Lecture # 10
Session 2003

Computation of $P(\boldsymbol{O}|\lambda)$

$$P(\boldsymbol{O}|\lambda) = \sum_{all \boldsymbol{Q}} P(\boldsymbol{O}, \boldsymbol{Q}|\lambda)$$

$$P(\boldsymbol{O}, \boldsymbol{Q}|\lambda) = P(\boldsymbol{O}|\boldsymbol{Q}, \lambda) P(\boldsymbol{Q}|\lambda)$$

- Consider the *fixed* state sequence: $\boldsymbol{Q} = q_1 q_2 \ldots q_T$

$$P(\boldsymbol{O}|\boldsymbol{Q}, \lambda) = b_{q_1}(o_1) b_{q_2}(o_2) \ldots b_{q_T}(o_T)$$

$$P(\boldsymbol{Q}|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \ldots a_{q_{T-1} q_T}$$

Therefore:

$$P(\boldsymbol{O}|\lambda) = \sum_{q_1, q_2, \ldots, q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \ldots a_{q_{T-1} q_T} b_{q_T}(o_T)$$

- Calculation required $\approx 2T \cdot N^T$ (there are N^T such sequences)
  For $N = 5, T = 100 \Rightarrow 2 \cdot 100 \cdot 5^{100} \approx 10^{72}$ computations!

The Viterbi Algorithm

1. Initialization:

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1), \qquad 1 \le i \le N \\ \psi_1(i) &= 0 \end{aligned}$$

2. Recursion:

$$\begin{aligned} \delta_t(j) &= \max_{1\le i\le N}[\delta_{t-1}(i)a_{ij}]b_j(o_t), \qquad 2 \le t \le T \qquad 1 \le j \le N \\ \psi_t(j) &= \operatorname*{argmax}_{1\le i\le N}[\delta_{t-1}(i)a_{ij}], \qquad 2 \le t \le T \qquad 1 \le j \le N \end{aligned}$$

3. Termination:

$$\begin{aligned} P^* &= \max_{1\le i\le N}[\delta_T(i)] \\ q_T^* &= \operatorname*{argmax}_{1\le i\le N}[\delta_T(i)] \end{aligned}$$

4. Path (state-sequence) backtracking:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \qquad t = T-1, T-2, \ldots, 1$$

Computation $\approx N^2 T$

Baum-Welch Re-estimation Formulas

$$\bar{\pi} = \text{expected number of times in state } s_i \text{ at } t = 1$$
$$= \gamma_1(i)$$

$$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } s_i \text{ to } s_j}{\text{expected number of transitions from state } s_i}$$
$$= \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\bar{b}_j(k) = \frac{\text{expected number of times in state } s_j \text{ with symbol } v_k}{\text{expected number of times in state } s_j}$$
$$= \frac{\sum_{\substack{t=1 \\ o_t = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$