8 HMM

8. (12 points) Consider the following Markov chain (an HMM without observations – the state is observed after every transition):

The conditional probability tables are:

X_1	$p(X_1)$
T	0.25
F	0.75

For X_2, X_3, X_4:

X_i	X_{i+1}	$p(X_{i+1} \mid X_i)$
T	T	0.25
T	F	0.75
F	T	0.75
F	F	0.25

Now add observations Y_1, Y_2, Y_3, Y_4, such that

X_i	Y_i	$p(Y_i \mid X_i)$
T	T	0.8
T	F	0.2
F	T	0
F	F	1

(b) Infer the most likely state sequence given the observation sequence

$$Y_1 = T,\ Y_2 = F,\ Y_3 = F,\ Y_4 = T.$$

You are welcome to work out the numerical values of the messages in Viterbi. But as a hint, you can use the properties of the observation probabilities to take shortcuts on needing to compute all the messages.

Solution:

Color Legend: To make the formulas easier to follow, Observations (Y) are purple, States (X) are blue, and Viterbi messages (δ) are orange. We use two shades of green for True (T and T) and two shades of red for False (F and F). The **brighter** shade emphasizes the specific target value being evaluated in the δ function and its corresponding conditions inside the probability formulas, while the **darker** shade represents the alternative/prior state values.

We will approach this problem using the Viterbi algorithm. Recall that Viterbi is a dynamic programming method in which we build a table containing every possible state at each time step, and for each entry, we choose the previous state that is most likely to transition into it. After filling the table, we pick the most likely final state and then backtrack to recover the full path. The quantity $\delta_t(s)$ denotes the relative probability of being in state s at time t, conditioned on the sequence of previous states. Also note that $\delta_t(s)$ is **not** the probability of being in a given state, but rather the probability of being in that state given the most likely path that would end there given our observations.

At $t = 1$ we have $Y_1 = T$, which tells us:

$$\delta_1(T) = p(X_1 = T)p(Y_1 = T \mid X_1 = T) = 0.25 \cdot 0.8 = 0.2$$

$$\delta_1(F) = p(X_1 = F)p(Y_1 = T \mid X_1 = F) = 0.75 \cdot 0.0 = 0$$

δ Table:

t	1	2	3	4
Y_t	T	F	F	T
$s = T$	0.2			
$s = F$	0			

Substituting the values from $t = 1$, we evaluate δ_2 for $Y_2 = F$:

$$\delta_2(T) = \left(\max \begin{Bmatrix} p(X_2 = T \mid X_1 = F)\,\delta_1(F), \\ p(X_2 = T \mid X_1 = T)\,\delta_1(T) \end{Bmatrix}\right) p(Y_2 = F \mid X_2 = T)$$

$$\delta_2(T) = \max\{0.75 \cdot 0,\ 0.25 \cdot 0.2\} \cdot 0.2 = \max\{0, 0.05\} \cdot 0.2 = 0.01$$

(Maximized by $X_1 = T$)

$$\delta_2(F) = \left(\max \begin{Bmatrix} p(X_2 = F \mid X_1 = F)\,\delta_1(F), \\ p(X_2 = F \mid X_1 = T)\,\delta_1(T) \end{Bmatrix}\right) p(Y_2 = F \mid X_2 = F)$$

$$\delta_2(F) = \max\{0.25 \cdot 0,\ 0.75 \cdot 0.2\} \cdot 1 = \max\{0, 0.15\} \cdot 1 = 0.15$$

(Maximized by $X_1 = T$)

δ Table:

t	1	2	3	4
Y_t	T	F	F	T
$s = T$	0.2	0.01		
$s = F$	0	0.15		

Next, for $t = 3$ where $Y_3 = F$:

$$\delta_3(T) = \left(\max \begin{Bmatrix} p(X_3 = T \mid X_2 = F)\,\delta_2(F), \\ p(X_3 = T \mid X_2 = T)\,\delta_2(T) \end{Bmatrix}\right) p(Y_3 = F \mid X_3 = T)$$

$$\delta_3(T) = \max\{0.75 \cdot 0.15,\ 0.25 \cdot 0.01\} \cdot 0.2 = \max\{0.1125, 0.0025\} \cdot 0.2 = 0.0225$$

(Maximized by $X_2 = F$)

$$\delta_3(F) = \left(\max \begin{Bmatrix} p(X_3 = F \mid X_2 = F)\,\delta_2(F), \\ p(X_3 = F \mid X_2 = T)\,\delta_2(T) \end{Bmatrix}\right) p(Y_3 = F \mid X_3 = F)$$

$$\delta_3(F) = \max\{0.25 \cdot 0.15,\ 0.75 \cdot 0.01\} \cdot 1 = \max\{0.0375, 0.0075\} \cdot 1 = 0.0375$$

(Maximized by $X_2 = F$)

δ Table:

t	1	2	3	4
Y_t	T	F	F	T
$s = T$	0.2	0.01	0.0225	
$s = F$	0	0.15	0.0375	

Finally, for $t = 4$ where $Y_4 = T$. As the hint suggested, we can use the properties of the observation probabilities to take a shortcut. Since $p(Y_4 = T \mid X_4 = F) = 0$, we immediately know that $\delta_4(F) = 0$. We only need to compute $\delta_4(T)$:

$$\delta_4(T) = \left(\max \begin{Bmatrix} p(X_4 = T \mid X_3 = F)\,\delta_3(F), \\ p(X_4 = T \mid X_3 = T)\,\delta_3(T) \end{Bmatrix}\right) p(Y_4 = T \mid X_4 = T)$$

$$\delta_4(T) = \max\{0.75 \cdot 0.0375,\ 0.25 \cdot 0.0225\} \cdot 0.8 = \max\{0.028125, 0.005625\} \cdot 0.8 = 0.0225$$

(Maximized by $X_3 = F$)

Final δ Table:

t	1	2	3	4
Y_t	T	F	F	T
$s = T$	0.2	0.01	0.0225	0.0225
$s = F$	0	0.15	0.0375	0

To find the most likely state sequence, we backtrack from the maximum probability at $t = 4$ (highlighted in red in the final table).

1. At $t = 4$, the maximum probability is $\delta_4(T)$, so $X_4 = T$.
2. The value for $\delta_4(T)$ was maximized by transitioning from $X_3 = F$.
3. The value for $\delta_3(F)$ was maximized by transitioning from $X_2 = F$.
4. The value for $\delta_2(F)$ was maximized by transitioning from $X_1 = T$.

Therefore, the most likely state sequence is:

$$X_1 = T,\ X_2 = F,\ X_3 = F,\ X_4 = T$$