Oct 30 Markov Models nb

Announcements and To-Dos

Announcements:

1. Skip 1a for now but it's worth a bit of E.C. if you get A^* working. I'll add a few edges to hard code in an addendum. 1a will end up as extra credit *only*, but you do need a working distance function (probably) for the rest.

Last time we learned:

1. Stationary distributions to Markov Models.

Hidden Markov Models: Roadmap

An assumption like $P(E_t|X_{0:t}, E_{0:t-1}) = P(E_t|X_t)$ is powerful, because we'll get to use large probability products and conditional probability tables just like we did with Bayesian Networks. In general, we have a handful of tasks to do on networks like this:

1. **Filtering:** Describing the *process*.
2. **Prediction:** Describing the future: X_{t+1} *given* the past.
3. **Smoothing:** Describing the past: (or the chain X).
4. **Most likely explanation:** Describing the past: (or the chain X).
5. **Learning:** Bayesian updates and improvements on *priors* and *posteriors*.

HMM: Filtering

Filtering: The goal is to predict X_{t+1} *given* all the evidence available $E_{1:t+1}$.

$$\begin{aligned} P(X_{t+1}|E_{1:t+1}) &= \alpha P(E_{t+1}|X_{t+1}, E_{1:t}) P(X_{t+1}, E_{1:t}) \\ &= \alpha P(E_{t+1}|X_{t+1}, E_{1:t}) \sum_{X_t} P(X_{t+1}|E_{1:t} X_{1:t}) P(X_t|E_{1:t}) \end{aligned}$$

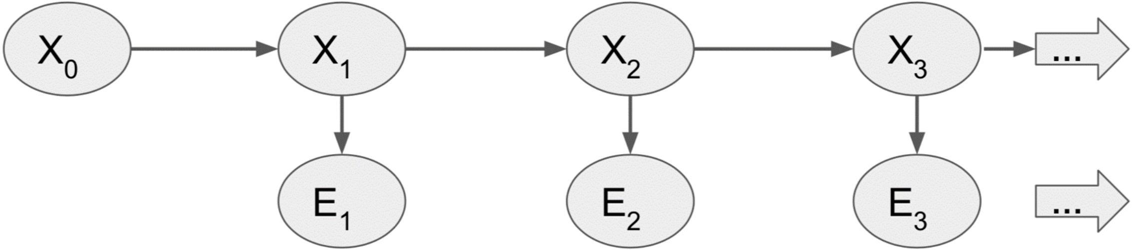
Sensor and Transition Models give independence!

$$\begin{aligned} &= \alpha P(E_{t+1}|X_{t+1}) \sum_{X_t} P(X_{t+1}|X_t) P(X_t|E_{1:t}) \\ &= \alpha \underbrace{P(E_{t+1}|X_{t+1})}_{Sensor} \sum_{X_t} \underbrace{P(X_{t+1}|X_t)}_{Transition} \underbrace{P(X_t|E_{1:t})}_{One\,prior\,time\,step} \end{aligned}$$

That last term is the same $X|E$ as the left-hand side, but for one *prior* time step. Sounds like a recursion or induction problem!

HMM: Filtering

Filtering: The goal is to predict X_{t+1} *given* all the evidence available $E_{1:t+1}$.
At $t = 0$:
$P(X_1|E_1) = P(E_1|X_1) \sum_{X_0} P(X_1|X_0)P(X_0)$
At $t = 1$:
$P(X_2|E_{1:2}) = P(E_2|X_2) \sum_{X_1} P(X_2|X_1)P(X_1|E_1)$
We continue FORWARD through the graph.



HMM: Prediction

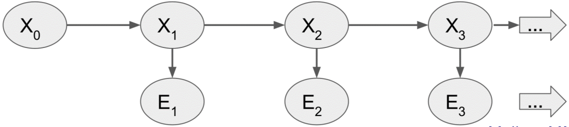
It turns out, this setup also allows us to *skip* steps and predict things in the future.

$$P(X_{t+1}|E_{1:t}) = \sum_{X_t} P(X_{t+1}|X_t)P(X_t|E_{1:t})$$

is a prediction *one* time step in the future. The same setup works if we skip ahead by k.

$$P(X_{t+k+1}|E_{1:t+1}) = \sum_{X_{t+k}} P(X_{t+k+1}|X_{t+k})P(X_{t+k}|E_{1:t})$$

This can also be done recursively, but now we have to sum over all possible outcomes of each unobserved step, just like in Bayes networks!



HMM: Prediction

So we want to predict the time in the future k...

$$P(X_{t+k+1}|E_{1:t+1}) = \sum_{X_{t+k}} P(X_{t+k+1}|X_{t+k})P(X_{t+k}|E_{1:t})$$

A one-step prediction is $k = 0$:

$$P(X_{t+1}|E_{1:t}) = \sum_{X_t} P(X_{t+1}|X_t)P(X_t|E_{1:t})$$

A two-step prediction is $k = 1$:

$$P(X_{t+2}|E_{1:t}) = \sum_{X_{t+1}} P(X_{t+2}|X_{t+1})P(X_{t+1}|E_{1:t})$$

but the last term *is* the one-step prediction!

$$P(X_{t+2}|E_{1:t}) = \sum_{X_{t+1}} P(X_{t+2}|X_{t+1}) \sum_{X_t} P(X_{t+1}|X_t)P(X_t|E_{1:t})$$

HMM: Prediction

The prediction of X two $(k = 1)$ time steps beyond where our evidence ended was:

$$P(X_{t+2}|E_{1:t}) = \sum_{X_{t+1}} P(X_{t+2}|X_{t+1}) \sum_{X_t} P(X_{t+1}|X_t) P(X_t|E_{1:t})$$

Making a k-step prediction just means doing FORWARD-steps up until we're out of evidence, and then following the Markov process to evolve $X_{t+1}|X_t$ until we reach the desired future time.