Smoothing

Smoothing is computation of distribution of past states given current evidence, i.e. $P(X_k|e_{1:t})$, $1 \leq k < t$.

Formula for backward message:

$$P(e_{k+1:t}|X_k) = \sum_{x_{k+1}} P(e_{k+1}|x_{k+1})P(e_{k+2:t}|x_{k+1})P(x_{k+1}|X_k)$$

First term is sensor model; third term is transition model; second is 'recursive call'.

Define $b_{k+1:t} = \text{BACKWARD}(b_{k+2:t}, e_{k+1:t})$.

The backward phase has to be initialised with $b_{t+1:t} = P(e_{t+1:t}|X_t) = 1$ (a vector of 1s) because probability of observing empty sequence is 1.

As before, all this is quite abstract, back to our example.

Where are we?

Last time . . .

▶ Time in reasoning about uncertainty
▶ Markov assumption, stationarity
▶ Algorithms for reasoning about temporal processes
▶ Filtering and prediction

Today . . .

▶ Time and uncertainty II
In umbrella example:

Plugged into the above equation this yields:

$$P(R_1|u_1, u_2) = \alpha P(R_1|u_1)P(u_2|R_1)$$

- Forward filtering process yielded $\langle 0.818, 0.182 \rangle$ for first term
- The second term can be obtained through backward recursion:
  $$P(u_2|R_1) = \sum_{r_2} P(u_2|r_2)P(r_2|R_1)$$
  $$= (0.9 \times 1 \times (0.7, 0.3)) + (0.2 \times 1 \times (0.3, 0.7)) = (0.69, 0.41)$$
- Plugged into the above equation this yields:
  $$P(R_1|u_1, u_2) = \alpha \langle 0.818, 0.182 \rangle \times (0.69, 0.41) \approx (0.883, 0.117)$$
- So our confidence that it rained on Day 1 increases when we see the umbrella on the second day as well as the first.
- A simple improved version of this that stores results runs in linear time (forward-backward algorithm)

Finding the most likely sequence

- Suppose $[true, true, false, true, true]$ is the umbrella sequence for first five days, what is the most likely weather sequence that caused it?
- Could use smoothing procedure to find posterior distribution for weather at each step and then use most likely weather at each step to construct sequence
- NO! Smoothing considers distributions over individual time steps, but we must consider joint probabilities over all time steps
- Actual algorithm is based on viewing each sequence as path through a graph (nodes=states at each time step)

Finding the most likely sequence

- In umbrella example:

  ![Diagram](attachment:image.png)

  $$\begin{align*}
  P(R_1, R_2, R_3, R_4, R_5 | u_1, u_2) &= \alpha P(R_1 | u_1) P(u_2) P(R_2 | R_1) P(R_3 | R_2) P(R_4 | R_3) P(R_5 | R_4) \\
  &= \alpha \langle 0.818, 0.182 \rangle \times \langle 0.69, 0.41 \rangle \\
  &\approx \langle 0.883, 0.117 \rangle
  \end{align*}$$
Finding the most likely sequence

- There is a recursive relationship between most likely paths to $x_{t+1}$ and most likely paths to each state $x_t$

$$
\max_{x_1, \ldots, x_t} P(x_1, \ldots, x_t, x_{t+1} | e_{1:t+1})
= \alpha P(e_{t+1} | x_{t+1}) \max_{x_t} P(X_{t+1} | x_t) \max_{x_1, \ldots, x_{t-1}} P(x_1, \ldots, x_{t-1}, x_t | e_{1:t})
$$

- This is like filtering only that the forward message is replaced by

$$
m_{1:t} = \max_{x_1, \ldots, x_{t-1}} P(x_1, \ldots, x_{t-1}, x_t | e_{1:t})
$$

- And summation is now replaced by maximisation

Hidden Markov Models

- So far, we have seen a general model for temporal probabilistic reasoning (independent of transition/sensor models)

- In this and the following lecture we are going to look at more concrete models and applications

- Hidden Markov Models (HMMs): temporal probabilistic model in which state of the process is described by a single variable

- Like our umbrella example (single variable $Rain_t$)

- More than one variable can be accommodated, but only by combining them into a single “mega-variable”

- Structure of HMMs allows for a very simple and elegant matrix implementation of basic algorithms

Summary

- The forward-backward algorithm

- Finding the most likely sequence (Viterbi algorithm)

- Talked about HMMs

- HMMs: single state variable, simplifies algorithms (see other courses for these)

- Huge significance, for example in speech recognition:

$$P(\text{words} | \text{signal}) = \alpha P(\text{signal} | \text{words}) P(\text{words})$$

- Vast array of applications, but also limits.

- Next time: Dynamic Bayesian Networks