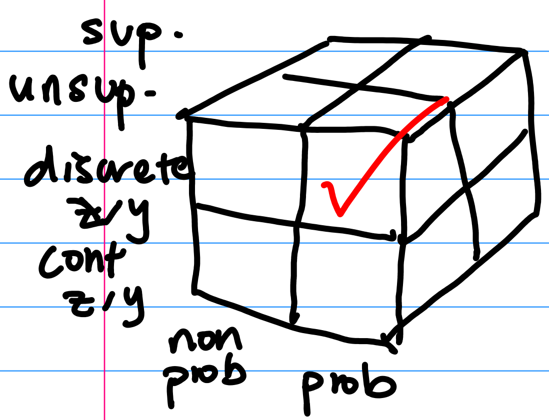
CS181 - Timeseries Models

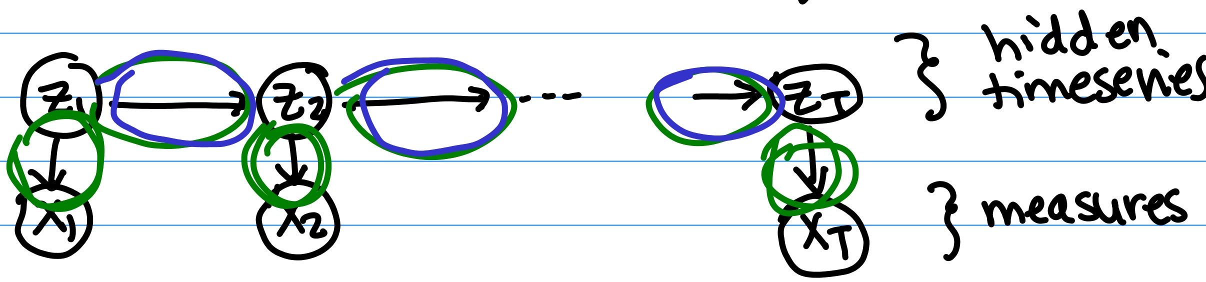


* Model Selection — NNs
* Expressive Models — BNs
* Objectives — Max Margin
  Decision Theory, RL

Notes: Practical due Friday

Last time(s): Bayesian Networks - defined, inference

Today: specific BN form: timeseries models
(Hidden Markov Models)



Three distributions we'll care about:

- $p_0(z)$: distribution of z_1 (where we start)
- $p_T(z'|z)$: distribution of next z given current z
- $p_{\Omega}(x|z)$: dist. of x given current z

Given this set-up, several questions we may ask:

- Filtering: $p(\underbrace{z_t}_{\text{query}} | \underbrace{x_1 \cdots x_t}_{\text{evidence}})$ real time prediction
- Smoothing: $p(z_t | x_1 \cdots x_T)$ afterwards inference
- p(seq) : $p(x_1 \cdots x_T)$ model selection
- best path : $\max_z p(z_1 \cdots z_T | x_1 \cdots x_T)$ best latents
- predict meas: $p(x_t | x_1 \cdots x_{t-1})$

How do we answer these common questions?

Full joint: $p(x_1 \cdots x_T, z_1 \cdots z_T) =$

$$p_0(z_1) \prod_{t=1}^{T-1} \underbrace{p_T(z_{t+1} \mid z_t)}_{\text{all the transitions}} \prod_{t=1}^{T} \underbrace{p_{\Omega}(x_t \mid z_t)}_{\text{all the emissions}}$$

Note: we can evaluate $p(z_1 \cdots z_T \mid x_1 \cdots x_T)$ up to constant factor by just looking at this joint.

BUT If we want $p(z_t \mid x_1 \cdots x_T)$, we need to marginalize out all the other z's!

