# CS181: Bayes Nets - Part III: Hidden Markov Models

(Prob)

- Unsupervised Learning: goal model $p(x)$
- Specifically: structure in $p(x)$: $x$ has multiple dims that have interesting conditional independence relationships

$x$: $A \rightarrow B \rightarrow C$

- Noted: relationships can be used to perform inference more efficiently: $p(\text{query var} \mid \text{evidence})$

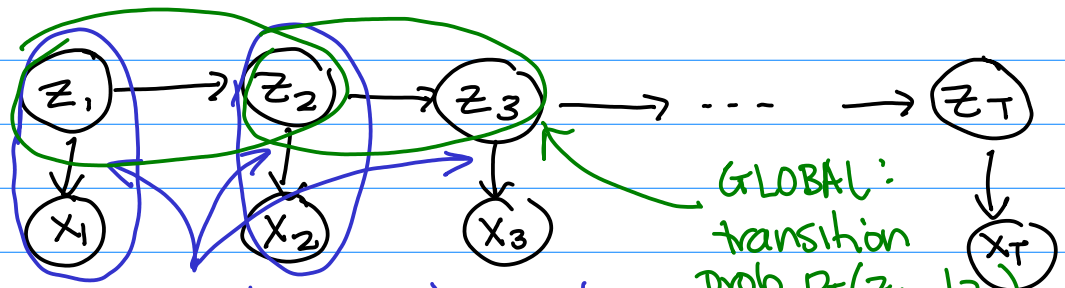
e.g. $p(A|C)$ or $p(B)$

TODAY: Specific type of BN: Hidden Markov Model
(Timeseries model)


(LOCAL)

hidden/ unobserved structure {

observed vector $\vec{x}_n$ {




GLOBAL: transition prob $p_T(z_{t+1}|z_t)$

GLOBAL emission prob $p_\Omega(x|z)$

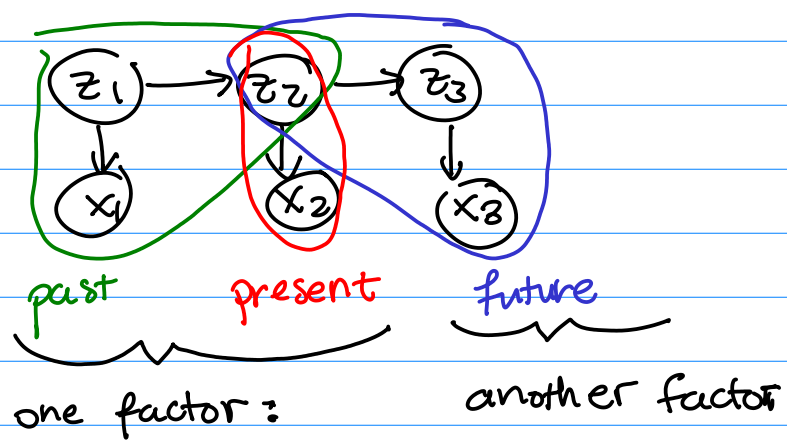
| | x = router1 | router2 | router 3 |
|---|---|---|---|
| z = hallway : | .7 | .2 | .1 |
| rec room : | .1 | .8 | .1 |
| diningroom : | .1 | .1 | .8 |

Why do we want this structure? What can we do with it?

1. Filtering: where am I now? $p(z_t \mid x_1 \cdots x_t)$
2. Smoothing: where was I at time t? $p(z_t \mid x_1 \cdots x_T)$
3. Prediction: what's my next meas? $p(x_{t+1} \mid x_1 \cdots x_t)$
4. prob(seq): is this a good model? $p(x_1 \cdots x_T)$
5. Best path: what path was most likely taken? $\underset{z_1 \cdots z_T}{\operatorname{argmax}}\, p(z_1 \cdots z_T \mid x_1 \cdots x_T)$

→ we've seen this type of problem! all have form $p(-\mid-)$ that requires various conditionals & marginals of the joint: $p(x_1 \cdots x_T, z_1 \cdots z_T)$

<u>intuition</u>:




<u>Formalizing</u>:

Lets consider $p(z_t, x_1 \cdots x_t)$

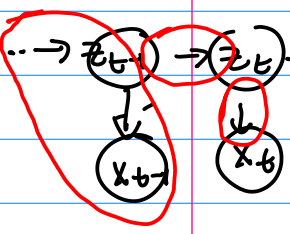
(aside: $p(z_t = k \mid x_1 \cdots x_t) \propto p(z_t = k, x_1 \cdots x_t)$

$$p(z_t = k \mid x_1 \cdots x_t) = \frac{p(z_t = k, x_1 \cdots x_t)}{\sum_j p(z_t = j, x_1 \cdots x_t)}$$

$p(x_1 \cdots x_t)$

$\alpha_t$ : $[K]$ $\alpha_t(z_t = k)$

$$\alpha_t(z_t) = p(z_t, x_1 \cdots x_t) = \sum_{z_{t-1}=1}^{K} p(z_{t-1}, z_t, x_1 \cdots x_t)$$



$$= \sum_{z_{t-1}=1}^{K} p(x_t \mid z_t)\, p(z_t \mid z_{t-1})\, p(z_{t-1}, x_1 \cdots x_{t-1})$$

$$= \underbrace{p(x_t \mid z_t)}_{\substack{K: \\ x_t \text{ obs,} \\ p(x_t \mid z_t = k)}} \sum_{z_{t-1}=1}^{K} \underbrace{p(z_t \mid z_{t-1})}_{\substack{K \times K \\ \text{transition} \\ \text{matrix}}} \underbrace{\alpha_{t-1}(z_{t-1})}_{K}$$

