4270 Lec19

HMM Graphical Model

$P(y_1, X(1)=j)$
$= P(X_1=j)\, P(y_1 \mid X_1=j)$

Defn: Forward message

$$\alpha_t(j) = P(y_1, \dots, y_t, X(t)=j) \qquad \underline{\alpha}_t = \begin{pmatrix} \alpha_t(1) \\ \vdots \\ \alpha_t(k) \end{pmatrix} \in \mathbb{R}^k$$

Claim: $\underline{\alpha}_t^T = \underline{\alpha}_{t-1}^T P D_{y_t}$ (recursively!)

Pf:

$$\alpha_t(j) = P(y_1, \dots, y_t, X(t)=j)$$

$$= \sum_i P(y_1, \dots, y_{t-1}, \underbrace{y_t, X(t-1)=i, X(t)=j}_{A})$$

$$= \sum_i P(y_1, \dots, y_{t-1}, X(t-1)=i)\, P(A \mid y_1, \dots, y_{t-1}, X(t-1)=i)$$

$$= \sum_i P(y_1, \dots, y_{t-1}, X(t-1)=i)\, P(y_t, X(t)=j \mid X(t-1)=i)$$

$$= \sum_i P(y_1, \dots, y_{t-1}, X(t-1)=i)\, P(y_t \mid X(t)=j) \cdot P(X(t)=i \mid X(t-1)=j)$$

$$= \sum_i \alpha_{t-1}(i)\, P(y_t \mid j)\, P_{ij}$$

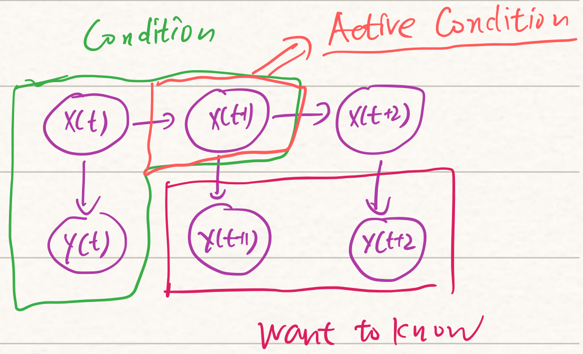
$$\Leftrightarrow \quad \alpha_t^T = \alpha_{t-1}^T P D_{y_t} \qquad \text{Forward Message}$$

Defn: Backward message

$$\beta_t(i) = P(y_{t+1}, \dots, y_n \mid X(t)=i) \qquad 1 \le t \le n-1$$

$$\left\{ \beta_t(i) = 1 \quad \forall i \quad \text{for } t=n \right.$$

Claim: $\beta_t = P D_{y_{t+1}} \beta_{t+1}$



Pf: $\beta_t(i) = P(y_{t+1}, \ldots, y_n \mid X(t)=i)$

$$= \sum_j P(y_{t+1}, \ldots, y_n, X(t+1)=j \mid X(t)=i)$$

$$= \sum_j P(y_{t+2} \ldots y_n \mid X(t+1)=j, \cancel{y_{t+1}}, \cancel{X(t)=i})\, P(y_{t+1}, X(t+1)=j \mid X(t)=i)$$

$\downarrow$ Markov Property!

$$= \sum_j P(y_{t+2}, \ldots, y_n \mid X(t+1)=j)\, P(y_{t+1} \mid X(t+1)=j)\, P(X(t+1)=j \mid X(t)=i)$$

$$= \sum_j \beta_{t+1}(j)\, P_{ij}\, P(y_{t+1} \mid X(t+1)=j)$$ $\Rightarrow$ Recursive Form!

In vector, $\beta_t^T = \beta_{t+1}^T D_{y_{t+1}} P^T$

$$\beta_t = P D_{y_{t+1}} \beta_{t+1}$$

Recall: $\alpha_t(i) = P(y_1, \ldots, y_t, X(t)=i)$

$\beta_t(i) = P(y_{t+1}, \ldots, y_n \mid X(t)=i)$

$$\Rightarrow P(y_1 \ldots y_n) = \sum_i P(y_1 \ldots y_n, X(t)=i)$$

$$= \sum_i \underbrace{P(y_1, \ldots, y_t, X(t)=i)}_{A}\, P(y_1, \ldots, y_n \mid A)$$

$$= \sum_i P(y_1 \ldots, y_t, X(t)=i)\, P(y_{t+1}, \ldots, y_n \mid X(t)=i)$$

$$= \underline{\alpha}_t^T \underline{\beta}_t \qquad \forall t = 1, 2, \dots, n$$

