Inference in Hidden Markov Models Part 2

Wenhu Chen

Lecture 11

Readings: RN 14.2.2.

Outline

Learning Goals

- Calculate the smoothing probability for a time step in a hidden Markov model.
- Describe the justification for a step in the derivation of the smoothing formulas.
- Describe the forward-backward algorithm.
- Describe the Viterbi algorithm.

The Umbrella Model

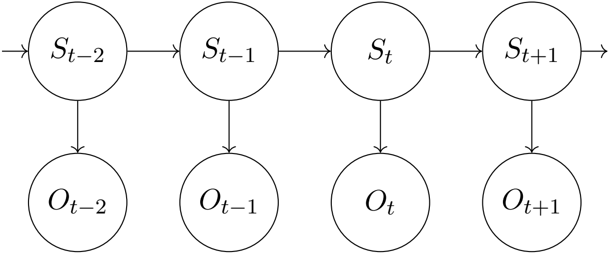
Let S_t be $true$ if it rains on day t and $false$ otherwise.

Let O_t be $true$ if the director carries an umbrella on day t and $false$ otherwise.

$$P(s_0) = 0.5$$

$$P(s_t|s_{t-1}) = 0.7$$
$$P(s_t|\neg s_{t-1}) = 0.3$$

$$P(o_t|s_t) = 0.9$$
$$P(o_t|\neg s_t) = 0.2$$



Smoothing

Given the observations from day 0 to day $t-1$, what is the probability that I am in a particular state on day k?

$$P(S_k | o_{0:(t-1)}), \text{ where } 0 \leq k \leq t-1$$

Smoothing through Backward Recursion

Calculating the smoothed probability $P(S_k|o_{0:(t-1)})$:

$$\begin{aligned} & P(S_k|o_{0:(t-1)}) \\ & = \alpha\, P(S_k|o_{0:k})\, P(o_{(k+1):(t-1)}|S_k) \\ & = \alpha\, f_{0:k}\, b_{(k+1):(t-1)} \end{aligned}$$

Calculate $f_{0:k}$ using forward recursion.
Calculate $b_{(k+1):(t-1)}$ using backward recursion.

Backward Recursion:

Base case:

$$b_{t:(t-1)} = \vec{1}.$$

Recursive case:

$$b_{(k+1):(t-1)} = \sum_{s_{k+1}} P(o_{k+1}|s_{k+1}) b_{(k+2):(t-1)} P(s_{k+1}|S_k).$$

A Smoothing Example

Consider the umbrella story.

Assume that $O_0 = true$, $O_1 = true$, and $O_2 = true$.

What is the probability that it rained on day 0 ($P(S_0|o_0 \wedge o_1 \wedge o_2)$) and the probability it rained on day 1 ($P(S_1|o_0 \wedge o_1 \wedge o_2)$)?

Here are the useful quantities from the umbrella story:

$$P(s_0) = 0.5$$
$$P(o_t|s_t) = 0.9, P(o_t|\neg s_t) = 0.2$$
$$P(s_t|s_{(t-1)}) = 0.7, P(s_t|\neg s_{(t-1)}) = 0.3$$

A Smoothing Example

Calculate $P(S_1|o_{0:2})$.

A Smoothing Example

Calculate $P(S_1|o_{0:2})$.

(1) What are the values of k and t?

$$P(S_1|o_{0:2}) = P(S_k|o_{0:(t-1)}) \Rightarrow k = 1, t = 3$$

A Smoothing Example

Calculate $P(S_1|o_{0:2})$.

(1) What are the values of k and t?

$$P(S_1|o_{0:2}) = P(S_k|o_{0:(t-1)}) \Rightarrow k = 1, t = 3$$

(2) Write the probability as a product of forward and backward messages.

$$\begin{aligned} &P(S_1|o_{0:2}) \\ &= \alpha\, P(S_1|o_{0:1}) * P(o_{2:2}|S_1) \\ &= \alpha\, f_{0:1} * b_{2:2} \end{aligned}$$

A Smoothing Example

Calculate $P(S_1|o_{0:2})$.