Forward-Backward Algorithm:

Start w/ some basic blocks:

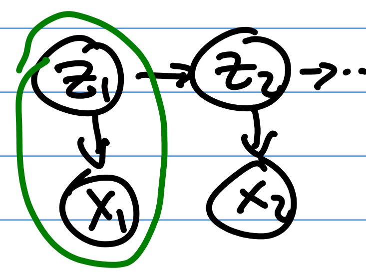
(1) $p_0(z_1) = \underbrace{[\;]}_{\text{vector of probs of each value of initial } z}$

(2) $p(\underline{x_1}) = \sum_{z_1} p_0(z_1)\, p_{\Omega}(x_1 \mid z_1)$

fixed ... fixed

vector w/ $\Pr(x_1 \mid z_1 = k)$

$= \sum_{z_1} [\;] \odot [\;]$



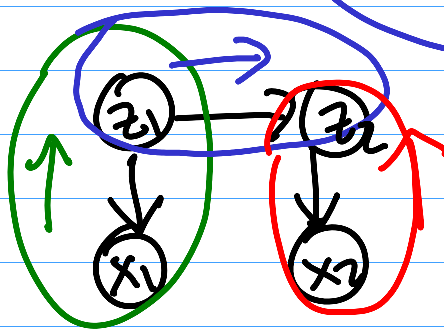
(3) $p(x_2 \mid x_1) \propto \sum_{z_1} \sum_{z_2} p_0(z_1)\, p_T(z_2 \mid z_1)\, p_{\Omega}(x_1 \mid z_1)\, p_{\Omega}(x_2 \mid z_2)$

$= \sum_{z_2} p_{\Omega}(x_2 \mid z_2) \sum_{z_1} p_0(z_1)\, p_T(z_2 \mid z_1)\, p_{\Omega}(x_1 \mid z_1)$

z_2, z_1

$$\sum_{z_2} p_{lv}(x_2|z_2) \sum_{z_1} p_T(z_2|z_1)\, \underbrace{p_0(z_1) \odot p_{lv}(x_1|z_1)}$$

adjust $p(z_2)$ given information about x_2: $p(x_1, x_2, z_2)$



What is the $p(z_2, x_1)$; haven't seen x_2

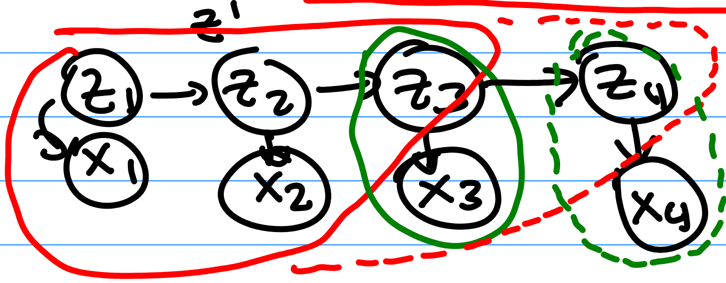
What is the prob of z_1, $p(z_1, x_1)$ where x_1 is fixed

Just before we summed out z_2, we had a vector that represented $p(z_2, x_1, x_2) \propto p(z_2 | x_1, x_2)$

(*) ↳ set up for a recursion!

$$\alpha_1 = p_0(z_1) \odot p_{lv}(x_1|z_1)$$

$$\alpha_t = \sum_{z'} \underline{\alpha_{t-1}(z')\, P_T(z|z')} \cdot \underline{P_{lv}(x_t|z)}$$



Forward pass: can be used to make predictions about next x, filter for current z:

$\alpha_t = p(z_t, x_1 \cdots x_t)$ so:

