Homework 5

CSCI 567: Machine Learning

Seyed Mohammad Asghari Pari

USCID: 5788788474

Hidden Markov Models

Solution to Question 1.1:

We want to calculate $P(\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta})$ where $\boldsymbol{O}_{1:6} = [AGCGTA]$. To this end, we use the formula

$$P(\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}) = \sum_{j=1}^{2} \alpha_6(j), \tag{1}$$

where

$$\alpha_1(j) = \pi_j P(o_1|Z_1 = s_j), \tag{2}$$

$$\alpha_t(j) = P(o_t|Z_t = s_j) \sum_{i=1}^{2} a_{ij}\alpha_{t-1}(i). \tag{3}$$

Note that since we have only two states s_1, s_2, we can simplify (2) and (3) for $j = 1, 2$ as follows:

$$\alpha_1(1) = \pi_1 P(o_1|Z_1 = s_1) = 0.7 \times P(o_1|Z_1 = s_1), \tag{4}$$

$$\alpha_1(2) = \pi_2 P(o_1|Z_1 = s_2) = 0.3 \times P(o_1|Z_1 = s_2), \tag{5}$$

$$\alpha_t(1) = P(o_t|Z_t = s_1) \sum_{i=1}^{2} a_{i1}\alpha_{t-1}(i) = P(o_t|Z_t = s_1)[0.8\alpha_{t-1}(1) + 0.4\alpha_{t-1}(2)], \tag{6}$$

$$\alpha_t(2) = P(o_t|Z_t = s_2) \sum_{i=1}^{2} a_{i2}\alpha_{t-1}(i) = P(o_t|Z_t = s_2)[0.2\alpha_{t-1}(1) + 0.6\alpha_{t-1}(2)]. \tag{7}$$

Now, we use (4)-(7) to calculate $\alpha_t(j)$, $t = 1, \ldots, 6$, $j = 1, 2$.

- $t = 1$: We have $o_1 = A$. Hence, from (4) and (5) we have

$$\alpha_1(1) = 0.7P(X_1 = A|Z_1 = s_1) = 0.7 \times 0.4 = 0.28, \tag{8}$$

$$\alpha_1(2) = 0.3P(X_1 = A|Z_1 = s_2) = 0.3 \times 0.2 = 0.06. \tag{9}$$

- $t = 2$: We have $o_2 = G$. Hence, from (6) and (7) we have

$$\begin{aligned}\alpha_2(1) &= P(X_2 = G|Z_2 = s_1)[0.8\alpha_1(1) + 0.4\alpha_1(2)] \\ &= 0.4[0.8 \times 0.28 + 0.4 \times 0.06] = 0.0992,\end{aligned} \tag{10}$$

$$\begin{aligned}\alpha_2(2) &= P(X_2 = G|Z_2 = s_2)[0.2\alpha_1(1) + 0.6\alpha_1(2)] \\ &= 0.2[0.2 \times 0.28 + 0.6 \times 0.06] = 0.0184.\end{aligned} \tag{11}$$

- $t = 3$: We have $o_3 = C$. Hence, from (6) and (7) we have

$$\begin{aligned}\alpha_3(1) &= P(X_3 = C|Z_3 = s_1)[0.8\alpha_2(1) + 0.4\alpha_2(2)] \\ &= 0.1[0.8 \times 0.0992 + 0.4 \times 0.0184] = 0.0087, \quad (12)\\ \alpha_3(2) &= P(X_3 = C|Z_3 = s_2)[0.2\alpha_2(1) + 0.6\alpha_2(2)] \\ &= 0.3[0.2 \times 0.0992 + 0.6 \times 0.0184] = 0.0093. \quad (13)\end{aligned}$$

- $t = 4$: We have $o_4 = G$. Hence, from (6) and (7) we have

$$\begin{aligned}\alpha_4(1) &= P(X_4 = G|Z_4 = s_1)[0.8\alpha_3(1) + 0.4\alpha_3(2)] \\ &= 0.4[0.8 \times 0.0087 + 0.4 \times 0.0093] = 0.0043, \quad (14)\\ \alpha_4(2) &= P(X_4 = G|Z_4 = s_2)[0.2\alpha_3(1) + 0.6\alpha_3(2)] \\ &= 0.2[0.2 \times 0.0087 + 0.6 \times 0.0093] = 0.0015. \quad (15)\end{aligned}$$