Note: If $t = n$, $\beta_t(i) = 1$ for $\forall i \in S$

Then we have:

$$P(y_1, \dots, y_n) = \underline{\alpha}_n^T \mathbb{1}$$

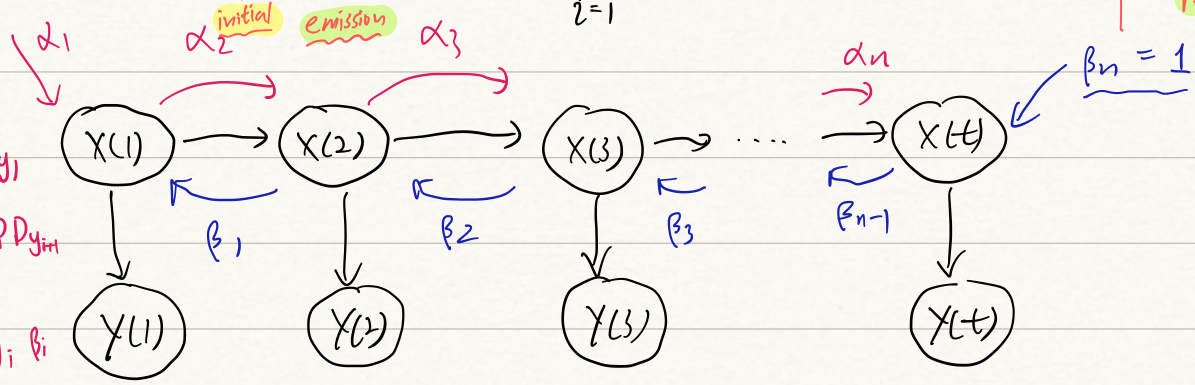
$\alpha_1(j) = P(y_1, X(1) = j) = \underbrace{q(j)}_{\text{initial}} \cdot \underbrace{P(y_1|j)}_{\text{emission}}$

$$= \sum_{i=1}^{K} \alpha_n(i)$$

set $\beta_n = 1$ to make sense the recursion of β

$$\begin{cases} \alpha_1^T = q^T D_{y_1} \\ \alpha_{i+1}^T = \alpha_i^T P D_{y_{i+1}} \end{cases}$$

$$\begin{cases} \beta_n = \mathbb{1} \\ \beta_{i-1} = P D_{y_i} \beta_i \end{cases}$$



Problem 2: Find the most likely hidden state sequence

$$(x_1^*, \dots, x_n^*) = \underset{x_1, \dots, x_n}{\arg\max} P(\underbrace{y_1, \dots, y_n}_{\text{observed}}, x_1, \dots, x_n) \quad \text{---} (*)$$

$|(x_1, \dots, x_n)| = k^n$

Similarly to Prob 1, (naive) exhaustive search require $O(k^n)$ operations.

$\alpha_1^T = q^T D_{y_1}$

Recall forward alg. $\alpha_1(j) = q(j) P(y_1|j) = Pr(y_1, x(1) = j)$

$$\downarrow$$

$$\alpha_t(j) = \left(\sum_i \alpha_{t-1}(i) P_{ij}\right) P(y_t|j)$$

Motivation : $\underbrace{a(b+c)}_{2} = \underbrace{ab+ac}_{3}$ steps

To solve $(*)$, change the sum to max !

Define Forward Message:

$$d_t(j) = \max_{x_1,\dots,x_{t-1}} P(y_1,\dots,y_t, \boxed{x_1,\dots,x_{t-1}}, X(t)=j)$$

$$= \sum_{x_1,\dots,x_{t-1}} P(y_1,\dots,y_t, x_1,\dots,x_{t-1}, X(t)=j)$$

Recall: $\alpha_t(j) = P(y_1,\dots,y_t, X(t)=j)$

Initialize: $d_1(j) = \alpha_1(j) = q(j)\, P(y_1 \mid j) = P(y_1, X(1)=j)$

Claim: $d_t(j) = \left(\max_i d_{t-1}(i)\, P_{ij}\right) P(y_t \mid j)$

$\alpha_t(j) = \left(\sum_i \alpha_{t-1}(i)\, P_{ij}\right) P(y_t \mid j)$

Pf: $d_t(j) = \max_{x_1,\dots,x_{t-1}} P(y_1,\dots,y_t, x_1,\dots,x_{t-1}, X(t)=j)$

cut

$$= \max_{x_1 \dots x_{t-2}, i} P(y_1,\dots,y_{t-1}, y_t, x_1,\dots,x_{t-2}, \boxed{X(t-1)=i}, X(t)=j)$$

$$= \max_{x_1,\dots,x_{t-2}, i} P(y_1,\dots,y_{t-1}, x_1,\dots,x_{t-2}, X(t-1)=i) \cdot P(y_t, X(t)=j \mid X(t-1)=i)$$

$$= \max_{x_1,\dots,x_{t-2}, i} P(y_1,\dots,y_{t-1}, x_1,\dots,x_{t-2}, X(t-1)=i)\; P(y_t \mid X(t)=j)\; P(X(t)=j \mid X(t-1)=i)$$

$$= \max_i \left\{ \left[\max_{x_1,\dots,x_{t-2}} P(y_1,\dots,y_{t-1}, x_1,\dots,x_{t-2}, X(t-1)=i) \right] \cdot P(y_t \mid X(t)=j)\; P_{ij} \right\}$$

$$= \max_i \left[d_{t-1}(i) P_{ij} \right] \cdot P(y_t \mid x(t) = j)$$

#

Evaluate the max prob. $\max_{x_1,\dots,x_n} P(y_1,\dots,y_n, x_1,\dots,x_n)$

$$= \max_j d_n(j)$$

But we still don't have argmax ! (Back track)