(1) What are the values of k and t?

$$P(S_1|o_{0:2}) = P(S_k|o_{0:(t-1)}) \Rightarrow k = 1, t = 3$$

(2) Write the probability as a product of forward and backward messages.

$$\begin{aligned} &P(S_1|o_{0:2}) \\ &= \alpha\, P(S_1|o_{0:1}) * P(o_{2:2}|S_1) \\ &= \alpha\, f_{0:1} * b_{2:2} \end{aligned}$$

(3) We already calculated $f_{0:1} = \langle 0.883, 0.117 \rangle$ in the last lecture. Next, we will calculate $b_{2:2}$ using backward recursion.

A Backward Recursion Example - Recursive Case

Calculate $b_{2:2} = P(o_{2:2}|S_1)$ where $k = 1, t = 3$.

A Backward Recursion Example - Recursive Case

Calculate $b_{2:2} = P(o_{2:2}|S_1)$ where $k = 1, t = 3$.

$$
\begin{aligned}
b_{2:2} &= P(o_{2:2}|S_1) \\
&= \sum_{s_2} P(o_2|s_2) * b_{3:2} * P(s_2|S_1) \\
&= \sum_{s_2} P(o_2|s_2) * P(o_{3:2}|s_2) * P(s_2|S_1) \\
&= \sum_{s_2} P(o_2|s_2) * P(o_{3:2}|s_2) * \langle P(s_2|s_1), P(s_2|\neg s_1)\rangle \\
&= \Big(P(o_2|s_2) * P(o_{3:2}|s_2) * \langle P(s_2|s_1), P(s_2|\neg s_1)\rangle \\
&\quad + P(o_2|\neg s_2) * P(o_{3:2}|\neg s_2) * \langle P(\neg s_2|s_1), P(\neg s_2|\neg s_1)\rangle \Big)
\end{aligned}
$$

A Backward Recursion Example - Recursive Case

Calculate $b_{2:2} = P(o_{2:2}|S_1)$ where $k = 1, t = 3$.

A Backward Recursion Example - Recursive Case

Calculate $b_{2:2} = P(o_{2:2}|S_1)$ where $k = 1, t = 3$.

$$
\begin{aligned}
b_{2:2} &= P(o_{2:2}|S_1) \\
&= \Big(P(o_2|s_2) * P(o_{3:2}|s_2) * \langle P(s_2|s_1), P(s_2|\neg s_1) \rangle \\
&\qquad + P(o_2|\neg s_2) * P(o_{3:2}|\neg s_2) * \langle P(\neg s_2|s_1), P(\neg s_2|\neg s_1) \rangle \Big) \\
&= \Big(0.9 * 1 * \langle 0.7, 0.3 \rangle + 0.2 * 1 * \langle 0.3, 0.7 \rangle \Big) \\
&= (0.9 * \langle 0.7, 0.3 \rangle + 0.2 * \langle 0.3, 0.7 \rangle) \\
&= (\langle 0.63, 0.27 \rangle + \langle 0.06, 0.14 \rangle) \\
&= \langle 0.69, 0.41 \rangle
\end{aligned}
$$

A Smoothing Example

Calculate $P(S_1|o_{0:2})$.

A Smoothing Example

Calculate $P(S_1|o_{0:2})$.

$$
\begin{aligned}
&P(S_1|o_{0:2}) \\
&= \alpha\, P(S_1|o_{0:1}) * P(o_{2:2}|S_1) \\
&= \alpha\, f_{0:1} * b_{2:2} \\
&= \alpha \langle 0.883, 0.117 \rangle * \langle 0.69, 0.41 \rangle \\
&= \alpha \langle 0.6093, 0.0480 \rangle \\
&= \langle 0.927, 0.073 \rangle
\end{aligned}
$$

A Smoothing Example

Calculate $P(S_0|o_{0:2})$.