$$p(z_t | x_1 \cdots x_t) \propto p(z_t, x_1 \cdots x_t)$$

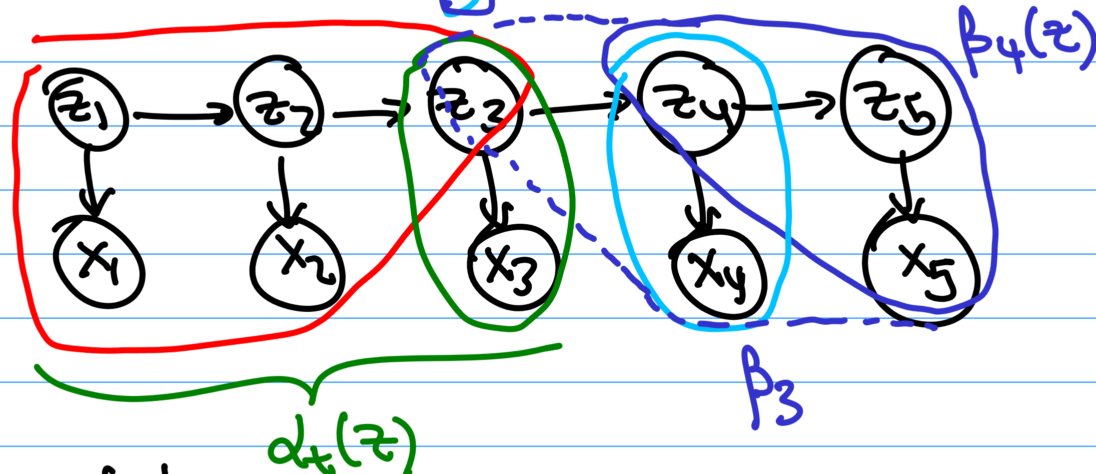
$$p(x_1 \cdots x_t) = \sum_{z_t} p(z_t, x_1 \cdots x_t) = \sum_{z_t} \alpha_t(z_t)$$

Why?? Note that
$p(x_1 \dots x_T, z_t) = \underbrace{p(x_1 \dots x_t, z_t)}_{\alpha} \cdot \underbrace{p(x_{t+1} \dots x_T \mid z_t)}_{\beta}$

Now, suppose I wanted to get
$p(x_{t+1} \dots x_T \mid z_t)$

$\beta_T(z) = 1$ (no info)

$$\beta_t(z) = \sum_{z'} \beta_{t+1}(z')\, p_T(z' \mid z)\, p_{\text{obs}}(x_{t+1} \mid z')$$



This is useful because
$$p(z_t \mid x_1 \dots x_T) \propto \underbrace{\alpha_t(z) \cdot \beta_t(z)}_{p(z_t, x_1 \dots x_T)}$$

✓ Using α & β: we can do filtering, smoothing, p(seq), predict x.

Remaining: $\max p(z_1 \dots z_T \mid x_1 \dots x_T)$
This needs a slightly different recursion...

Let $\delta_1(z_1) = p_0(z_1)\, p_{\text{obs}}(x_1 \mid z_1)$] start

$$\delta_2(z_2) = \underbrace{\left[\max_{z_1} \delta_1(z_1)\, p_T(z_2 \mid z_1)\right]}_{\text{best path to get to a specific } z_2} \underbrace{p_{\text{obs}}(x_2 \mid z_2)}_{\text{new obs at } t=2}$$

...

$$\delta_t(z_t) = \underbrace{\left[\max_{z_{t-1}} \delta_{t-1}(z_{t-1})\, p_T(z_t|z_{t-1})\right]}_{\text{best path to } z_t} p_{ob}(x_t|z_t)$$

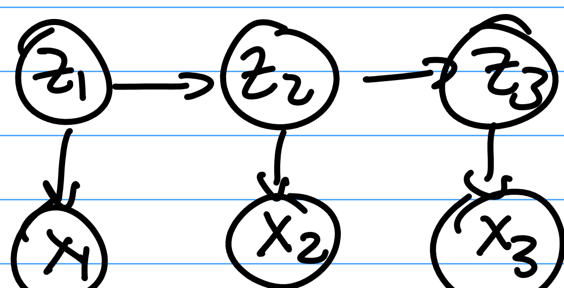
When you reach the end, choose value z_T^* based on highest $\delta_T(z_T)$, and then work backwards

$$z_t^* = \arg\max \{ \underbrace{\delta_t(z_t)}_{\text{ways to get to } z_t} p_T(z^*_{t+1}|z_t) \}$$

note that this is already fixed as we work backwards

where you'd need to be to get to z_{t+1}^*

Last Note on Learning:



} hiddens

global: p_0, p_T, p_{ob}

(*) block coordinate ascent: given z, solving for MLE p_0, p_T, p_{ob} is easy!!, given p_0, p_T, p_{ob}, finding max z or $p(z)$ is also do-able...