Markov Model

↳ State로 이루어진 sequence를 상태전이 확률 행렬로 표현

ex) M T W T F S S

Sequence 비 → 해 → 해 → 해 → 비 → 비 → 해

State = {비, 해}

state transition probability matrix

$$\text{From}\quad \begin{array}{c|cc|c} & \multicolumn{2}{c|}{\text{To}} & \\ & 비 & 해 & \\ \hline 비 & 1 & 2 & /3 \\ 해 & 1 & 2 & /3 \end{array}$$

rowsum으로 나누어 P 구하기

$$\begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \qquad P(비 \to 비) = \frac{1}{3}$$

가정)

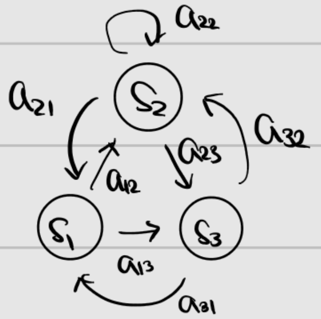
시간 t 에서의 관측은 가장 최근 r개의 관측만 의존

ex) r=1 이라면 $P(S_t | S_{t-1} \cdots S_1) = P(S_t | S_{t-1})$

$$S_1 \to S_2 \cdots \to \boxed{S_{t-1} \to S_t}$$

상태가 우리는 {비, 해} 로 2개. state → k개 k×k matrix

	S_1	S_2	S_3
S_1	a_{11}	a_{12}	a_{13}
S_2	a_{21}	a_{22}	a_{23}
S_3	a_{21}	a_{32}	a_{33}



Hidden Markov Model

	M	T	W	T	F	S	S	
Sequence 1 (Hidden)	비 →	해 →	해 →	해 →	비 →	비 →	해	$s_1 s_2 \cdots s_t$
	↓	↓	↓	↓	↓	↓	↓	
Sequence 2 (observable)	쇼핑	산책	산책	쇼핑	연구	연구	연구	$o_1 o_2 \cdots o_t$

S : {비, 해} O : {쇼핑, 산책, 연구}

A: state transition probability matrix

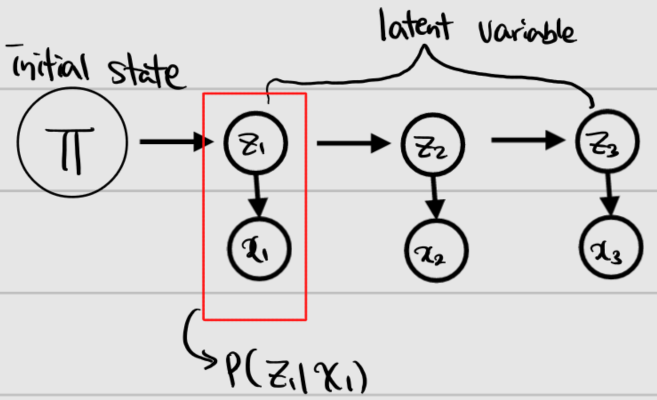
B: emission probability matrix

$$A = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \end{array}$$

$$B = \begin{array}{c} \\ o_1 \\ o_2 \\ o_3 \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \end{array}$$

π : initial state probability matrix

$$\pi = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array} \begin{array}{c} s_{t=0} \\ \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} \end{array}$$



Observation X

- Discrete or Continuous
- $X_1\ X_2 \cdots X_T$ time $1 \sim T$
- $x_i \in \{c_1\ c_2 \cdots c_m\}$

Latent state Z

- vector variable with K elements
  ex) GMM에서 K개 cluster
- Continuous : Kalman filter