- $t = 5$: We have $o_5 = T$. Hence, from (6) and (7) we have

$$\begin{aligned}\alpha_5(1) &= P(X_5 = T|Z_5 = s_1)[0.8\alpha_4(1) + 0.4\alpha_4(2)] \\ &= 0.1[0.8 \times 0.0043 + 0.4 \times 0.0015] = 3.98 \times 10^{-4}, \quad (16)\\ \alpha_5(2) &= P(X_5 = T|Z_5 = s_2)[0.2\alpha_4(1) + 0.6\alpha_4(2)] \\ &= 0.3[0.2 \times 0.0043 + 0.6 \times 0.0015] = 5.17 \times 10^{-4}. \quad (17)\end{aligned}$$

- $t = 6$: We have $o_6 = A$. Hence, from (6) and (7) we have

$$\begin{aligned}\alpha_6(1) &= P(X_6 = A|Z_6 = s_1)[0.8\alpha_5(1) + 0.4\alpha_5(2)] \\ &= 0.4[0.8 \times 3.98 \times 10^{-4} + 0.4 \times 5.17 \times 10^{-4}] = 2.1 \times 10^{-4}, \quad (18)\\ \alpha_6(2) &= P(X_6 = A|Z_6 = s_2)[0.2\alpha_5(1) + 0.6\alpha_5(2)] \\ &= 0.2[0.2 \times 3.98 \times 10^{-4} + 0.6 \times 5.17 \times 10^{-4}] = 7.79 \times 10^{-5}. \quad (19)\end{aligned}$$

Now, we use (1) to calculate $P(\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6})$ as follows

$$P(\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}) = \alpha_6(1) + \alpha_6(2) = 2.88 \times 10^{-4}. \quad (20)$$

Solution to Question 1.2:

First note that

$$\begin{aligned}\operatorname*{argmax}_{\boldsymbol{z}_{1:6}} P(\boldsymbol{Z}_{1:6} = \boldsymbol{z}_{1:6}|\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}) &= \operatorname*{argmax}_{\boldsymbol{z}_{1:6}} \frac{P(\boldsymbol{Z}_{1:6} = \boldsymbol{z}_{1:6}, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta})}{P(\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta})} \\ &= \operatorname*{argmax}_{\boldsymbol{z}_{1:6}} P(\boldsymbol{Z}_{1:6} = \boldsymbol{z}_{1:6}, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}), \quad (21)\end{aligned}$$

where the last equality is correct because the denominator does not depend on $\boldsymbol{z}_{1:6}$. Therefore, we need to find most likely path and to this end, we use the following formula,

$$\delta_t(j) = \max_{i=1,2} \delta_{t-1}(i) a_{ij} P(X_t = o_t|Z_t = s_j). \quad (22)$$

Note that since we have only two states s_1, s_2, we can simplify (2) and (3) for $j = 1, 2$ as follows:

$$\begin{aligned}\delta_1(1) &= \pi_1 P(o_1|Z_1 = s_1) = 0.7 \times P(o_1|Z_1 = s_1), \quad (23)\\ \delta_1(2) &= \pi_2 P(o_1|Z_1 = s_2) = 0.3 \times P(o_1|Z_1 = s_2), \quad (24)\\ \delta_t(1) &= \max\{\delta_{t-1}(1)a_{11}P(X_t = o_t|Z_t = s_1), \delta_{t-1}(2)a_{21}P(X_t = o_t|Z_t = s_1)\} \\ &= \max\{\underbrace{0.8\delta_{t-1}(1)P(X_t = o_t|Z_t = s_1)}_{\mathbb{T}_t(1,1)}, \underbrace{0.4\delta_{t-1}(2)P(X_t = o_t|Z_t = s_1)}_{\mathbb{T}_t(2,1)}\}, \quad (25)\\ \delta_t(2) &= \max\{\delta_{t-1}(1)a_{12}P(X_t = o_t|Z_t = s_2), \delta_{t-1}(2)a_{22}P(X_t = o_t|Z_t = s_2)\} \\ &= \max\{\underbrace{0.2\delta_{t-1}(1)P(X_t = o_t|Z_t = s_2)}_{\mathbb{T}_t(1,2)}, \underbrace{0.6\delta_{t-1}(2)P(X_t = o_t|Z_t = s_2)}_{\mathbb{T}_t(2,2)}\}. \quad (26)\end{aligned}$$

Now, we use (25)-(26) to calculate $\delta_t(j), t = 1, \ldots, 6, j = 1, 2$.