A Smoothing Example

Calculate $P(S_0|o_{0:2})$.

$$k = 0, t = 3$$

$$\begin{aligned} b_{1:2} &= P(o_{1:2}|S_0) \\ &= (P(o_1|s_1) * P(o_{2:2}|s_1) * \langle P(s_1|s_0), P(s_1|\neg s_0)\rangle \\ &\quad + P(o_1|\neg s_1) * P(o_{2:2}|\neg s_1) * \langle P(\neg s_1|s_0), P(\neg s_1|\neg s_0)\rangle) \\ &= (0.9 * 0.69 * \langle 0.7, 0.3\rangle + 0.2 * 0.41 * \langle 0.3, 0.7\rangle) \\ &= \langle 0.4593, 0.2437\rangle \end{aligned}$$

A Smoothing Example

Calculate $P(S_0|o_{0:2})$.

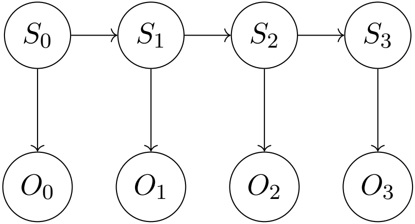
$$k = 0, t = 3$$

$$\begin{aligned}
b_{1:2} &= P(o_{1:2}|S_0) \\
&= (P(o_1|s_1) * P(o_{2:2}|s_1) * \langle P(s_1|s_0), P(s_1|\neg s_0)\rangle \\
&\quad + P(o_1|\neg s_1) * P(o_{2:2}|\neg s_1) * \langle P(\neg s_1|s_0), P(\neg s_1|\neg s_0)\rangle) \\
&= (0.9 * 0.69 * \langle 0.7, 0.3\rangle + 0.2 * 0.41 * \langle 0.3, 0.7\rangle) \\
&= \langle 0.4593, 0.2437\rangle
\end{aligned}$$

$$\begin{aligned}
P(S_0|o_{0:2}) &= \alpha\, f_{0:0} * b_{1:2} \\
&= \alpha\langle 0.818, 0.182\rangle * \langle 0.4593, 0.2437\rangle \\
&= \langle 0.894, 0.106\rangle
\end{aligned}$$

Smoothing: Example 2

Consider a hidden Markov model with 4 time steps.



$P(s_0) = 0.4$

$P(s_t|s_{t-1}) = 0.7$
$P(s_t|\neg s_{t-1}) = 0.2$

$P(o_t|s_t) = 0.9$
$P(o_t|\neg s_t) = 0.2$

Calculate $P(S_2|o_0 \wedge o_1 \wedge o_2 \wedge \neg o_3)$.

Smoothing (time k)

How can we derive the formula for $P(S_k|o_{0:(t-1)}), 0 \leq k \leq t-1$?

$$
\begin{aligned}
& P(S_k|o_{0:(t-1)}) \\
& = P(S_k|o_{(k+1):(t-1)} \wedge o_{0:k}) \\
& = \alpha \, P(S_k|o_{0:k}) \, P(o_{(k+1):(t-1)}|S_k \wedge o_{0:k}) \\
& = \alpha \, P(S_k|o_{0:k}) \, P(o_{(k+1):(t-1)}|S_k) \\
& = \alpha f_{0:k} b_{(k+1):(t-1)}
\end{aligned}
$$

Calculate $f_{0:k}$ through forward recursion.

Calculate $b_{(k+1):(t-1)}$ through backward recursion.

Q: Smoothing Derivation

Q #1: What is the justification for the step below?

$$P(S_k|o_{0:(t-1)})$$
$$= P(S_k|o_{(k+1):(t-1)} \wedge o_{0:k})$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Smoothing Derivation

Q #1: What is the justification for the step below?

$$P(S_k | o_{0:(t-1)})$$
$$= P(S_k | o_{(k+1):(t-1)} \wedge o_{0:k})$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

$\rightarrow$ Correct answer is (B) Re-writing the expression.