<Initial>

↱ Multinomial distribution

$$P(Z_1) \sim Mult(\pi_1, \pi_2 \cdots \pi_K)$$

<Transition>

$z_1 \rightarrow z_2 \rightarrow z_3$

$$P(z_t \mid z_{t-1}^{(i)}=1) \sim Mult(a_{i1}\ a_{i2} \cdots a_{iK})$$

$$P(z^{(j)}_t \mid z^{(i)}_{t-1}) = a_{ij}$$

<Emission>

$$P(x_t \mid z_t^{(i)}=1) \sim Mult(b_{i1}, b_{i2} \cdots b_{im})$$

$$P(x_t^{(j)}=1 \mid z_t^{(i)}=1) = b_{ij}$$

Main Q

<Evaluation>

1. Given π, a, b, X

$\rightarrow P(X \mid M, \pi, a, b)$

<Decoding>

2. Given π, a, b, X

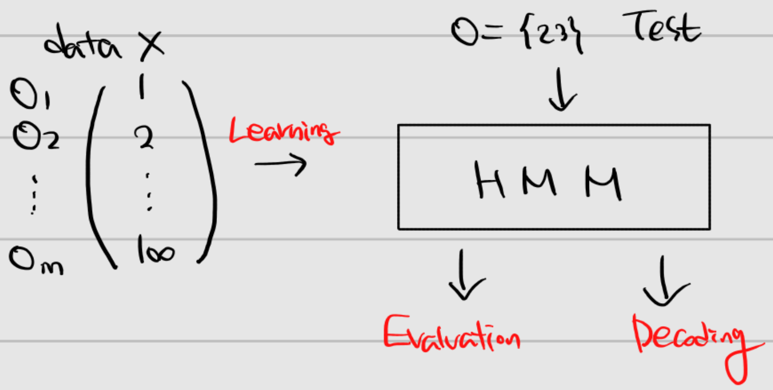
$\rightarrow \mathrm{argmax}_Z P(Z \mid X, M, a, b)$

latent의 most probable

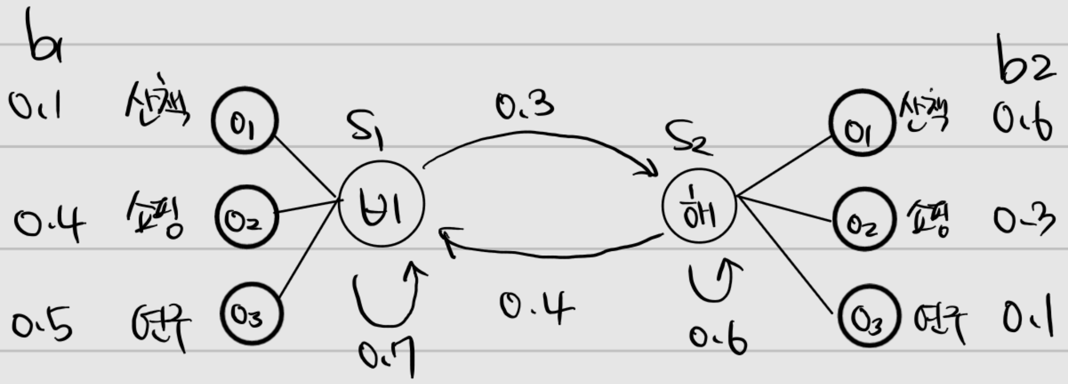
<Learning>

3. Given X

$\mathrm{argmax}_{\pi, a, b} P(X \mid M, \pi, a, b)$



① Evaluation



(초기 비 → $\pi_1 = 0.6$
해 → $\pi_2 = 0.4$)

P(산책, 산책, 연구, 쇼핑) = ?

<Forward>

		T=1	T=2	T=3	T=4
State	S_1 비	$\alpha_1(1)$	$\alpha_2(1)$	$\alpha_3(1)$	$\alpha_4(1)$
	S_2 해	$\alpha_1(2)$	$\alpha_2(2)$	$\alpha_3(2)$	$\alpha_4(2)$
Observation		O_1 = 산책	O_2 = 산책	O_3 = 연구	O_4 = 쇼핑

a_{11}, a_{12}, a_{21}, a_{22}

$$\alpha_1(1) = \pi_1 \cdot b_1(\text{산책}) = 0.6 \times 0.1 = 0.06$$
$$\alpha_1(2) = \pi_2 \cdot b_2(\text{산책}) = 0.4 \times 0.6 = 0.24$$

$$\alpha_2(1) = (\alpha_1(1) \cdot a_{11} + \alpha_1(2) \cdot a_{21}) \cdot b_1(\text{산책}) = 0.0138$$
$$\alpha_2(2) = (\alpha_1(1) \cdot a_{12} + \alpha_1(2) \cdot a_{22}) \cdot b_2(\text{산책}) = 0.0972$$

