CSCI 567 HW 5

Amir Erfan Eshratifar
USC ID: 2875431498

1 Hidden Markov Models

1. The target sequence we are looking for is "AGCGTA". For the initial state we have:
   At t=1:

$$\alpha_1(j) = P(X_1 = x_1|Z_1 = s_j)\pi_j$$
$$\alpha_1(1) = \pi_1 P(X_1 = A|Z_1 = 1) = 0.7 * 0.4 = 0.28$$
$$\alpha_1(2) = \pi_2 P(X_1 = A|Z_1 = 2) = 0.3 * 0.2 = 0.06$$

For the next states we have:

$$\alpha_t(j) = P(X_t = x_t|Z_t = s_j)\sum_i a_{ij}\alpha_{t-1}(j)$$

At t=2:

$$\begin{aligned}
\alpha_2(1) &= P(X_2 = G|Z_2 = 1)\sum_i a_{i1}\alpha_1(1) \\
&= b_{1G}(a_1 1 * \alpha_1(1) + a_2 1 * \alpha_1(2)) = 0.4 * (0.8 * 0.28 + 0.4 * 0.06) = 0.0992 \\
\alpha_2(2) &= P(X_2 = G|Z_2 = 2)\sum_i a_{i2}\alpha_1(2) \\
&= b_{2G}(a_2 1 * \alpha_1(1) + a_2 2 * \alpha_1(2)) = 0.2 * (0.2 * 0.28 + 0.6 * 0.06) = 0.0184
\end{aligned}$$

At t=3:

$$\begin{aligned}
\alpha_3(1) &= P(X_3 = C|Z_3 = 1)\sum_i a_{i1}\alpha_2(i) \\
&= b_{1G}(a_{11} * \alpha_2(1) + a_2 1 * \alpha_2(2)) = 0.1 * (0.8 * 0.0992 + 0.4 * 0.0184) = 0.008672 \\
\alpha_3(2) &= P(X_3 = C|Z_3 = 2)\sum_i a_{i2}\alpha_2(i) \\
&= b_{2G}(a_{21} * \alpha_2(1) + a_2 2 * \alpha_2(2)) = 0.3 * (0.2 * 0.0992 + 0.6 * 0.0184) = 0.009264
\end{aligned}$$

At t=4:

$$\begin{aligned}
\alpha_4(1) &= P(X_4 = G|Z_4 = 1)\sum_i a_{i1}\alpha_3(i) \\
&= b_{1G}(a_{11} * \alpha_3(1) + a_{21} * \alpha_3(2)) = 0.1 * (0.8 * 0.008672 + 0.4 * 0.009264) = 0.00425728 \\
\alpha_4(2) &= P(X_4 = G|Z_4 = 2)\sum_i a_{i2}\alpha_3(i) \\
&= b_{2G}(a_{21} * \alpha_3(1) + a_{22} * \alpha_3(2)) = 0.3 * (0.2 * 0.008672 + 0.6 * 0.009264) = 0.00145856
\end{aligned}$$

At t=5:

$$\begin{aligned}
\alpha_5(1) &= P(X_5 = T|Z_5 = 1)\sum_i a_{i1}\alpha_4(i) \\
&= b_{1T}(a_{11} * \alpha_4(1) + a_{21} * \alpha_4(2)) = 0.1 * (0.8 * 0.00425728 + 0.4 * 0.00145856) = 0.000398925 \\
\alpha_5(2) &= P(X_5 = T|Z_5 = 2)\sum_i a_{i2}\alpha_4(i) \\
&= b_{2T}(a_{21} * \alpha_4(1) + a_{22} * \alpha_4(2)) = 0.3 * (0.2 * 0.00425728 + 0.6 * 0.00145856) = 0.000517978
\end{aligned}$$

At t=6:

$$\begin{aligned}
\alpha_6(1) &= P(X_6 = A|Z_6 = 1)\sum_i a_{i1}\alpha_5(i) \\
&= b_{1A}(a_{11} * \alpha_5(1) + a_{21} * \alpha_4(2)) = 0.1 * (0.8 * 0.000398925 + 0.4 * 0.000517978) = 0.000210532 \\
\alpha_5(2) &= P(X_6 = A|Z_6 = 2)\sum_i a_{i2}\alpha_5(i) \\
&= b_{2A}(a_{21} * \alpha_5(1) + a_{22} * \alpha_5(2)) = 0.3 * (0.2 * 0.000398925 + 0.6 * 0.000517978) = 7.81143E - 05
\end{aligned}$$

