Inference in Hidden Markov Models Part 1

Alice Gao

Lecture 14

Readings: RN 15.1, 15.2.1, 15.2.2. PM 8.5.1 - 8.5.3.

Outline

Learning Goals

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

- 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
- speech and handwriting recognition

The Umbrella Story

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

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 un-observable state at time t.
  Let O_t denote the signal/observation at time t.

What are the random variables in the umbrella world?

Transition Model

How does the current state depend on the previous states?

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

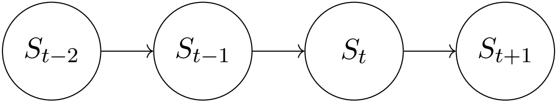
$$P(S_t | S_{t-1} \land S_{t-2} \land S_{t-3} \land \cdots \land S_0)$$

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

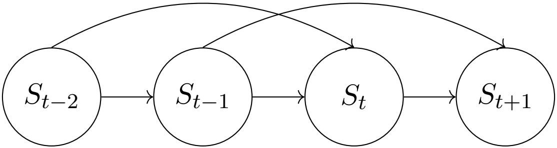
First-order Markov process:



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:



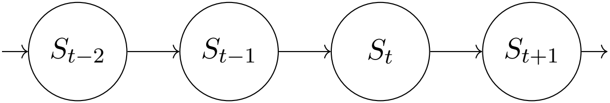
The transition model:

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

Transition model for the umbrella story

The Markov assumption:

The future is independent of the past given the present.



The transition model:

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

Stationary Process

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

Stationary process:

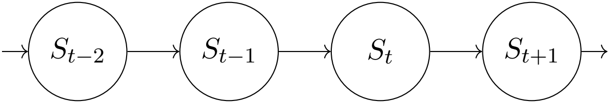
- The dynamics does not change over time.
- The conditional probability distribution for each time step remains the same.

What are the advantages of using a stationary model?

Transition model for the umbrella story

$$P(s_0) = 0.5$$

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



Sensor model

How does the evidence variable O_t at time t depend on the previous and current states?

(Sensor) Markov assumption:

Each state is sufficient to generate its observation.

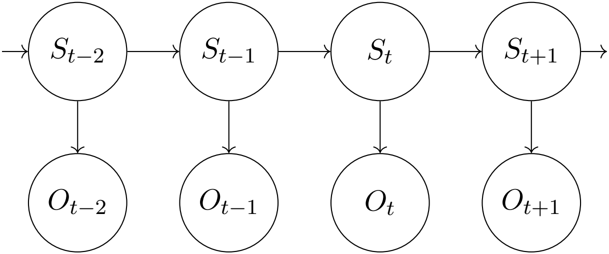
$$P(O_t | S_t \land S_{t-1} \land \cdots \land S_0 \land O_{t-1} \land O_{t-2} \land \cdots \land O_0) = P(O_t | S_t)$$

Complete model for the umbrella story

$$P(s_0) = 0.5$$

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

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



Hidden Markov Model

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

Common Inference Tasks

- **Filtering:** Which state am I in right now?
- **Prediction:** Which state will I be in tomorrow?
- **Smoothing:** Which state was I in yesterday?
- **Most likely explanation:** Which sequence of states is most likely to have generated the observations?

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 day 0 to day k,
what is the probability that I am in a particular state on day k?

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

Filtering through Forward Recursion

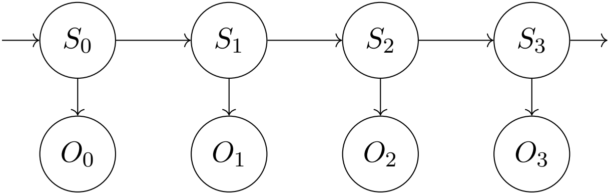
Let $f_{0:k} = P(S_k | o_{0:k})$.

Base case:

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

Recursive case:

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

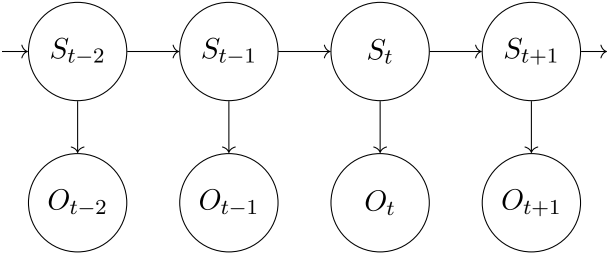


The Umbrella Story

$$P(s_0) = 0.5$$

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

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



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

$$P(s_0|o_0) = \alpha\, P(o_0|s_0)P(s_0) = \alpha\, 0.9 * 0.5 = \alpha\, 0.45$$
$$P(\neg s_0|o_0) = \alpha\, P(o_0|\neg s_0)P(\neg s_0) = \alpha\, 0.2 * 0.5 = \alpha\, 0.1$$
$$P(s_0|o_0) = 0.45/(0.45 + 0.1) = 0.818$$
$$P(\neg s_0|o_0) = 1 - 0.818 = 0.182$$

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

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.

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

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

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

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

Filtering (day k)

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

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

Filtering (day k) 1/6

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

Filtering (day k) 2/6

What is the justification for the step below?

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

(A) Bayes' rule

(B) Re-writing the expression

(C) The chain/product rule

(D) The Markov assumption

(E) The sum rule

Filtering (day k) 3/6

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

Filtering (day k) 4/6

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

Filtering (day k) 5/6

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 (day k) 6/6

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

Revisiting the Learning Goals

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

- 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.