a_{v,u} \cdot b_v(x_i) \cdot c_{\mathrm{v},x_i}(y_i) \qquad \forall u \in \{1, \ldots, N-1\} \\
&\qquad where\ b_v(x_i) = p(x_i | z_i = v), \\
&\qquad c_{\mathrm{v},x_i}(y_i) = p(y_i | z_i = v, x_i)
\end{aligned}
$$

Backward probability algorithm

Base case:

$$\beta_u(n) = a_{u,STOP} \cdot \mathrm{b_u}(x_n) \cdot c_{u,x_n}(y_n) \qquad \forall u \in \{1, \ldots, N-1\}$$

Recursive case:

$For\ i = n-1, \ldots, 1:$

$$
\begin{aligned}
\beta_u(i) &= p(x_i, \ldots, x_n, y_i, \ldots, y_n | z_i = u; \theta) \\
&= p(x_i, \ldots x_n, y_i, \ldots, y_n, z_{i+1} = v | z_i = u) \\
&= p(x_i, y_i, z_{i+1} = v, \ldots, x_{i+1}, \ldots, x_n, y_{i+1}, \ldots, y_n | z_i = u) \\
&= \sum_v a_{u,v} \cdot b_u(x_i) \cdot \ c_{u,x_\mathrm{i}}(y_i) \cdot \beta_v(i+1) \qquad \forall u \in \{1, \ldots, N-1\}
\end{aligned}
$$

Time complexity

- Length of observation pairs = n
- Set of possible states at each position = T
- At each position, there are T forward and T backward probabilities to compute
  - Time complexity = $O(2T)$
- For each forward and backward probabilities, there are T number of operations
  - Time complexity = $O(T)$

Hence, given the length of observation pairs = n, the overall time complexity is $O(2nT^2)$, which is then simplified to $O(nT^2)$.