Q: Smoothing Derivation

Q #2: What is the justification for the step below?

$$\begin{aligned} &= P(S_k | o_{(k+1):(t-1)} \land o_{0:k}) \\ &= \alpha\, P(S_k | o_{0:k})\, P(o_{(k+1):(t-1)} | S_k \land o_{0:k}) \end{aligned}$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Smoothing Derivation

Q #2: What is the justification for the step below?

$$= P(S_k | o_{(k+1):(t-1)} \wedge o_{0:k})$$
$$= \alpha \, P(S_k | o_{0:k}) \, P(o_{(k+1):(t-1)} | S_k \wedge o_{0:k})$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

→ Correct answer is (A) Bayes' rule.

Q: Smoothing Derivation

Q #3: What is the justification for the step below?

$$
\begin{aligned}
&= \alpha\, P(S_k | o_{0:k})\, P(o_{(k+1):(t-1)} | S_k \wedge o_{0:k}) \\
&= \alpha\, P(S_k | o_{0:k})\, P(o_{(k+1):(t-1)} | S_k)
\end{aligned}
$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Smoothing Derivation

Q #3: What is the justification for the step below?

$$= \alpha \, P(S_k | o_{0:k}) \, P(o_{(k+1):(t-1)} | S_k \wedge o_{0:k})$$
$$= \alpha \, P(S_k | o_{0:k}) \, P(o_{(k+1):(t-1)} | S_k)$$

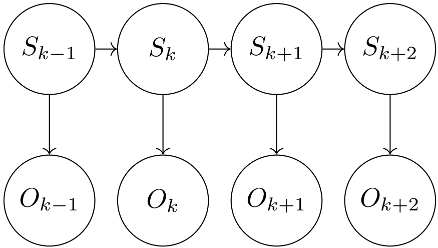
(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule



→ Correct answer is (D) The Markov assumption.

Backward Recursion Formula Derivations

How did we derive the formula for backward recursion?

$$
\begin{aligned}
& P(o_{(k+1):(t-1)} | S_k) \\
&= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} \land s_{(k+1)} | S_k) && (1) \\
&= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)} \land S_k) * P(s_{(k+1)} | S_k) && (2) \\
&= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k) && (3) \\
&= \sum_{s_{(k+1)}} P(o_{(k+1)} \land o_{(k+2):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k) && (4) \\
&= \sum_{s_{(k+1)}} P(o_{(k+1)} | s_{(k+1)}) * P(o_{(k+2):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k) && (5)
\end{aligned}
$$

Q: Backward Recursion Derivation

Q #4: What is the justification for the step below?

$$P(o_{(k+1):(t-1)} | S_k)$$
$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} \wedge s_{(k+1)} | S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Backward Recursion Derivation

Q #4: What is the justification for the step below?

$$P(o_{(k+1):(t-1)}|S_k)$$
$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} \wedge s_{(k+1)}|S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

→ Correct answer is (E) The sum rule.

Q: Backward Recursion Derivation

Q #5: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} \wedge s_{(k+1)} | S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)} \wedge S_k) P(s_{(k+1)} | S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Backward Recursion Derivation

Q #5: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} \wedge s_{(k+1)} | S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)} \wedge S_k) P(s_{(k+1)} | S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

→ Correct answer is (C) The chain/product rule.

Q: Backward Recursion Derivation

Q #6: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)} \wedge S_k) P(s_{(k+1)} | S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)}) P(s_{(k+1)} | S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Backward Recursion Derivation

Q #6: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)} \wedge S_k) P(s_{(k+1)} | S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)}) P(s_{(k+1)} | S_k)$$

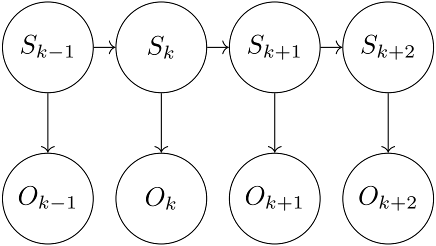
(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule



→ Correct answer is (D) The Markov assumption.