$\downarrow$

Now we have the maximum val. of $P(y_1,\dots,y_n, x_1,\dots,x_n)$ over all $x_1,\dots,x_n$ but not yet $\underbrace{x_1^*,\dots,x_n^*}_{\text{the argmax term}}$

$\downarrow$

Backtracking $\quad x_n^* = \arg\max_j d_n(j)$

$$\boxed{1 \le t \le n-1 : \quad x_t^* = \arg\max_i d_t(i) P_{i, x_{t+1}^*}}$$

Why this makes sense?

$\Downarrow$

suppose we find $x_n^* = \arg\max d_n(j)$

$\downarrow$

we want a 'certificate'

s.t $d_{n-1}(i) P_{ij} P(y_n \mid j) = d_n(x_n^*)$

$\downarrow$

$d_n(x_n^*) = \max_i (d_{n-1}(i) P_{i,x_n^*}) P(y_n \mid x_n^*)$

$\Updownarrow$

$x_{n-1}^* = \arg\max_i d_{n-1}(i) P_{i,x_n^*}$

Motivation: we already know x_n^* first, then we just check $d_{n-1}(i) P_{i,x_n^*}$ to achieve x_{n-1}^*

$\downarrow$

$\vdots$

$\downarrow$

x_1^*

$\boxed{\text{Viterbi Algorithm}}$ $\rightarrow$ $(x_1^*,\dots,x_n^*) = \arg\max_{x_1,\dots,x_n} P(y_1,\dots,y_n, x_1,\dots,x_n)$

① $d_t(j) = \max P(y_1, \dots, y_t, x_1, \dots, x_{t-1}, X(t)=j)$

$\hookrightarrow d_1(j) = q(j)\, P(y_1 | j)$

$\downarrow$ recursive

$d_n(j) = \max_i \{d_{n-1}(i) P_{i,j}\} P(y_t | j)$

② $\max P(y_1, \dots, y_n, x_1, \dots, x_n)$

$= \max_j d_n(j)$

③ $x_n^* = \arg\max_j d_n(j)$

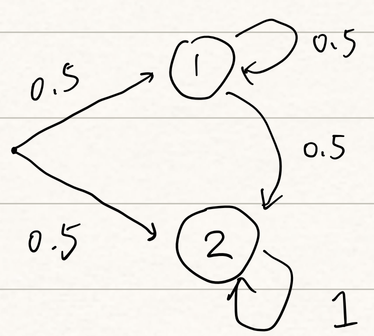
$x_{n-1}^* = \arg\max_i d_{n-1}(i) P_{i, x_n^*} \underbrace{\cancel{P(y_t | x_n^*)}}_{\text{const}}$

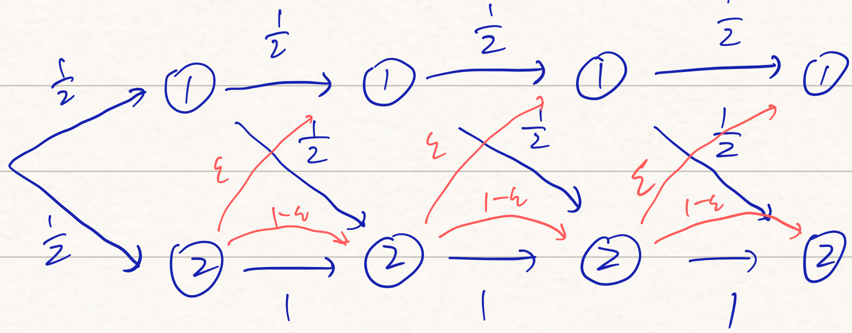
$\vdots$

x_1^*

Example [Viterbi]

Underlying hidden state sequence $\{X(t)\}_{t=1}^{\infty}$ has the following state transition diagram. $\cap \{1,2\}$





$$P(y \mid j) = \mathcal{N}(y; \mu_j, \sigma^2) \qquad \mu_j = \begin{cases} 1, & j=2 \\ 3, & j=1 \end{cases}$$

① σ^2 is very large : we don't care the observation

↓

Gaussian flat

⇓

only care about the hidden MC !

⇓

$(2,2,2,2,2,2,2,2)$

⇓

$$\arg\max P(y_1, y_2, \dots, y_8, X_1, \dots, X_8)$$

$$\approx \arg\max P(X_1, \dots, X_8)$$

$$\Rightarrow (X_1^*, \dots, X_8^*) = (2, \dots, 2)$$

② σ^2 small $\quad (1,1,1,1,1,2,2,2) \rightsquigarrow$ one bias

VS

$(1,1,2,2,2,2,2)$ X

two bias