Last lecture:

- Markov models
- Hidden Markov Models

Markov model as Bayes Net:

$$\begin{tikzcd}
X_0 \arrow[r] \arrow[rrr, bend left] & \cdots \arrow[r] \arrow[r, bend left] & X_{t-2} \arrow[r] \arrow[rr, bend left] & X_{t-1} \arrow[r] \arrow[rr, bend left] & X_t \arrow[r] & \cdots
\end{tikzcd}$$

1st order: $P(X_t \mid X_{0:t-1}) = P(X_t \mid X_{t-1})$

2nd order $\sim \quad = P(X_t \mid X_{t-1}, X_{t-2})$

Joint probability of a <u>specific</u> state sequence:

$$P(X_0 = x_0, X_1 = x_1, \dots, X_t = x_t)$$

Just use rule for Bayes Nets:

$$P(X_{0:T}) = P(X_0) \prod_{t=1}^{T} P(X_t \mid pa(X_t))$$

$$pa(X_t) = \begin{cases} X_{t-1} & \text{1st order} \\ X_{t-1}, X_{t-2} & \text{2nd order} \\ \vdots & \end{cases}$$

How do we compute $P(X_t = s)$?

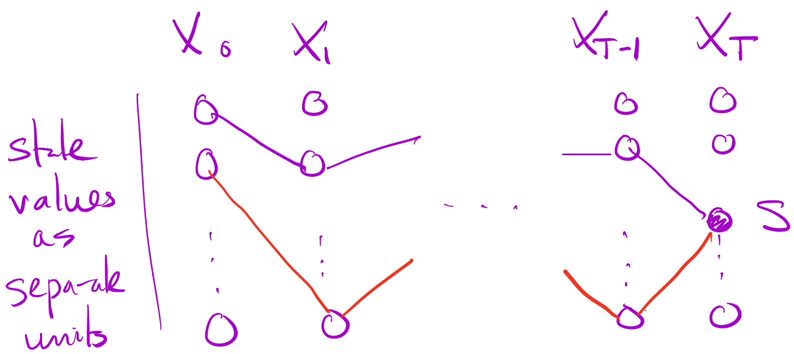
$=$ probability that $X$ is in state $s$ at time $t$.

Note this is not the same as before:

$P(X_{0:T})$

which is $P(X_0 = x_1, X_1 = x_2, \ldots, X_T = s)$

We need to consider all sequences that <u>could</u> have lead to $X_t = S$.




Brute force: We know $P(X_{0:T})$ so just add up all the probabilities of the paths that end at $X_T = S$, i.e. marginalize.

$$P(X_T = S) = \sum P(X_{1:T} = q)$$

$q \in$ set of state sequences that end at $s$ at time $T$.

But this is obviously inefficient.

$O(N^T)$ for $|S| = N$, i.e. $N$ states.

A better way: use induction to avoid redundant computation.

For each state, define $p_t(i) = P(X_t = S_i)$

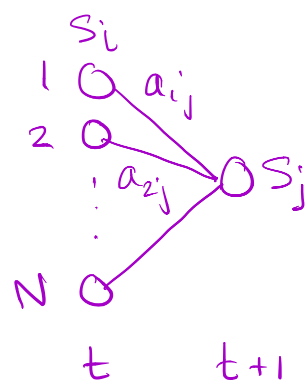
$= \text{prob state is } S_i \text{ at time } t.$

Induction:

$$\forall i \quad P_0(i) = \begin{cases} 1 & \text{if } S_i = \text{start state} \\ 0 & \text{otherwise} \end{cases}$$

$$\forall j \quad P_{t+1}(j) = P(X_{t+1} = S_j)$$

$$= \sum_{i=1}^{N} P(X_{t+1} = S_j \wedge X_t = S_i)$$




sum over all the ways we could have gotten to $S_i$ from last time step.

$$= \sum_{i=1}^{N} \underbrace{P(X_{t+1} = S_j \mid X_t = S_i)}_{\swarrow} P(X_t = S_i)$$

$$= \sum_{i=1} a_{ij} \, P_t(i)$$

$\hookrightarrow$ Markov transition probability.

$\Rightarrow$ Can now compute $P_t(i)$ for all states.

$O(N^2 T)$

# Hidden Markov Models (HMMs)

Eg:

$$\begin{tikzcd} \text{words} & X_0 \arrow[r] \arrow[d] & \cdots & X_{t-1} \arrow[r] \arrow[d] & X_t \arrow[r] \arrow[d] & \cdots & \text{hidden causal states} \\ \text{acoustic features} & E_0 & & E_{t-1} & E_t & \cdots & \text{observable evidence} \end{tikzcd}$$

Joint prob is same as Bayes Net:

$$P(X_{0:T}, E_{0:T})$$

$$= P(E_0) \prod_{i=1}^{t} P(E_i \mid X_i) P(X_i \mid X_{i-1})$$

$\boxed{?}$ How would you model spelling with HMMs?

HMM generalizations: DBNs, Dynamic Bayes Nets.

## Inference tasks

"filtering" — posterior prob. over current state given all evidence thus far

$P(X_t | E_{0:t})$

e.g.: prob $X_t$ = word i given observed features

prediction — posterior over future state

$P(X_{t+k} | E_{0:t})$, $k>0$

"smoothing" — $P(X_k | E_{0:t})$, $0 \le k < t$

Can future evidence affect beliefs about past states? Yes. This is a temporal version of explaining away.

"Given what we know now, we think that..."

most likely hidden sequence (Viterbi "alignment")

$$\arg\max_{X_{0:t}} P(X_{0:t} | E_{0:t})$$

Note: most likely hidden sequence is not sequence of most likely states.

Just look at one in detail.

Filtering: the "forward" algorithm

Like before: want recursive, efficient estimation of

$$\underbrace{P(X_{t+1} \mid E_{0:t+1})} = f\Big(\underbrace{E_{t+1}}_{\text{new evidence / observation}},\ \underbrace{P(X_t \mid E_{0:t})}_{\text{already computed (assumed by induction)}}\Big)$$

Note this is a posterior distribution

$$P(X_{t+1} \mid E_{0:t+1}) = P(X_{t+1} \mid E_{0:t}, E_{t+1})$$

$$= \alpha\, P(E_{t+1} \mid X_{t+1}, E_{0:t})\, P(X_{t+1} \mid E_{0:t}) \qquad \text{by Bayes' Rule}$$

$$= \underbrace{\alpha}_{\text{norm. const.}}\ \underbrace{P(E_{t+1} \mid X_{t+1})}_{\text{sensor model}}\ \underbrace{P(X_{t+1} \mid E_{0:t})}_{\text{one-step prediction}} \qquad \text{cond. indep.}$$

sensor model → we know this

one-step prediction → we need to derive this

Use marginalization again:

$$P(X_{t+1} \mid E_{0:t}) = \sum_{X_t} P(X_{t+1}, X_t \mid E_{0:t})$$

$$= \sum_{X_t} \underbrace{P(X_{t+1} \mid X_t, \underset{\text{cond indep.}}{\cancel{E_{0:t}}})}_{\text{transition model}}\ \underbrace{P(X_t \mid E_{0:t})}_{\text{current state model}}$$