Inference in Hidden Markov Models Part 2

Alice Gao

Lecture 15

Readings: RN 15.2.3.

Outline

Learning Goals

By the end of the lecture, you should be able to

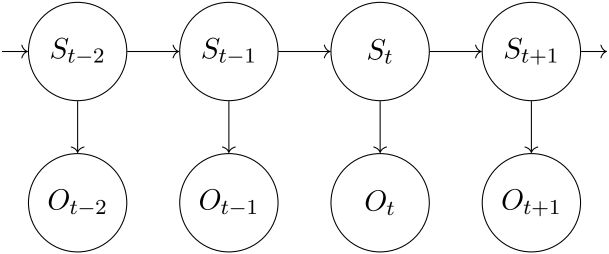
- Calculate the smoothing probability for a time step in a hidden Markov model.
- Describe the justification for a step in the derivation of the smoothing formulas.
- Describe the forward-backward algorithm.

The Umbrella Model

$$P(s_0) = 0.5$$

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

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



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?

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

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 $f_{0:k}$ through forward recursion.
Calculate $b_{(k+1):(t-1)}$ through 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 $O_0 = t$, $O_1 = t$, and $O_2 = t$.

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

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?

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

(2) Write the probability as a product of two 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 $f_{0:1} = \langle 0.883, 0.117 \rangle$.
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 $P(S_0|o_{0:2})$.

$$k = 0, t = 3$$

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

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

Smoothing (day k)

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

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

Smoothing Derivation 1/3

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

Smoothing Derivation 2/3

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

Smoothing Derivation 3/3

What is the justification for the step below?

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

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Backward Recursion Formula Derivations

How did we derive the formula for backward recursion?

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

Backward Recursion Derivation 1/5

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

Backward Recursion Derivation 2/5

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

Backward Recursion Derivation 3/5

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

Backward Recursion Derivation 4/5

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

Backward Recursion Derivation 5/5

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

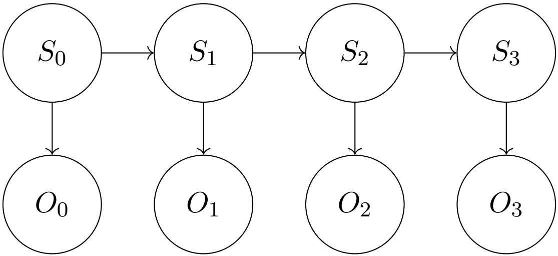
(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

The Forward-Backward Algorithm

Consider a hidden Markov model with 4 time steps.
We can calculate the smoothed probabilities using
one forward pass and one backward pass through the network.



Revisiting the Learning Goals

By the end of the lecture, you should be able to

- 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.