Inference in Hidden Markov Models Part 1

Wenhu Chen

Lecture 10

Readings: RN 14.1 & 14.2.1, PM 8.5.1 - 8.5.3.

Outline

Learning Goals

- Construct a hidden Markov model given a real-world scenario.
- Explain the independence assumptions in a hidden Markov model.
- Calculating the filtering probability for a time step in a hidden Markov model.
- Describe the justification for a step in the derivation of the filtering formulas.

Inference in a Changing World

So far, we can reason probabilistically in a static world.
However, the world evolves over time.

In an evolving world, we have to reason about a sequence of events.

Applications of sequential belief networks:

- weather predictions
- stock market predictions
- patient monitoring
- robot localization

  → A robot is trying to figure out where it is.
- speech and handwriting recognition

Speech Recognition

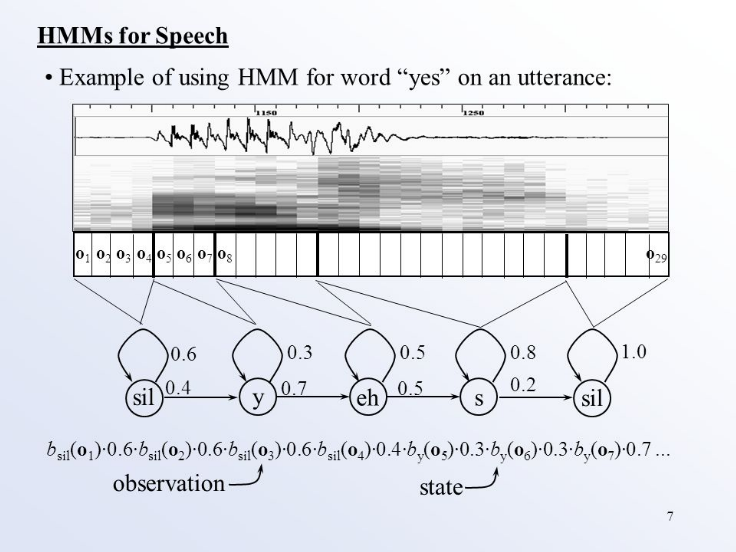


Figure: Speech Recognition

Running Example: the Umbrella Story

You are a security guard stationed at a secret underground facility.

You want to know whether it's raining today.

Unfortunately, your only access to the outside world occurs each morning when you see the director coming in with, or without, an umbrella.

States and Observations

- The world contains a series of time slices.
- Each time slice contains a set of random variables,
  Let S_t denote the unobservable state at time t.

  Let O_t denote the signal/observation at time t.

What are the random variables in the umbrella world?

States and Observations

- The world contains a series of time slices.
- Each time slice contains a set of random variables,
  Let S_t denote the unobservable state at time t.

  Let O_t denote the signal/observation at time t.

What are the random variables in the umbrella world?

$\rightarrow$ S_t denotes whether it rains at time t.

O_t denotes whether the director carries an umbrella at time t.

Transition Model

How does the current state depend on the previous states?

In general, every state may depend on all the previous states.

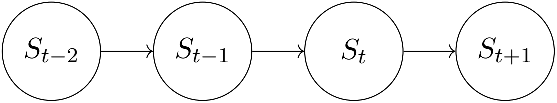
$$P(S_t | S_{t-1} \land S_{t-2} \land S_{t-3} \land \cdots \land S_0)$$

Problem: As t increases, the conditional probability distribution can be unboundedly large.

Solution: Let the current state depend on a fixed number of previous states.

K-order Markov Chain

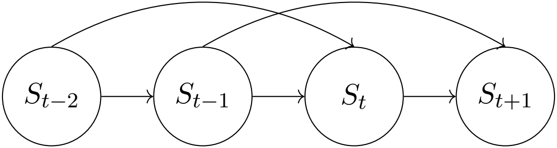
First-order Markov process:



The transition model:

$$P(S_t|S_{t-1} \land S_{t-2} \land S_{t-3} \land \cdots \land S_0) = P(S_t|S_{t-1})$$

Second-order Markov process:



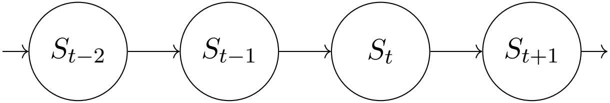
The transition model:

$$P(S_t|S_{t-1} \land S_{t-2} \land S_{t-3} \land \cdots \land S_0) = P(S_t|S_{t-1} \land S_{t-2})$$

The Markov Assumption

The Markov assumption:
The future is independent of the past given the present.