⋮

P(산, 산, 연, 쇼) = 0.0111

<Backward>

State	S_1 비	$\beta_1(1)$	$\beta_2(1)$	$\beta_3(1)$	1	1 → End
	S_2 해	$\beta_1(2)$	$\beta_2(2)$	$\beta_3(2)$	1	1 → End
	O	산	산	연	쇼	

a_{11}, a_{12}, a_{21}, a_{22}

← 방향

$$\beta_3(1) = \beta_4(1) \cdot a_{11} \cdot b_1(\text{쇼핑}) + \beta_4(2) \cdot a_{12} \cdot b_2(\text{쇼핑}) = 0.37$$

$$\beta_3(2) = \beta_4(1) \cdot a_{21} \cdot b_1(\text{쇼핑}) + \beta_4(2) \cdot a_{22} \cdot b_2(\text{쇼핑}) = 0.34$$

$\beta_3 \rightarrow \beta_4$만 사용, $\beta_2 \rightarrow \beta_3$만사용 ... 바로 다음것만 사용

P(산, 산, 연, 쇼)

$= \beta_1(1) \cdot \pi_1 \cdot b_1(\text{산}) + \beta_1(2) \cdot \pi_1 \cdot b_2(\text{산}) = 0.111$

∴ forward, backward 값은 같음

Evaluation → forward, backward 아무거나 OK

② Decoding

$O = (o_1 = \text{산} \quad o_2 = \text{산} \quad o_3 = \text{연} \quad o_4 = \text{쇼})$

$H = (\ ? \quad ? \quad ? \quad ? \ )$

Hidden state를 예측

		T=1	T=2	T=3	T=4
State	S_1 비	$V_1(1)$	$V_2(1) \xrightarrow{a_{11}}$	$V_3(1)$	$V_4(1)$
	S_2 해	$V_1(2)$	$V_2(2) \xrightarrow{a_{22}}$	$V_3(2)$	$V_4(2)$
Observation		O_1= 산책	O_2= 산책	O_3= 연구	O_4= 쇼핑

Viterbi 사용 $V_1(1) = \pi_1 \cdot b_1(\text{산}) = 0.6 \times 0.1 = 0.06$

$V_1(2) = \pi_2 \cdot b_2(\text{산}) = 0.4 \times 0.6 = 0.24$

max중 큰값의 위치

1 or 2

→ V_1에서 오는것, V_2에서 오는것의 max

$\tau_2(1) = 2$

$V_2(1) = \max(V_1(1) \cdot a_{11}, V_1(2) \cdot a_{21}) \cdot b_1(\text{산}) = \max(0.04, 0.09) \cdot 0.1 = 0.009$

$V_2(2) = \max(V_1(1) \cdot a_{12}, V_1(2) \cdot a_{22}) \cdot b_2(\text{산}) = \max(0.01, 0.14) \cdot 0.6 = 0.0864$

$\tau_2(2) = 2$

S_1 비 $V_1(1)=0.06$ $V_2(1)=0.0096$ $V_3(1)=0.01728$ $V_4(1)=0.00484$

S_2 해 $V_2(1)=0.24$ $V_2(2)=0.0864$ $V_3(1)=0.00518$ $V_4(2)=0.00156$

해 해 비 비

↳ V의 max로 결정 Hidden= (해, 해, 비, 비) 로 예측

forward → 합, viberti → max

③ Learning

Input : HMM(λ) Architecture

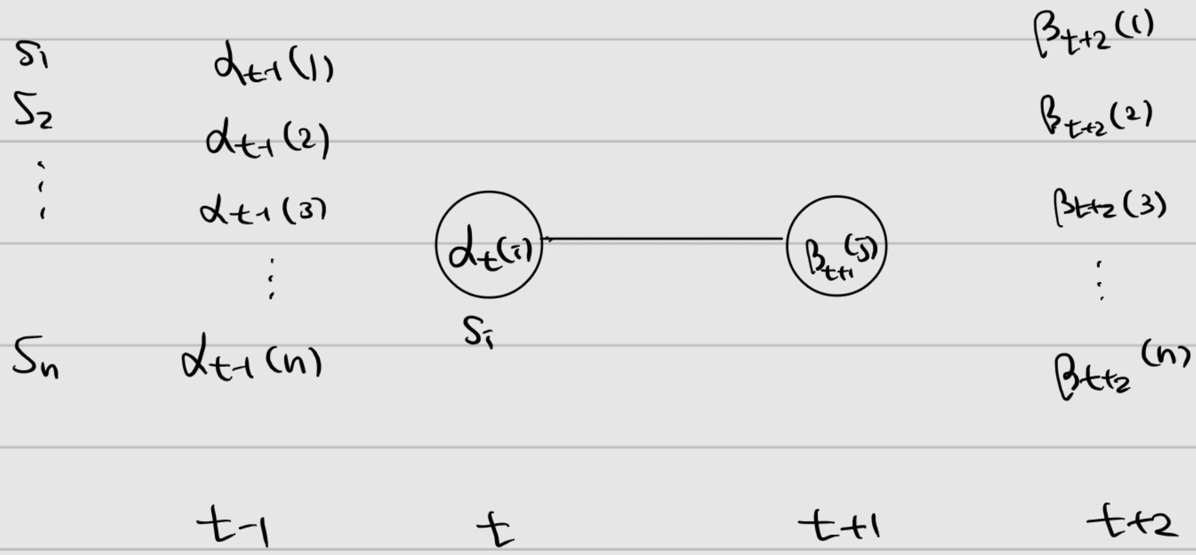
output : HMM(λ^*) = $\{A^*, B^*, \pi^*\}$