- $t = 1$: We have $o_1 = A$. Hence, from (23)-(24) we have

$$\delta_1(1) = 0.7P(X_1 = A|Z_1 = s_1) = 0.7 \times 0.4 = 0.28, \tag{27}$$
$$\delta_1(2) = 0.3P(X_1 = A|Z_1 = s_2) = 0.3 \times 0.2 = 0.06. \tag{28}$$

- $t = 2$: We have $o_2 = G$. Hence, from (25)-(26) we have

$$\begin{aligned}
\mathbb{T}_2(1,1) &= 0.8\delta_1(1)P(X_2 = G|Z_2 = s_1) = 0.8 \times 0.28 \times 0.4 = 0.0896\\
\mathbb{T}_2(2,1) &= 0.4\delta_1(2)P(X_2 = G|Z_2 = s_1) = 0.4 \times 0.06 \times 0.2 = 0.0048\\
\delta_2(1) &= \max\{\mathbb{T}_2(1,1), \mathbb{T}_2(2,1)\} = 0.0896 &\quad (29)\\
\mathbb{T}_2(1,2) &= 0.2\delta_1(1)P(X_2 = G|Z_2 = s_2) = 0.2 \times 0.28 \times 0.4 = 0.0224\\
\mathbb{T}_2(2,2) &= 0.6\delta_1(2)P(X_2 = G|Z_2 = s_2) = 0.6 \times 0.06 \times 0.2 = 0.0072\\
\delta_2(2) &= \max\{\mathbb{T}_2(1,2), \mathbb{T}_2(2,2)\} = 0.0224. &\quad (30)
\end{aligned}$$

Since $\mathbb{T}_2(1,1) > \mathbb{T}_2(2,1)$, if the most probable state at $t = 2$ is s_1, then the most probable state at time $t = 1$ is also s_1. Further, since $\mathbb{T}_2(1,2) > \mathbb{T}_2(2,2)$, if the most probable state at $t = 2$ is s_2, then the most probable state at time $t = 1$ is s_1.

- $t = 3$: We have $o_3 = C$. Hence, from (25)-(26) we have

$$\begin{aligned}
\mathbb{T}_3(1,1) &= 0.8\delta_2(1)P(X_3 = C|Z_3 = s_1) = 0.8 \times 0.0896 \times 0.1 = 0.0072\\
\mathbb{T}_3(2,1) &= 0.4\delta_2(2)P(X_3 = C|Z_3 = s_1) = 0.4 \times 0.0224 \times 0.3 = 0.0027\\
\delta_3(1) &= \max\{\mathbb{T}_3(1,1), \mathbb{T}_3(2,1)\} = 0.0072 &\quad (31)\\
\mathbb{T}_3(1,2) &= 0.2\delta_2(1)P(X_3 = C|Z_3 = s_2) = 0.2 \times 0.0896 \times 0.1 = 0.0012\\
\mathbb{T}_3(2,2) &= 0.6\delta_2(2)P(X_3 = C|Z_3 = s_2) = 0.6 \times 0.0224 \times 0.3 = 0.0040\\
\delta_3(2) &= \max\{\mathbb{T}_3(1,2), \mathbb{T}_3(2,2)\} = 0.0040. &\quad (32)
\end{aligned}$$

Since $\mathbb{T}_3(1,1) > \mathbb{T}_3(2,1)$, if the most probable state at $t = 3$ is s_1, then the most probable state at time $t = 2$ is also s_1. Further, since $\mathbb{T}_3(1,2) < \mathbb{T}_3(2,2)$, if the most probable state at $t = 3$ is s_2, then the most probable state at time $t = 2$ is also s_2.

- $t = 4$: We have $o_4 = G$. Hence, from (25)-(26) we have

$$\begin{aligned}
\mathbb{T}_4(1,1) &= 0.8\delta_3(1)P(X_4 = G|Z_4 = s_1) = 0.8 \times 0.0072 \times 0.4 = 0.0023\\
\mathbb{T}_4(2,1) &= 0.4\delta_3(2)P(X_4 = G|Z_4 = s_1) = 0.4 \times 0.0040 \times 0.2 = 3.22 \times 10^{-4}\\
\delta_4(1) &= \max\{\mathbb{T}_4(1,1), \mathbb{T}_4(2,1)\} = 0.0023 &\quad (33)\\
\mathbb{T}_4(1,2) &= 0.2\delta_3(1)P(X_4 = G|Z_4 = s_2) = 0.2 \times 0.0072 \times 0.4 = 5.73 \times 10^{-4}\\
\mathbb{T}_4(2,2) &= 0.6\delta_3(2)P(X_4 = G|Z_4 = s_2) = 0.6 \times 0.0040 \times 0.2 = 4.83 \times 10^{-4}\\
\delta_4(2) &= \max\{\mathbb{T}_4(1,2), \mathbb{T}_4(2,2)\} = 5.73 \times 10^{-4}. &\quad (34)
\end{aligned}$$

Since $\mathbb{T}_4(1,1) > \mathbb{T}_4(2,1)$, if the most probable state at $t = 4$ is s_1, then the most probable state at time $t = 3$ is also s_1. Further, since $\mathbb{T}_4(1,2) > \mathbb{T}_4(2,2)$, if the most probable state at $t = 4$ is s_2, then the most probable state at time $t = 3$ is s_1.