Every day, our slate is wiped clean. We can start fresh. Every day is a new beginning.



The transition model:

$$P(S_t | S_{t-1} \wedge S_{t-2} \wedge S_{t-3} \wedge \cdots \wedge S_0) = P(S_t | S_{t-1})$$

Stationary Process

Is there a different conditional probability distribution
for each time step?

Stationary process:

- The dynamics does not change over time.
- The conditional probability distribution for each time step remains the same.

What are the advantages of using a stationary model?

$\rightarrow$ Simple to specify.

Natural: the dynamics typically does not change.
If it changes, it's due to another feature that we can model.

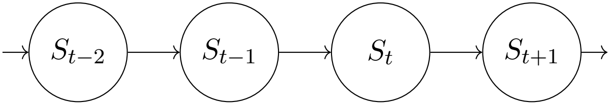
A finite number of parameters gives an infinite network.

Transition model for the umbrella story

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

$$\boxed{P(s_0) = 0.5}$$

$$\boxed{\begin{array}{l} P(s_t | s_{t-1}) = 0.7 \\ P(s_t | \neg s_{t-1}) = 0.3 \end{array}}$$



Warm-up Example

What's the chance of raining in day 1?

$$p(s_1 = T) = 0.5 * 0.7 + 0.5 * 0.3 = 0.5$$
$$p(s_1 = F) = 0.5 * 0.3 + 0.5 * 0.7 = 0.5$$

What's the chance of raining in day 2?

$$p(s_2 = T) = 0.5 * 0.7 + 0.5 * 0.3 = 0.5$$
$$p(s_2 = F) = 0.5 * 0.3 + 0.5 * 0.7 = 0.5$$

On day K?

$$p(s_K = T) = p(s_K = F) = 0.5$$

Sensor model

How does the evidence variable O_t at time t depend on the previous and current states $S_0, S_1, \ldots, S_t$?

(Sensor) Markov assumption:

Each state is sufficient to generate its observation.

$$P(O_t | S_t \land S_{t-1} \land \cdots \land S_0 \land O_{t-1} \land O_{t-2} \land \cdots \land O_0) = P(O_t | S_t)$$

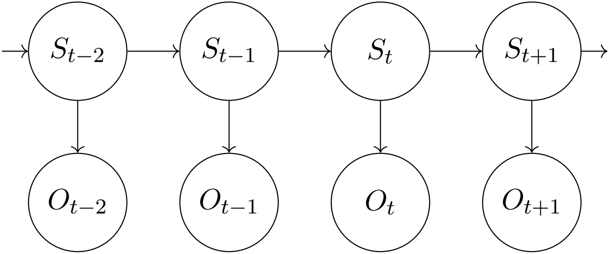
Complete model for the umbrella story

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.

$$\boxed{P(s_0) = 0.5}$$

$$\boxed{\begin{array}{l} P(s_t|s_{t-1}) = 0.7 \\ P(s_t|\neg s_{t-1}) = 0.3 \end{array}} \qquad \boxed{\begin{array}{l} P(o_t|s_t) = 0.9 \\ P(o_t|\neg s_t) = 0.2 \end{array}}$$



Warm-up Example

What's the chance of the director carrying an umbrella on day 1?

$$p(O_1 = T) = 0.5 * 0.9 + 0.5 * 0.2 = 0.55$$
$$p(O_1 = F) = 0.5 * 0.1 + 0.5 * 0.8 = 0.45$$

What's the chance of the director carrying an umbrella on day 2?

$$p(O_2 = T) = 0.5 * 0.9 + 0.5 * 0.2 = 0.55$$
$$p(O_2 = F) = 0.5 * 0.1 + 0.5 * 0.8 = 0.45$$

On day K?

$$p(O_K = T) = 0.55, p(O_K = F) = 0.45$$

Hidden Markov Model

Hidden Markov Model:

- A Markov process
- The state variables are unobservable
- The evidence variables, which depend on the states, are observable

Common Inference Tasks

- **Filering:** Which state am I in right now?

