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

Let \( a, b \) index two HMMs on \( F^n x K^n \)

\[ a: p_a(s_i), p_a(s_i | s_{i-1}), p_a(x_i | s_{i}) \]

\[ b: p_b(s_i), p_b(s_i | s_{i-1}), p_b(x_i | s_{i}) \]

Given: a feature sequence \( x \in F^n, x = (x_0, \ldots, x_n) \)

Question: which model has generated \( x \)?

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

Task: Compute

\[
p(x) = \sum_{s \in K^n} p(x, s) = \prod_{i=1}^{n} p(s_i) p(x_i | s_i) \prod_{i=2}^{n} p(s_i | s_{i-1}) p(x_i | s_{i})
\]

Denote:

\[
\Psi_i(s_i, s_{i-1}) = p(s_i | s_{i-1}) \cdot p(x_i | s_i)
\]

\[
\Psi_i(s_i) = p(s_i) \cdot p(x_i | s_i)
\]

\[ \Psi_n(s_n) = 1 \]

Expression in the sum reads

\[
\Psi_i(s_i) \cdot \Psi_2(s_2) \cdots \Psi_1(s_1, s_2) \cdots \Psi(s_{n-1}, s_n) \cdot \Psi_n(s_n)
\]

Summation can be performed iteratively from right to left

\[
\Psi_{i-1}(s_{i-1}) = \sum_{s_i \in K} \Psi_i(s_{i-1}, s_i) \cdot \Psi_i(s_i)
\]

\[
p(x) = \sum_{s \in K} \Psi_1(s_1) \Psi_2(s_2) \cdots \Psi_n(s_n)
\]
or from left to right
\[ \Psi_i(s_i) = \sum_{s_{i-1} \in K} \Psi_{i-1}(s_{i-1}) P_i(s_{i-1}, s_i) \]
\[ P(x) = \sum_{s_n \in K} \Psi_n(s_n) y_n(s_n) \]
In matrix-vector form we have
\[ \Psi(x) = \Psi_n \cdot P_n \cdot P_{n-1} \cdots P_2 \cdot P_1 \cdot \Psi_1 \]
Complexity: \( O(n|K|^2) \)

Remark. The intermediate results have the following statistical meaning
\[ \Psi_i(s_i) = P(x_{i-1}, x_n | s_i) \]
\[ \Psi_i(s_i) = P(x_{i-1}, x_n, s_i) \]

4. Recognising the most probable sequence of hidden states

We observe \( x = (x_1, \ldots, x_n) \) for a known HMM

Question: Which sequence has generated \( x \)? If the answer is

\[ S^* = \arg \max_{S \in K^n} p(x|S) \text{, we have to solve} \]

\[ S^* = \arg \max_{S \in K^n} \log p(x|S) \]

\[ = \arg \max_{S \in K^n} \left\{ \log [p(S; x_{i=1})] + \sum_{i=2}^n \log [p(S; x_{i=1}, x_{i-1}) p(x_i|S_{i-1})] \right\} \]

Denote:
\[ M_i(s_{i-1}, s_i) = \log [p(S; x_{i-1}, x_i)] \]
\[ \Psi_i(s_i) = \log [p(S; x_{i-1}, x_i)] \]
\[ \Psi_n(s_n) = 0 \]
Now, translate the algorithm from Sec 3 by replacing operators
\( x \rightarrow +, + \rightarrow \max \). E.g. maximising dynamically from right to left

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

\[
\max_{s_i \in K} \log p(x, s) = \max_{s_i \in K} \left[ \Psi_i(s_i) + \Psi_i(s_i) \right]
\]

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

\[
\Psi_i(s_i) \in \arg \max_{s_i \in K} \left[ \Psi_i(s_i) + \Psi_i(s_i) \right].
\]

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

\[\text{Complexity } O(n \cdot k^2)\]

\( s: \text{Recognising the sequence of most probable hidden states} \)

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

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

a) If \( c(s, s') = \mathbb{1}[s \neq s'] \), then \( s^* \in \arg \max_{s_i \in K} p(x, s) \)

b) If \( c(s, s') = \sum_{i=1}^m \mathbb{1}[s_{i} \neq s_{i}] \), i.e. Hamming distance, then
\[ s^* \in \arg \min_{s \in K} \sum_{i=1}^{n} p(x_i, s') \sum_{s_i'} [1 - \delta(s_i', s_i)] \Rightarrow \]

\[ s_i^* \in \arg \max_{s_i \in K} p(x_i, s_i) \quad \forall \; i = 1, \ldots, n \]

**Problem:** Compute prob's \( p(x, s_i = k) \) \( \forall \; i = 1, \ldots, n \) \( \forall \; k \in K \)

Recall quantities \( \psi_i, \psi_i^* \) from Sec. 3

\[ \psi_i(s_i) = p(x_{i+1}, \ldots, x_n | s_i) \]

\[ \psi_i^*(s_i) = p(x_{i+1}, \ldots, x_n | s_i) \]

Since \( p(x|s_i) = p(x_{i+1}, x_i | s_i) p(x_{i+1}, \ldots, x_n | s_i) \) holds for \( \text{HMM}_i \), we have

\[ p(x, s_i) = \psi_i(s_i) \psi_i^*(s_i) \]

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