- $t = 5$: We have $o_5 = T$. Hence, from (25)-(26) we have

$$\begin{aligned}
\mathbb{T}_5(1,1) &= 0.8\delta_4(1)P(X_5 = T|Z_5 = s_1) = 0.8 \times 0.0023 \times 0.1 = 1.83 \times 10^{-4}\\
\mathbb{T}_5(2,1) &= 0.4\delta_4(2)P(X_5 = T|Z_5 = s_1) = 0.4 \times 5.73 \times 10^{-4} \times 0.3 = 6.88 \times 10^{-5}\\
\delta_5(1) &= \max\{\mathbb{T}_5(1,1), \mathbb{T}_5(2,1)\} = 1.83 \times 10^{-4} &\quad (35)\\
\mathbb{T}_5(1,2) &= 0.2\delta_4(1)P(X_5 = T|Z_5 = s_2) = 0.2 \times 0.0023 \times 0.1 = 4.58 \times 10^{-5}\\
\mathbb{T}_5(2,2) &= 0.6\delta_4(2)P(X_5 = T|Z_5 = s_2) = 0.6 \times 5.73 \times 10^{-4} \times 0.3 = 1.03 \times 10^{-4}\\
\delta_5(2) &= \max\{\mathbb{T}_5(1,2), \mathbb{T}_5(2,2)\} = 1.03 \times 10^{-4}. &\quad (36)
\end{aligned}$$

Since $\mathbb{T}_5(1,1) > \mathbb{T}_5(2,1)$, if the most probable state at $t = 5$ is s_1, then the most probable state at time $t = 4$ is also s_1. Further, since $\mathbb{T}_5(1,2) < \mathbb{T}_5(2,2)$, if the most probable state at $t = 5$ is s_2, then the most probable state at time $t = 4$ is also s_2.

- $t = 6$: We have $o_6 = A$. Hence, from (25)-(26) we have

$$\mathbb{T}_6(1,1) = 0.8\delta_5(1)P(X_6 = A|Z_6 = s_1) = 0.8 \times 1.83 \times 10^{-4} \times 0.4 = 5.87 \times 10^{-5}$$
$$\mathbb{T}_6(2,1) = 0.4\delta_5(2)P(X_6 = A|Z_6 = s_1) = 0.4 \times 1.03 \times 10^{-4} \times 0.2 = 8.25 \times 10^{-6}$$
$$\delta_6(1) = \max\{\mathbb{T}_6(1,1), \mathbb{T}_6(2,1)\} = 5.87 \times 10^{-5} \tag{37}$$
$$\mathbb{T}_6(1,2) = 0.2\delta_5(1)P(X_6 = A|Z_6 = s_2) = 0.2 \times 1.83 \times 10^{-4} \times 0.4 = 1.46 \times 10^{-5}$$
$$\mathbb{T}_6(2,2) = 0.6\delta_5(2)P(X_6 = A|Z_6 = s_2) = 0.6 \times 1.03 \times 10^{-4} \times 0.2 = 1.23 \times 10^{-5}$$
$$\delta_6(2) = \max\{\mathbb{T}_6(1,2), \mathbb{T}_6(2,2)\} = 1.46 \times 10^{-5}. \tag{38}$$

  Since $\mathbb{T}_6(1,1) > \mathbb{T}_6(2,1)$, if the most probable state at $t = 6$ is s_1, then the most probable state at time $t = 5$ is also s_1. Further, since $\mathbb{T}_6(1,2) > \mathbb{T}_6(2,2)$, if the most probable state at $t = 6$ is s_2, then the most probable state at time $t = 5$ is s_1.

Now, since $\delta_6(1) > \delta_6(2)$, the most probable state at $t = 6$ is s_1. If we trace this path back, we can see that the most probable state at time $t = 5$ is also s_1. Continuing this procedure, we find the most probable path as $\boldsymbol{z}^*_{1:6} = [s_1, s_1, s_1, s_1, s_1, s_1]$.

Solution to Question 1.3:

We want to find x maximizing $x^* = \operatorname{argmax}_x P(X_7 = x|\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta})$ where $\boldsymbol{O}_{1:6} = [AGCGTA]$. To this end, note that

$$\begin{aligned}\underset{x}{\operatorname{argmax}}\, P(X_7 = x|\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}) &= \underset{x}{\operatorname{argmax}}\, \frac{P(X_7 = x, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta})}{P(\boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta})} \\ &= \underset{x}{\operatorname{argmax}}\, P(X_7 = x, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}),\end{aligned} \tag{39}$$

where the last equality is correct because the denominator does not depend on x.