  $\rightarrow$ The posterior distribution over the most recent state given all evidence to date.

Common Inference Tasks

- **Filtering:** Which state am I in right now?

  $\rightarrow$ The posterior distribution over the most recent state given all evidence to date.

- **Prediction:** Which state will I be in tomorrow?

  $\rightarrow$ The posterior distribution over the future state given all evidence to date.

Common Inference Tasks

- **Filtering:** Which state am I in right now?

  → The posterior distribution over the most recent state given all evidence to date.

- **Prediction:** Which state will I be in tomorrow?

  → The posterior distribution over the future state given all evidence to date.

- **Smoothing:** Which state was I in yesterday?

  → The posterior distribution over a past state, given all evidence to date.

Common Inference Tasks

- **Filtering:** Which state am I in right now?

  → The posterior distribution over the most recent state given all evidence to date.

- **Prediction:** Which state will I be in tomorrow?

  → The posterior distribution over the future state given all evidence to date.

- **Smoothing:** Which state was I in yesterday?

  → The posterior distribution over a past state, given all evidence to date.

- **Most likely explanation:** Which sequence of states is most likely to have generated the observations? → Find the sequence of states that is most likely to have generated all the evidence to date.

Algorithms for the inference tasks

A HMM is a Bayesian network.
We can perform inference using the variable elimination algorithm!

More specialized algorithms:

- The forward-backward algorithm: filtering and smoothing
- The Viterbi algorithm: most likely explanation

Filtering

Given the observations from time 0 to time k, what is the probability that I am in a particular state at time k?

$$P(S_k | o_{0:k})$$

Warm-up Question

I already know that the manager brought umbrella yesterday, but he did not bring umbrella today. What's the chance of raining for today, e.g. $p(S_1|o_0, \neg o_1)$?

$$
\begin{aligned}
p(s_0) &= 0.5 \\
P(s_t|s_{t-1}) &= 0.7 \\
P(s_t|\neg s_{t-1}) &= 0.3 \\
P(o_t|s_t) &= 0.9 \\
P(o_t|\neg s_t) &= 0.2
\end{aligned}
$$

Is s_1 independent of o_0 given o_1?

Warm-up Question

I already know that the manager brought umbrella yesterday, but he did not bring umbrella today. What's the chance of raining for today, e.g. $p(S_1|o_0, \neg o_1)$?

$$
\begin{aligned}
p(s_0) &= 0.5 \\
P(s_t|s_{t-1}) &= 0.7 \\
P(s_t|\neg s_{t-1}) &= 0.3 \\
P(o_t|s_t) &= 0.9 \\
P(o_t|\neg s_t) &= 0.2
\end{aligned}
$$

$$
\begin{aligned}
p(S_1|o_0, \neg o_1) &\propto p(S_1, o_0, \neg o_1) \\
&= \sum_{S_0} p(S_0) p(o_0|S_0) p(S_1|S_0) p(\neg o_1|S_1)
\end{aligned}
$$

Warm-up Question

I already know that the manager brought umbrella yesterday, but he did not bring umbrella today. What's the chance of raining for today, e.g. $p(S_1|o_0, \neg o_1)$?

$$\begin{aligned} p(s_0) &= 0.5 \\ P(s_t|s_{t-1}) &= 0.7 \\ P(s_t|\neg s_{t-1}) &= 0.3 \\ P(o_t|s_t) &= 0.9 \\ P(o_t|\neg s_t) &= 0.2 \end{aligned}$$

$$\begin{aligned} p(S_1|o_0, \neg o_1) &\propto p(S_1, o_0, \neg o_1) \\ &= \sum_{S_0} p(S_0) p(o_0|S_0) p(S_1|S_0) p(\neg o_1|S_1) \end{aligned}$$