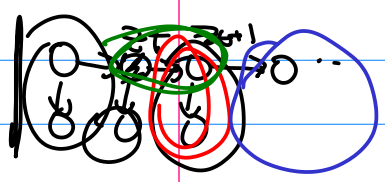
$$\alpha_t(z_t) = \sum_{z_1} \sum_{z_2} \cdots \sum_{z_{t-1}} p(\underbrace{z_1 \cdots z_{t-1}}_{\text{exponentially many combinations}}, z_t, x_1 \cdots x_t)$$

but given structure, efficient dynamic program

FORWARD PASS
(considers the past)

Next: BACKWARD PASS (future)

for smoothing problem: $p(z_t, x_1 \cdots x_T)$



$$= p(x_1 \cdots x_t, z_t)\, p(x_{t+1} \cdots x_T \mid z_t, x_1 \cdots x_t)$$ true in general

$$= \underbrace{p(x_1 \cdots x_t, z_t)}_{\substack{\text{FORWARD PASS} \\ \alpha \checkmark}} \underbrace{p(x_{t+1} \cdots x_T \mid z_t)}_{\substack{\text{BACKWARD PASS} \\ \beta}}$$ using cond. ind.

$$\beta_t(z_t) = p(x_{t+1} \cdots x_T \mid z_t)$$

vector size $K$

$$= \sum_{z_{t+1}=1}^{K} p(x_{t+1} \cdots x_T, z_{t+1} \mid z_t)$$ generally allowed, explicitly marginalizing

$$= \sum p(x_{t+1} \mid z_{t+1})\, p(x_{t+2} \cdots x_T \mid z_{t+1})\, p(z_{t+1} \mid z_t)$$ using struct.

$$= \sum_{z_{t+1}=1}^{K} p(x_{t+1} \mid z_{t+1})\, p(z_{t+1} \mid z_t)\, \beta_{t+1}(z_{t+1})$$

convention: $\beta_T(z_T) = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

## Return to our tasks:

1. filtering: $p(z_t \mid x_1 \ldots x_t) \propto \alpha_t(z_t)$
2. smoothing: $p(z_t \mid x_1 \ldots x_T) \propto \alpha_t(z_t)\, \beta_t(z_t)$
3. prob seq: $\underline{p(x_1 \ldots x_T)} = \sum_{z_T=1}^{K} \alpha_T(z_T)$

$p(z_1 \ldots z_T, x_1 \ldots x_T)$

$K^T$ !! $\{ \sum_{z_1} \sum_{z_2} \ldots \sum_{z_T} p(z_1 \ldots z_T, x_1 \ldots x_T)$

(recall: $\alpha_T(z_T) = p(z_T, x_1 \ldots x_T)$)

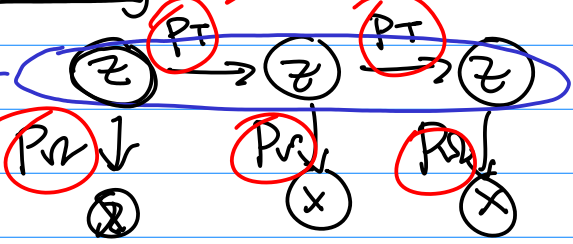
4. prediction: $p(x_{t+1} \mid x_1 \ldots x_t) = \dfrac{p(x_1 \ldots x_{t+1})}{p(x_1 \ldots x_t)} = \dfrac{\sum_{z_{t+1}} \alpha_{t+1}(z_{t+1})}{\sum_{z_t} \alpha_t(z_t)}$

## Note on Learning:

globals: M step

$p(z_{t+1} \mid z_t)$

locals: E-step




emission $p(x \mid z)$

if $p_T$, $p_\pi$ are known, we can compute

$\{ p(z_t \mid x_1 \ldots x_T)$
$p(z_t, z_{t+1} \mid x_1 \ldots x_T)$

if we have these, we can update $p_T$, $p_\pi$

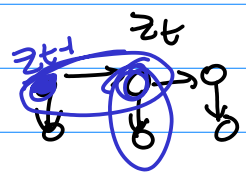
(recall: EM opt. lower bound on $p(x)$)

Note on finding the most likely path:

$$\underbrace{\delta_t(z_t)}_{\text{vec of length } K}^{=k} = \max_{z_1 \cdots z_{t-1}} p(x_1 \cdots x_t, z_1 \cdots z_{t-1}, z_t)$$

→ intuitively: ~~what~~ if we end up at $z_t = k$, what was the most likely path there?

$$\delta_t(z_t)^{=k} = \max_{z_{t-1}} \boxed{\delta_{t-1}(z_{t-1})} \, p(z_t \mid z_{t-1}) \, p(x_t \mid z_t)$$




could you get to $z_{t-1} = k'$ in a likely way

could you transition to $z_t$

would you have seen $x_t$?