2. The target sequence we are looking for is "AGCGTA". For the initial state we have:
At t=1:

$$\begin{aligned}
\sigma_1(s = 1) &= P(X_1 = A|Z_1 = 1)\pi(1) = 0.4 * 0.7 = 0.28 \\
\sigma_1(s = 2) &= P(X_1 = A|Z_1 = 2)\pi(2) = 0.2 * 0.3 = 0.06
\end{aligned}$$

At t=2:

$$\begin{aligned}
&\sigma_2(s = 1) = 0.0896 \\
&\max\Big(P(X_2 = G|Z_2 = 1)P(Z_2 = 1|Z_1 = 1)\sigma_1(1), P(X_2 = G|Z_2 = 1)P(Z_2 = 1|Z_1 = 2)\sigma_1(2)\Big) \\
&= \max(0.4 * 0.8 * 0.28, 0.4 * 0.4 * 0.06) = 0.0896 \\
&\sigma_2(s = 2) = 0.0112 \\
&\max\Big(P(X_2 = G|Z_2 = 2)P(Z_2 = 2|Z_1 = 1)\sigma_1(1), P(X_2 = G|Z_2 = 2)P(Z_2 = 2|Z_1 = 2)\sigma_1(2)\Big) \\
&= \max(0.2 * 0.2 * 0.28, 0.2 * 0.6 * 0.06) = 0.0112
\end{aligned}$$

At t=3:

$$\begin{aligned}
&\sigma_3(s = 1) = 0.007168 \\
&\max\Big(P(X_3 = C|Z_3 = 1)P(Z_3 = 1|Z_2 = 1)\sigma_2(1), P(X_3 = C|Z_3 = 1)P(Z_3 = 1|Z_2 = 2)\sigma_2(2)\Big) \\
&= \max(0.1 * 0.8 * 0.0896, 0.1 * 0.4 * 0.0112) = 0.007168 \\
&\sigma_3(s = 2) = 0.005376 \\
&\max\Big(P(X_3 = C|Z_3 = 2)P(Z_3 = 2|Z_2 = 1)\sigma_2(1), P(X_3 = C|Z_3 = 2)P(Z_3 = 2|Z_2 = 2)\sigma_2(2)\Big) \\
&= \max(0.3 * 0.2 * 0.0896, 0.3 * 0.6 * 0.0112) = 0.005376
\end{aligned}$$

At t=4:

$$\begin{aligned}
&\sigma_4(s=1) = 0.00229376 \\
&\max\Big(P(X_4=G|Z_4=1)P(Z_4=1|Z_3=1)\sigma_3(1), P(X_4=G|Z_4=1)P(Z_4=1|Z_3=2)\sigma_3(2)\Big) \\
&= \max\left(0.3*0.2*0.007168, 0.3*0.6*0.005376\right) = 0.00229376 \\
&\sigma_4(s=2) = 0.00064512 \\
&\max\Big(P(X_4=G|Z_4=2)P(Z_4=2|Z_3=1)\sigma_3(1), P(X_4=G|Z_4=2)P(Z_4=2|Z_3=2)\sigma_3(2)\Big) \\
&= \max\left(0.2*0.2*0.007168, 0.2*0.6*0.005376\right) = 0.00064512
\end{aligned}$$

At t=5:

$$\begin{aligned}
&\sigma_5(s=1) = 0.0001835 \\
&\max\Big(P(X_5=T|Z_5=1)P(Z_5=1|Z_4=1)\sigma_4(1), P(X_5=T|Z_5=1)P(Z_5=1|Z_4=2)\sigma_4(2)\Big) \\
&= \max\left(0.1*0.8*\sigma_4(s=1), 0.1*0.4*\sigma_4(s=2)\right) = 0.0001835 \\
&\sigma_5(s=2) = 0.0001376 \\
&\max\Big(P(X_5=T|Z_5=2)P(Z_5=2|Z_4=1)\sigma_4(1), P(X_5=T|Z_5=2)P(Z_5=2|Z_4=2)\sigma_4(2)\Big) \\
&= \max\left(0.3*0.2*\sigma_4(s=1), 0.3*0.6*\sigma_4(s=2)\right) = 0.0001835
\end{aligned}$$