Operations: (3 mult * 2 + 1 add) * 2 = 14 ops

Filtering through Enumeration

I already know that the manager's behavior in these past K days, what's the chance of raining for today, e.g. $p(S_k|o_0, \cdots, o_k)$?

$$
\begin{aligned}
p(s_0) &= 0.5 \\
P(s_t|s_{t-1}) &= 0.7 \\
P(s_t|\neg s_{t-1}) &= 0.3 \\
P(o_t|s_t) &= 0.9 \\
P(o_t|\neg s_t) &= 0.2
\end{aligned}
$$

$$
\begin{aligned}
p(S_k|o_0, \cdots, o_k) &\propto p(S_k, o_0, \cdots, o_k) \\
&= \sum_{S_k} \cdots \sum_{S_0} p(S_0) p(o_0|S_0) \cdots p(o_k|S_K)
\end{aligned}
$$

Filtering through Enumeration

I already know that the manager's behavior in these past K days, what's the chance of raining for today, e.g. $p(S_k|o_0, \cdots, o_k)$?

$$\begin{aligned} p(s_0) &= 0.5 \\ P(s_t|s_{t-1}) &= 0.7 \\ P(s_t|\neg s_{t-1}) &= 0.3 \\ P(o_t|s_t) &= 0.9 \\ P(o_t|\neg s_t) &= 0.2 \end{aligned}$$

$$\begin{aligned} p(S_k|o_0, \cdots, o_k) &\propto p(S_k, o_0, \cdots, o_k) \\ &= \sum_{S_k} \cdots \sum_{S_0} p(S_0)p(o_0|S_0) \cdots p(o_k|S_K) \end{aligned}$$

Operations: $O(K \times 2^K)$ ops

Filtering through Forward Recursion

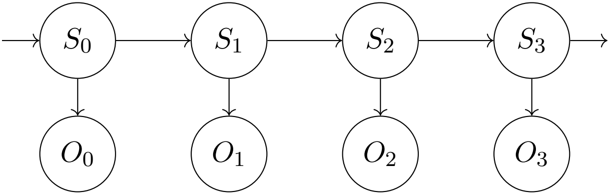
Let $f_{0:k} = P(S_k | o_{0:k})$.

Base case:

$$f_{0:0} = \alpha \, P(o_0 | S_0) P(S_0)$$

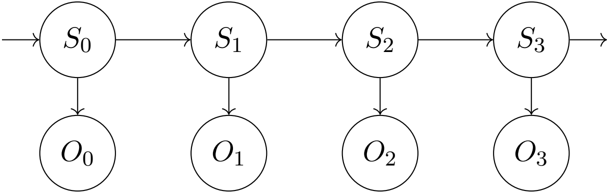
Recursive case:

$$f_{0:k} = \alpha \, P(o_k | S_k) \sum_{s_{k-1}} P(S_k | s_{k-1}) f_{0:(k-1)}$$



Filtering through Forward Recursion

$$f_{0:k} = \alpha\, P(o_k|S_k) \sum_{s_{k-1}} P(S_k|s_{k-1}) f_{0:(k-1)}$$



Operations:
$f_{0:0}$ has 2 mult ops,
$f_{0:1}$ has (2 mult + 1 add + 1 mult) * 2 ops
$f_{0:2}$ has (2 mult + 1 add + 1 mult) * 2 ops

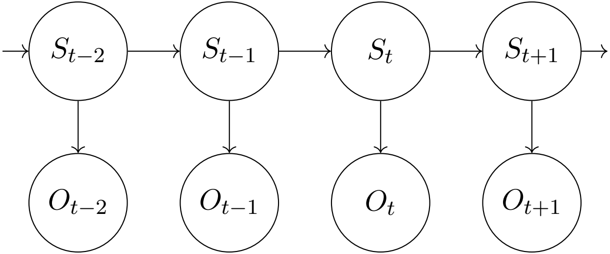
Total Operation: $O(k)$ ops

The Umbrella Story

$$P(s_0) = 0.5$$

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

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



A Filtering Example

Consider the umbrella story.