$\gamma_t(i)$: HMM(λ), O 주어졌을 때, t 시점 State= S_i 일 확률

$\xi_t(i,j)$ 〃 S_i, t+1 시점 S_j 일 확률

S_1	$\alpha_{t-1}(1)$		$\beta_{t+1}(1)$
S_2	$\alpha_{t-1}(2)$	$\alpha_t(i)$ $\beta_t(i)$	$\beta_{t+1}(2)$
$\vdots$	$\alpha_{t-1}(3)$	S_i	$\beta_{t+1}(3)$
	$\vdots$		$\vdots$
S_n	$\alpha_{t-1}(n)$ forward →	← backward	$\beta_{t+1}(n)$

$$\gamma_t(i) = P(q_t = S_i \mid O, \lambda) = \frac{\alpha_t(i)\,\beta_t(i)}{\sum_{j=1}^{n} \alpha_t(j)\,\beta_t(j)} = \frac{S_i \text{ 일 확률}}{S_1 S_2 \cdots S_n \text{ 일 확률}}$$



$$\xi_t(i,j) = \frac{\alpha_t(i)\, \overset{\text{i,j 상태연결}}{\underline{a_{ij}\, b_j (O_{t+1})}}\, \beta_{t+1}(j)}{\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_t(i)\, a_{ij}\, b_j (O_{t+1})\, \beta_{t+1}(j)}$$

E-step → α, β 계산 → $\gamma_t(i)$, $\xi_t(i,j)$

M-step → $\gamma_t(i)$, $\xi_t(i,j)$ 사용하여 HMM(λ) 개선

Baum-Welch Algorithm 활용

π_i^{new} = t가 1일때 S_i에 있을 확률 = $\gamma_1(i)$

Counting으로 estimation

$$a_{ij}^{new} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (1 \le i,j \le n) : \frac{S_i \to S_j \text{ 로 전이할 기댓값}}{S_i \text{에서 전이할 기댓값}}$$

$$b_i(v_k)^{new} = \frac{S_i\text{에서 } v_k \text{ 관측 확률}}{S_i \text{ 에 있을 확률}} = \frac{O_t = v_k \text{인 모든 } t\text{에 대해 } S_i\text{에 있을 확률의 합}}{t=1\ S_i \text{ 있을확률} + \cdots\ t=T \text{ 일때 } S_i\text{에 있을 확률}}$$

$$= \frac{\sum_{t=1,\, O_t=v_k}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

Obtaining π, a, b Given X, M

L: Loaded dice, F: Fair dice

LFLL ... L F → Dealer만 아는 사실(Z)
1 3 2 4 ... 5 6 → Dealer, 관객 모두 아는 사실(X)

$$\left.\begin{aligned} P(L \to L) &= P(Z_{t+1} = L \mid Z_t = L) = a \\ P(\underset{\text{주사위 눈}}{X_t = 1} \mid Z_t = L) &= b \end{aligned}\right\} \text{MLE, MAP} \cdots \text{ 사용}$$

$$P(X, Z) = P(X_1 X_2 \cdots X_t, Z_1 Z_2 \cdots Z_t)$$

$$= P(Z_1) \cdot P(X_1 | Z_1) \cdot P(Z_2 | Z_1) \cdot P(X_2 | Z_2) \cdots P(Z_t | Z_{t-1}) \cdot P(X_t | Z_t)$$

가정)

L :