Q: Backward Recursion Derivation

Q #7: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)}|s_{(k+1)}) * P(s_{(k+1)}|S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1)} \wedge o_{(k+2):(t-1)}|s_{(k+1)}) * P(s_{(k+1)}|S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Backward Recursion Derivation

Q #7: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1)} \wedge o_{(k+2):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

→ Correct answer is (B) Re-writing the expression.

Q: Backward Recursion Derivation

Q #8: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1)} \wedge o_{(k+2):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1)} | s_{(k+1)}) * P(o_{(k+2):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k)$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

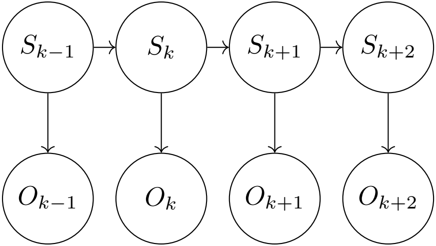
(E) The sum rule

Q: Backward Recursion Derivation

Q #8: What is the justification for the step below?

$$= \sum_{s_{(k+1)}} P(o_{(k+1)} \wedge o_{(k+2):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k)$$

$$= \sum_{s_{(k+1)}} P(o_{(k+1)} | s_{(k+1)}) * P(o_{(k+2):(t-1)} | s_{(k+1)}) * P(s_{(k+1)} | S_k)$$



(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

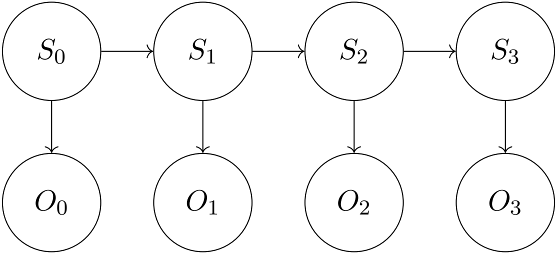
(D) The Markov assumption

(E) The sum rule

→ Correct answer is (D) The Markov assumption.

The Forward-Backward Algorithm

For a hidden Markov model with any number of time steps, we can calculate the smoothed probabilities using one forward pass and one backward pass through the network.



$$
\begin{aligned}
& P(S_k | o_{0:(t-1)}) \\
& = \alpha \, P(S_k | o_{0:k}) \, P(o_{(k+1):(t-1)} | S_k) \\
& = \alpha \, f_{0:k} \, b_{(k+1):(t-1)}
\end{aligned}
$$

Finding most likely explanation

We have observed all the states $o_{0:t-1}$ and want to decode all the hidden states $s_{0:t-1}$. Here we make a more general assumption:

- S_t is not boolean variable, $S_t \in \{0, 1, 2, \cdots, n-1\}$.
- The time spans from 0 to $t-1$.
- The transition matrix $A \in \mathbb{R}^{n \times n}$ and emission matrix $O^{n \times o}$ are already given, where o is the possible observations.

$$\hat{s}_0, \cdots, \hat{s}_{t-1} = \arg\max_{S_0:S_{t-1}} p(S_0, \cdots, S_{t-1} | o_{0:t-1})$$

Brutal-Force Decoding

Loop through all the possible $S_{0:t-1}$, and then compute their likelihood $p(S_{0:t-1}|o_{0:t-1})$ to find the maximum.

- $S_0 = T, S_1 = T, S_2 = T, \cdots, S_{t-1} = T$
- $S_0 = T, S_1 = T, S_2 = T, \cdots, S_{t-1} = F$
- $\ldots$
- $S_0 = F, S_1 = F, S_2 = F, \cdots, S_{t-1} = F$

The complexity is $O(n^t)$, which is extremely expensive.