Assume that $O_0 = t$ and $O_1 = t$.

Let's calculate $f_{0:0}$ and $f_{0:1}$ using forward recursion.

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 Filtering Example - Base Case of Forward Recursion

Calculate $f_{0:0} = P(S_0|o_0)$.

A Filtering Example - Base Case of Forward Recursion

Calculate $f_{0:0} = P(S_0|o_0)$.

$$
\begin{aligned}
&\alpha = 1/p(o_0), constant \\
&P(s_0|o_0) = \alpha\, P(o_0|s_0)P(s_0) = \alpha\, 0.9 * 0.5 = \alpha\, 0.45 \\
&P(\neg s_0|o_0) = \alpha\, P(o_0|\neg s_0)P(\neg s_0) = \alpha\, 0.2 * 0.5 = \alpha\, 0.1 \\
&P(s_0|o_0) = 0.45/(0.45 + 0.1) = 0.818 \\
&P(\neg s_0|o_0) = 1 - 0.818 = 0.182
\end{aligned}
$$

A more compact approach:

$$
\begin{aligned}
P(S_0|o_0) &= \alpha\, P(o_0|S_0)P(S_0) \\
&= \alpha\, \langle 0.9, 0.2\rangle * \langle 0.5, 0.5\rangle \\
&= \alpha\, \langle 0.45, 0.1\rangle \\
&= \langle 0.818, 0.182\rangle
\end{aligned}
$$

A Filtering Example - Recursive Case of Forward Recursion

Calculate $f_{0:1} = \alpha\, P(o_1|S_1) \sum_{s_0} P(S_1|s_0) f_{0:0}$

where $f_{0:0} = \langle 0.818, 0.182 \rangle$.

A Filtering Example - Recursive Case of Forward Recursion

Calculate $f_{0:1} = \alpha\, P(o_1|S_1) \sum_{s_0} P(S_1|s_0) f_{0:0}$

where $f_{0:0} = \langle 0.818, 0.182 \rangle$.

First, let's expand the formula.

$$
\begin{aligned}
& P(S_1|o_{0:1}) \\
& = \alpha P(o_1|S_1) \sum_{s_0} P(S_1|s_0) P(s_0|o_0) \\
& = \alpha P(o_1|S_1) * \left(P(S_1|s_0) * P(s_0|o_0) + P(S_1|\neg s_0) * P(\neg s_0|o_0) \right) \\
& = \alpha \langle P(o_1|s_1), P(o_1|\neg s_1) \rangle \\
& \qquad * \Big(\langle P(s_1|s_0), P(\neg s_1|s_0) \rangle * P(s_0|o_0) \\
& \qquad\qquad + \langle P(s_1|\neg s_0), P(\neg s_1|\neg s_0) \rangle * P(\neg s_0|o_0) \Big)
\end{aligned}
$$

A Filtering Example - Recursive Case of Forward Recursion

Calculate $f_{0:1} = \alpha \, P(o_1|S_1) \sum_{s_0} P(S_1|s_0) f_{0:0}$

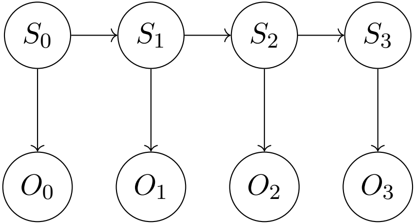
where $f_{0:0} = \langle 0.818, 0.182 \rangle$.

$$
\begin{aligned}
P(S_1|o_{0:1}) = \alpha \langle P(o_1|s_1), P(o_1|\neg s_1) \rangle \\
* \, (\langle P(s_1|s_0), P(\neg s_1|s_0) \rangle * P(s_0|o_0) \\
+ \langle P(s_1|\neg s_0), P(\neg s_1|\neg s_0) \rangle * P(\neg s_0|o_0))
\end{aligned}
$$

$$
\begin{aligned}
&= \alpha \langle 0.9, 0.2 \rangle \, (\langle 0.7, 0.3 \rangle * 0.818 + \langle 0.3, 0.7 \rangle * 0.182) \\
&= \alpha \langle 0.9, 0.2 \rangle \, (\langle 0.5726, 0.2454 \rangle + \langle 0.0546, 0.1274 \rangle) \\
&= \alpha \langle 0.9, 0.2 \rangle * \langle 0.6272, 0.3728 \rangle \\
&= \alpha \langle 0.56448, 0.07456 \rangle \\
&= \langle 0.883, 0.117 \rangle
\end{aligned}
$$

