Theoretical justification for the forward procedure

Problem 1: Computing $P(\mathcal{O}|\lambda)$

For the total probability of the observation sequence with model λ, we sum across all possible state sequences:

$$\begin{aligned} P(\mathcal{O}|\lambda) &= \sum_{\text{all } X} P(\mathcal{O}, X|\lambda) \\ &= \sum_{i=1}^{N} P(\boldsymbol{x}_1^T, \boldsymbol{o}_1^T|\lambda) \end{aligned} \tag{33}$$

where $\boldsymbol{x}_1^T$ and $\boldsymbol{o}_1^T$ are T-frame state and observation sequences. The RHS term upto any frame t is

$$P(\boldsymbol{x}_1^t, \boldsymbol{o}_1^t|\lambda) = \sum_{i=1}^{N} P(\boldsymbol{x}_1^{t-1}, \boldsymbol{o}_1^{t-1}|\lambda)\, P(x_t, o_t|\boldsymbol{x}_1^{t-1}, \boldsymbol{o}_1^{t-1}, \lambda), \tag{34}$$

and we define the *forward likelihood* as

$$\alpha_t(i) = P(\boldsymbol{x}_1^t, \boldsymbol{o}_1^t|\lambda). \tag{35}$$

This enables us to re-write eq. 34, for $x_t = j$:

$$\begin{aligned}\alpha_t(j) &= \sum_{i=1}^{N} \alpha_{t-1}(i)\, P(x_t, o_t | \boldsymbol{x}_1^{t-1}, \boldsymbol{o}_1^{t-1}, \lambda) \\ &= \sum_{i=1}^{N} \alpha_{t-1}(i)\, P(x_t | \boldsymbol{x}_1^{t-1}, \boldsymbol{o}_1^{t-1}, \lambda) \times \\ &\qquad P(o_t | x_t, \boldsymbol{x}_1^{t-1}, \boldsymbol{o}_1^{t-1}, \lambda). \end{aligned} \tag{36}$$

Now, we apply simplifying assumptions to yield,

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i)\, P(x_t = j | x_{t-1} = i, \lambda)\, P(o_t | x_t = j, \lambda), \tag{37}$$

where the probability of current state $x_t = j$ depends only on previous state $x_{t-1} = i$, and the current observation o_t depends only on the current state.

Ref: B. Gold & N. Morgan, *Speech and Audio Signal Processing*, New York: Wiley, pp.346–347, 2000.