Since x can have one of four values A, C, G, T, we calculate $P(X_7 = x, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta})$ for all these values and choose the one maximizing that term. Note that we have already calculated $\alpha_6(1), \alpha_6(2)$ and we can use them to find $\alpha_7(1), \alpha_7(2)$.

- $x = A$. In this case, we have,

$$\begin{aligned}\alpha_7(1) &= P(X_7 = A|Z_7 = s_1)[0.8\alpha_6(1) + 0.4\alpha_6(2)] \\ &= 0.4[0.8 \times 2.1 \times 10^{-4} + 0.4 \times 7.79 \times 10^{-5}] = 7.96 \times 10^{-5},\end{aligned} \tag{40}$$
$$\begin{aligned}\alpha_7(2) &= P(X_7 = A|Z_7 = s_2)[0.2\alpha_6(1) + 0.6\alpha_6(2)] \\ &= 0.2[0.2 \times 2.1 \times 10^{-4} + 0.6 \times 7.79 \times 10^{-5}] = 1.77 \times 10^{-5}.\end{aligned} \tag{41}$$

  Hence, we have

$$P(X_7 = A, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}) = \alpha_7(1) + \alpha_7(2) = 9.73 \times 10^{-5}. \tag{42}$$

- $x = C$. In this case, we have,

$$\begin{aligned}\alpha_7(1) &= P(X_7 = C|Z_7 = s_1)[0.8\alpha_6(1) + 0.4\alpha_6(2)] \\ &= 0.1[0.8 \times 2.1 \times 10^{-4} + 0.4 \times 7.79 \times 10^{-5}] = 1.99 \times 10^{-5},\end{aligned} \tag{43}$$
$$\begin{aligned}\alpha_7(2) &= P(X_7 = C|Z_7 = s_2)[0.2\alpha_6(1) + 0.6\alpha_6(2)] \\ &= 0.3[0.2 \times 2.1 \times 10^{-4} + 0.6 \times 7.79 \times 10^{-5}] = 2.66 \times 10^{-5}.\end{aligned} \tag{44}$$

  Hence, we have

$$P(X_7 = C, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}) = \alpha_7(1) + \alpha_7(2) = 4.66 \times 10^{-5}. \tag{45}$$

- $x = G$. In this case, we have,

$$\begin{aligned}\alpha_7(1) &= P(X_7 = G|Z_7 = s_1)[0.8\alpha_6(1) + 0.4\alpha_6(2)] \\ &= 0.4[0.8 \times 2.1 \times 10^{-4} + 0.4 \times 7.79 \times 10^{-5}] = 7.96 \times 10^{-5},\end{aligned} \tag{46}$$
$$\begin{aligned}\alpha_7(2) &= P(X_7 = G|Z_7 = s_2)[0.2\alpha_6(1) + 0.6\alpha_6(2)] \\ &= 0.2[0.2 \times 2.1 \times 10^{-4} + 0.6 \times 7.79 \times 10^{-5}] = 1.77 \times 10^{-5}.\end{aligned} \tag{47}$$

  Hence, we have

$$P(X_7 = G, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}) = \alpha_7(1) + \alpha_7(2) = 9.73 \times 10^{-5}. \tag{48}$$

- $x = T$. In this case, we have,

$$\begin{aligned}\alpha_7(1) &= P(X_7 = T|Z_7 = s_1)[0.8\alpha_6(1) + 0.4\alpha_6(2)] \\ &= 0.1[0.8 \times 2.1 \times 10^{-4} + 0.4 \times 7.79 \times 10^{-5}] = 1.99 \times 10^{-5}, \qquad (49) \\ \alpha_7(2) &= P(X_7 = T|Z_7 = s_2)[0.2\alpha_6(1) + 0.6\alpha_6(2)] \\ &= 0.3[0.2 \times 2.1 \times 10^{-4} + 0.6 \times 7.79 \times 10^{-5}] = 2.66 \times 10^{-5}. \qquad (50)\end{aligned}$$

  Hence, we have

$$P(X_7 = T, \boldsymbol{X}_{1:6} = \boldsymbol{O}_{1:6}, \boldsymbol{\Theta}) = \alpha_7(1) + \alpha_7(2) = 4.66 \times 10^{-5}. \qquad (51)$$

As can be seen, the maximizers are $x^* = A, G$.