3. Recognising the generating model – computing the emission probability of a feature sequence

Let a, b index two HMMS on $F^n \times K^n$

a: $p_a(s_1)$, $p_a(s_i | s_{i-1})$, $p_a(x_i | s_i)$

b: $p_b(s_1)$, $p_b(s_i | s_{i-1})$, $p_b(x_i | s_i)$

Given: a feature sequence $x \in F^n$; $x = (x_1, \dots, x_n)$

Question: which model has generated x?

Any reasonable answer is based on comparing $p_a(x)$, $p_b(x)$ $\Rightarrow$

Task: Compute

$$p(x) = \sum_{s \in K^n} p(x, s)$$

$$= \sum_{s_1 \in K} \dots \sum_{s_n \in K} p(s_1)\, p(x_1 | s_1) \prod_{i=2}^{n} p(s_i | s_{i-1})\, p(x_i | s_i)$$

Denote: $P_i(s_{i-1}, s_i) = p(s_i | s_{i-1}) \cdot p(x_i | s_i)$

$$\varphi_1(s_1) = p(s_1)\, p(x_1 | s_1)$$

$$\psi_n(s_n) \equiv 1$$

Expression in the sum reads

$$\varphi_1(s_1) \cdot P_2(s_1, s_2) \cdot \ldots \cdot P(s_{n-1}, s_n) \cdot \psi_n(s_n)$$

Summation can be performed iteratively from right to left

$$\psi_{i-1}(s_{i-1}) = \sum_{s_i \in K} P_i(s_{i-1}, s_i)\, \psi_i(s_i)$$

$$p(x) = \sum_{s_1 \in K} \varphi_1(s_1)\, \psi_1(s_1)$$

or from left to right

$$\varphi_i(s_i) = \sum_{s_{i-1}\in K} \varphi_{i-1}(s_{i-1}) P_i(s_{i-1}, s_i)$$

$$p(x) = \sum_{s_n \in K} \varphi_n(s_n) \psi_n(s_n)$$

In matrix-vector form we have

$$p(x) = \vec{\varphi}_1 \cdot P_2 \cdot P_3 \cdot \ldots \cdot P_n \cdot \vec{\psi}_n$$

Complexity: $O(n|K|^2)$

Remark The intermediate results have the following statistical meaning

$$\psi_i(s_i) = p(x_{i+1}, .., x_n \mid s_i)$$

$$\varphi_i(s_i) = p(x_1, .., x_i, s_i)$$

4. Recognising the most probable sequence of hidden states

We observe $x = (x_1, .., x_n)$ for a known HMM

Question Which sequence has generated x? If the answer is $s^* \in \operatorname{argmax}_{s\in K^n} p(x,s)$, whe have to solve

$$s^* \in \operatorname*{argmax}_{s\in K^n} \log p(x,s)$$

$$= \operatorname*{argmax}_{s\in K^n} \left\{ \log[p(s_1) p(x_1|s_1)] + \sum_{i=2}^{n} \log[p(s_i|s_{i-1}) p(x_i|s_i)] \right\}$$

Denote: $M_i(s_{i-1}, s_i) = \log[p(s_i|s_{i-1}) p(x_i|s_i)]$

$$\varphi_1(s_1) = \log[p(s_1) p(x_1|s_1)]$$

$$\psi_n(s_n) \equiv 0$$

Now, translate the algorithm from Sec 3 by replacing operators $\times \mapsto +$, $+ \mapsto \max$. E.g. maximising dynamically from right to left

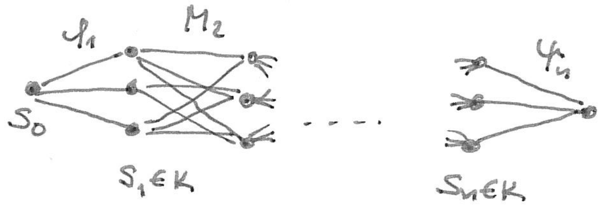
$$\psi_{i-1}(s_{i-1}) = \max_{s_i \in K} \left[M_i(s_{i-1}, s_i) + \psi_i(s_i) \right]$$

$$\max_{s \in K^n} \log p(x,s) = \max_{s_1 \in K} \left[\varphi_1(s_1) + \psi_1(s_1) \right]$$

Since we are looking for $s^* \in \arg\max_{s \in K} p(x,s)$: introduce pointers for $\arg\max_{s_i \in K} [\ldots]$ and backtrack an optimiser s^* starting from

$$s_1^* \in \arg\max_{s_1 \in K} \left[\varphi_1(s_1) + \psi_1(s_1) \right].$$

Similarly, we can optimise from left to right. Both variants search the best ~~graph~~ path in the graph



Complexity $\mathcal{O}(n|K|^2)$

5. Recognising the sequence of most probable hidden states

Is $s^* \in \arg\max_{s \in K^n}$ always the best answer to the question posed in Sec. 4?

The answer depends on $p(x,s)$ and the loss $\ell(s,s')$

a) If $\ell(s,s') = \mathbb{1}\{s' \neq s\}$, then $s^* \in \arg\max_{s \in K^n} p(x,s)$

b) If $\ell(s,s') = \sum_{i=1}^{n} \mathbb{1}\{s_i' \neq s_i\}$, i.e. Hamming distance, then

$$s^* \in \operatorname*{argmin}_{s \in K^n} \sum_{s' \in K^n} p(x, s') \sum_{i=1}^{n} [1 - \delta(s'_i, s_i)] \quad \Rightarrow$$

$$s_i^* \in \operatorname*{argmax}_{s_i \in K} p(x, s_i) \quad \forall\, i = 1, .., n$$

<u>Problem</u>: Compute prob's $p(x, s_i = k)$ $\forall\, i = 1,..,n$, $\forall\, k \in K$

Recall quantities φ, ψ from Sec. 3

$$\varphi_i(s_i) = p(x_1, .., x_i, s_i)$$
$$\psi_i(s_i) = p(x_{i+1}, .., x_n \mid s_i)$$

Since $p(x \mid s_i) \overset{!}{=} p(x_1, .., x_i \mid s_i)\, p(x_{i+1}, .., x_n \mid s_i)$ holds for HMMs, we have

$$p(x, s_i) = \varphi_i(s_i)\, \psi_i(s_i)$$

<u>Complexity</u>: $\mathcal{O}(2n|K|^2)$