Dynamic Programming

Assuming we have a sequence $S_{0:k}$ ending at $S_k = j$, we can define a function $r(S_k = j, S_{0:k-1})$ as:

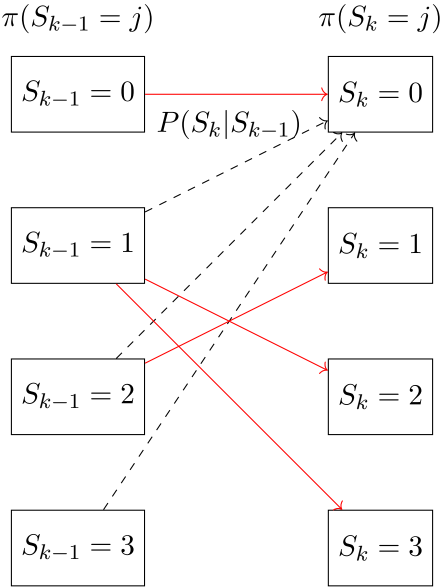
$$r(S_k = j, S_{0:k-1}) = P(S_{0:k-1}, S_k = j | o_{0:k})$$

Define an auxiliary function $\pi_k(j)$ as:

$$\begin{aligned}\pi_k(j) &= \max_{S_{0:k-1}} r(S_k = j, S_{0:k-1}) \\ &= \max_{S_{0:k-1}; s.t. S_{k-1}=j} P(S_{0:k-1}, S_k = j | o_{0:k})\end{aligned}$$

By definition, we have:
$\pi_k(j)$ denotes the maximum probability of any sequence $S_{0:k}$ ending with $S_k = j$ under the current observations.

Dynamic Programming



Dynamic Programming

Base case:
For 0-th step, we have:

$$\pi_0(j) = \alpha P(S_0) P(o_0 | S_0)$$

Recursive definition:
For any $k \in \{1, \cdots, t-1\}$, for any j, we have:

$$\pi_k(j) = P(o_k | S_k = j) \max_z [\pi_{k-1}(z) P(S_k = j | S_{k-1} = z)]$$
$$\phi_k(j) = \arg\max_z [\pi_{k-1}(z) P(S_k = j | S_{k-1} = z)]$$

Viterbi Algorithm

Given: π_0 and probabilities P. Return $\hat{s}$ as the output.

- For $k = 1, \cdots, t-1$
  - For $j = 0, \cdots n-1$

$$\pi_k(j) = P(o_k | S_k = j) \max_z [\pi_{k-1}(z) P(S_k = j | S_{k-1} = z)]$$

$$\phi_k(j) = \arg\max_z [\pi_{k-1}(z) P(S_k = j | S_{k-1} = z)]$$

  Find last state $\hat{s}_{t-1} = \arg\max_j \pi_{t-1}(j)$.

- For $k = t-1, \cdots, 1$

$$\hat{s}_{k-1} = \phi_k(\hat{s}_k)$$

- Return $\hat{s} = \hat{s}_0, \cdots, \hat{s}_{t-1}$.

Viterbi Algorithm

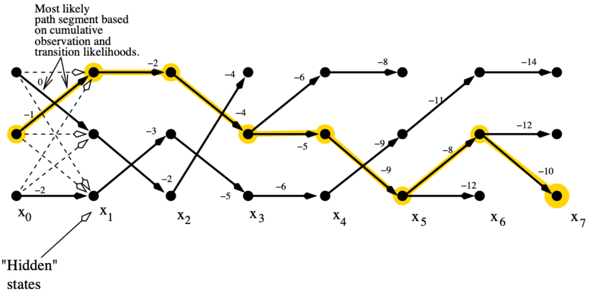


Figure: viterbi algorithm visualization.

Given the length of the sequence as t, and the number of states as n, the time complexity is $O(t \times n^2)$

Revisiting Learning Goals

- Calculate the smoothing probability for a time step in a hidden Markov model.
- Describe the justification for a step in the derivation of the smoothing formulas.
- Describe the forward-backward algorithm.
- Describe the Viterbi algorithm.