At t=6:

$$\begin{aligned}
&\sigma_6(s=1) = 0.00005872 \\
&\max\Big(P(X_6=A|Z_6=1)P(Z_6=1|Z_5=1)\sigma_5(1), P(X_6=A|Z_6=1)P(Z_6=1|Z_5=2)\sigma_5(2)\Big) \\
&= \max\left(0.8*0.4*\sigma_5(s=1), 0.4*0.4*\sigma_5(s=2)\right) = 0.00005872 \\
&\sigma_6(s=2) = 0.000016512 \\
&\max\Big(P(X_6=A|Z_6=2)P(Z_6=2|Z_5=1)\sigma_5(1), P(X_6=A|Z_6=2)P(Z_6=2|Z_5=2)\sigma_5(2)\Big) \\
&= \max\left(0.2*0.2*\sigma_5(s=1), 0.2*0.6*\sigma_5(s=2)\right) = 0.000016512
\end{aligned}$$

By tracing the sequence back we can the state path:

$$z^* = [z_1^*, z_2^*, z_3^*, z_4^*, z_5^*, z_6^*] = [1,1,1,1,1,1]$$

3.

$$\begin{aligned}
x^* &= argmax_x P(X_7 = x|X_{1:6} = O_{1:6}) \\
&= argmax_x \frac{P(X_7 = x, X_{1:6} = O_{1:6})}{P(X_{1:6} = O_{1:6})} \\
&= argmax_x \frac{P(X_7 = x, X_{1:6} = O_{1:6})}{const} \\
&= argmax_x P(X_7 = x, X_{1:6} = O_{1:6}) \\
&= \sum_i P(X_{1:7}, Z_7 = s_i) = \sum_i \alpha_7(i)
\end{aligned}$$

For x = 'A' we have:

$$\alpha_7(1) = P(X_7 = A|Z_7 = 1)\sum_i a_{i1}\alpha_6(i) = 7.98686E-05$$
$$\alpha_7(2) = P(X_7 = A|Z_7 = 2)\sum_i a_{i2}\alpha_6(i) = 1.7795E-05$$

For x = 'C' we have:

$$\alpha_7(1) = P(X_7 = C|Z_7 = 1)\sum_i a_{i1}\alpha_6(i) = 1.99672E-05$$
$$\alpha_7(2) = P(X_7 = C|Z_7 = 2)\sum_i a_{i2}\alpha_6(i) = 2.66925E-05$$

For x = 'G' we have:

$$\alpha_7(1) = P(X_7 = G|Z_7 = 1)\sum_i a_{i1}\alpha_6(i) = 7.98686E-05$$
$$\alpha_7(2) = P(X_7 = G|Z_7 = 2)\sum_i a_{i2}\alpha_6(i) = 1.7795E-05$$

For x = 'T' we have:

$$\alpha_7(1) = P(X_7 = T|Z_7 = 1)\sum_i a_{i1}\alpha_6(i) = 1.99672E-05$$
$$\alpha_7(2) = P(X_7 = T|Z_7 = 2)\sum_i a_{i2}\alpha_6(i) = 2.66925E-05$$

For the following joint probabilities we have:

$$P(X_7 = A, X_{1:6} = O_{1:6}) = \sum_{i=1}^{2}\alpha_7(i) = 9.76637E-05$$
$$P(X_7 = C, X_{1:6} = O_{1:6}) = \sum_{i=1}^{2}\alpha_7(i) = 4.66597E-05$$
$$P(X_7 = G, X_{1:6} = O_{1:6}) = \sum_{i=1}^{2}\alpha_7(i) = 9.76637E-05$$
$$P(X_7 = T, X_{1:6} = O_{1:6}) = \sum_{i=1}^{2}\alpha_7(i) = 4.66597E-05$$

Therefore, x^* can be either 'A' or 'G'.