(520|600).666 Information Extraction

Homework # 2

Due Thursday, February 15, 2024.

1. Consider the HMM of Chapter 2, Figure 2.8, with state space $\mathcal{S} = \{1, 2, 3\}$ and output alphabet $\mathcal{Y} = \{\mathbf{0}, \mathbf{1}\}$. In the notation of Chapter 2, let the *transition* probabilities for the output-producing transitions and null transitions be given, respectively, by

$$p(s'|s) = \begin{bmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{3} \\ \frac{3}{4} & \frac{1}{4} & 0 \end{bmatrix} \quad \text{and} \quad q(s'|s) = \begin{bmatrix} 0 & 0 & \frac{1}{6} \\ \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix}, \quad s, s' \in \mathcal{S}.$$

   Let the probabilities associated with the output producing transition from state s to state s' be given by

$$q(\mathbf{0}|s \to s') = \begin{bmatrix} 1 & \frac{1}{2} & 1 \\ 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad q(\mathbf{1}|s \to s') = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 1 & 1 & \frac{2}{3} \\ 1 & 1 & 1 \end{bmatrix},$$

   and note that $q(\mathbf{1}|s \to s')$ is simply $1 - q(\mathbf{0}|s \to s')$. Let the initial state be $s_0 = 1$.

   Perform the following calculations *by hand*.[1]

   (a) Draw a state diagram of this HMM, attaching state-labels to all nodes, probabilities $r(y, s'|s)$ or $q(s'|s)$ to all arcs and output-labels to all non-null arcs.

   (b) Draw a 4-stage trellis for this HMM, showing *only* the paths that can result in the output $y_1y_2y_3y_4 = \mathbf{0110}$.

   (c) Compute the *forward* probabilities $\alpha_i(s)$ for the output sequence **0110**, and indicate them on the trellis drawn above for (b).

   (d) Compute the marginal probability of the output: $P(y_1y_2y_3y_4 = \mathbf{0110}|s_0 = 1)$.

   (e) Redraw the trellis with all the paths that can give rise to **0110**, and indicate on it the *Viterbi* probabilities $\gamma_i(s)$ for each time and state, and the trace-back pointer.

   (f) Identify the most likely path given $y_1y_2y_3y_4 = \mathbf{0110}$ and color it.

[1] Using a calculator is fine. Make sure you retain a sufficient number of significant digits to preserve numerical precision, or retain your answers as integral *fractions*. The latter is guaranteed not to run into precision problems. A collateral benefit of doing these calculations by hand is that this can serve as a *test input* for your Project #1 code, wherein you will perform the same computations on real-life data.

2. Using basic probability and the properties of hidden Markov models, show that we can use the forward probabilities, $\alpha_i(s)$, to efficiently compute the marginal probability of an observation sequence. You can use ChatGPT to solve this problem if you want. However, make sure to attach all prompts needed to obtain the correct answer as well as any incorrect or incomplete responses you obtained along the way.

Carefully review §2.2-§2.6 from Chapter 2 again after finishing the homework. Hopefully, you will have a clearer understanding of the mathematical notation after having worked out this example.