Example: Filtering

Consider a hidden Markov model with 4 time steps.



$$\boxed{P(s_0) = 0.4}$$

$$\boxed{\begin{array}{l} P(s_t|s_{t-1}) = 0.7 \\ P(s_t|\neg s_{t-1}) = 0.2 \end{array}}$$

$$\boxed{\begin{array}{l} P(o_t|s_t) = 0.9 \\ P(o_t|\neg s_t) = 0.2 \end{array}}$$

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

$\rightarrow$ i.e. $\alpha f_{0:2}$

Filtering (time k)

How did we derive the formula for $P(S_k|o_{0:k})$?

$$
\begin{aligned}
&P(S_k|o_{0:k}) \\
&= P(S_k|o_k \wedge o_{0:(k-1)}) && (1)\\
&= \alpha P(o_k|S_k \wedge o_{0:(k-1)})P(S_k|o_{0:(k-1)}) && (2)\\
&= \alpha P(o_k|S_k)P(S_k|o_{0:(k-1)}) && (3)\\
&= \alpha P(o_k|S_k)\sum_{s_{k-1}} P(S_k \wedge s_{k-1}|o_{0:(k-1)}) && (4)\\
&= \alpha P(o_k|S_k)\sum_{s_{k-1}} P(S_k|s_{k-1} \wedge o_{0:(k-1)})P(s_{k-1}|o_{0:(k-1)}) && (5)\\
&= \alpha P(o_k|S_k)\sum_{s_{k-1}} P(S_k|s_{k-1})P(s_{k-1}|o_{0:(k-1)}) && (6)
\end{aligned}
$$

Q: Filtering (time k)

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

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

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Q: Filtering (time k)

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

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

(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.

Filtering (time k)

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

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

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Filtering (time k)

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

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

(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 (A) Bayes' rule.

Filtering (time k)

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

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

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Filtering (time k)

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

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

(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 (D) The Markov assumption.

Filtering (time k)

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

$$
\begin{aligned}
&= \alpha P(o_k|S_k) P(S_k|o_{0:(k-1)}) \\
&= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k \wedge s_{k-1}|o_{0:(k-1)})
\end{aligned}
$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Filtering (time k)

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

$$= \alpha P(o_k|S_k) P(S_k|o_{0:(k-1)})$$
$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k \wedge s_{k-1}|o_{0:(k-1)})$$

(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.

Filtering (time k)

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

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k \wedge s_{k-1}|o_{0:(k-1)})$$

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k|s_{k-1} \wedge o_{0:(k-1)}) P(s_{k-1}|o_{0:(k-1)})$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Filtering (time k)

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

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k \wedge s_{k-1}|o_{0:(k-1)})$$

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k|s_{k-1} \wedge o_{0:(k-1)}) P(s_{k-1}|o_{0:(k-1)})$$

(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 (C) The chain/product rule.

Filtering (time k)

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

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k|s_{k-1} \wedge o_{0:(k-1)}) P(s_{k-1}|o_{0:(k-1)})$$

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k|s_{k-1}) P(s_{k-1}|o_{0:(k-1)})$$

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Filtering (time k)

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

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k|s_{k-1} \wedge o_{0:(k-1)}) P(s_{k-1}|o_{0:(k-1)})$$

$$= \alpha P(o_k|S_k) \sum_{s_{k-1}} P(S_k|s_{k-1}) P(s_{k-1}|o_{0:(k-1)})$$

(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 (D) The Markov assumption.

Revisiting Learning Goals

- Construct a hidden Markov model given a real-world scenario.
- Explain the independence assumptions in a hidden Markov model.
- Calculating the filtering probability for a time step in a hidden Markov model.
- Describe the justification for a step in the derivation of the filtering formulas.