1	2	3	4	5	6
$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{5}{10}$

$P(L \to L) = \frac{7}{10}$

X: 1, 6, 6 Z = LLL or FFF

$$P(166, LLL) : \underset{\text{초기 L 선택}}{\frac{1}{2}} \times \underset{P(X=1)}{\frac{1}{10}} \times \underset{P(L\to L)}{\frac{7}{10}} \times \underset{P(X=6)}{\frac{5}{10}} \times \underset{P(L\to L)}{\frac{7}{10}} \times \underset{P(X=6)}{\frac{5}{10}} = 0.0061$$

$$P(166, FFF) : \frac{1}{2} \times \frac{1}{6} \times \frac{1}{2} \times \frac{1}{6} \times \frac{1}{2} \times \frac{1}{6} = 0.0005$$

$$\Rightarrow P(166, LLL) > P(166, FFF)$$

가능한경우의수 2^3 (LLL LLF ...) ⇒ 3개의 Sequence도 너무 계산↑

∴ Use X, Marginalize Z

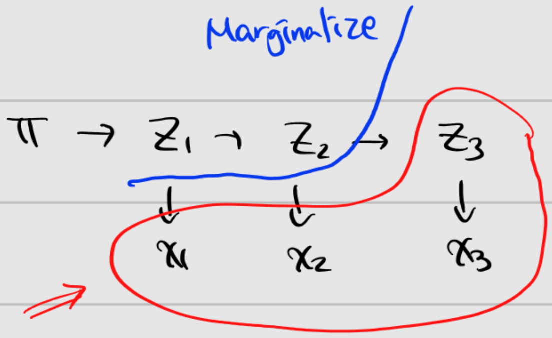
GMM에서 $P(X|\theta) = \sum_z P(X, Z|\theta)$

HMM → $P(X|\pi, a, b) = \sum_z P(X, Z|\pi, a, b)$

$$P(X) = \sum_Z P(X, Z) = \sum_{z_1}\sum_{z_2}\cdots\sum_{z_t} P(x_1, x_2 \cdots x_t, z_1, z_2 \cdots z_t)$$

$$= \sum_{z_1}\sum_{z_2}\cdots\sum_{z_t} \pi_z \prod_{t=2}^{T} a_{z_{t-1} z_t} \prod_{t=1}^{T} b_{z_t x_t}$$

$$\begin{cases} P(z_t \mid z_{t-1}) = a \\ P(x_t \mid z_t) = b \end{cases}$$



$$P(x_1\, x_2 \cdots x_t,\, z_t^k = 1)$$

$$= \sum_{z_{t-1}} P(\underbrace{x_1\, x_2 \cdots x_{t-1}}_{A},\, \underbrace{x_t}_{C},\, \underbrace{z_{t-1}}_{A},\, \underbrace{z_t^k = 1}_{B})$$

Factorization 사용

↳ $P(A, B, C) = P(A) \cdot P(B|A) \cdot P(C|A, B)$

$$\sum_{z_{t-1}} P(x_1 x_2 \cdots x_{t-1}, z_{t-1}) \cdot P(z_t^k = 1 \mid x_1 x_2 \cdots x_{t-1} z_{t-1}) \cdot P(x_t \mid z_t^k = 1, x_1 x_2 \cdots x_{t-1}, z_{t-1})$$

$$= \sum_{z_{t-1}} P(x_1\, x_2 \cdots x_{t-1}, z_{t-1}) \cdot P(z_t^k = 1 \mid z_{t-1}) \cdot P(x_t \mid z_t^k = 1)$$

z_t 결정시 X 영향X
Bayes Ball 생각하기
(굴러가지 않음)

z_t 만 결정되면
나머지 영향X
(굴러가지 않음)

$$\therefore P(x_1 x_2 \cdots x_t , z_t^k=1) = \sum_{z_{t-1}} P(x_1 x_2 \cdots x_{t-1}\, z_{t-1}) \cdot P(z_t^k=1 \mid z_{t-1}) \cdot P(x_t \mid z_t^k=1)$$

$$= \underbrace{P(x_t \mid z_t^k=1)}_{b_{z_t x_t}} \sum_{z_{t-1}} P(x_1\, x_2 \cdots x_{t-1}\, z_{t-1}) \cdot \underbrace{P(z_t^k=1 \mid z_{t-1})}_{a_{z_t z_{t-1}}}$$

$$P(x_1 x_2 \cdots x_t , z_t^k=1) = b_{z_t x_t} \sum_{z_{t-1}} P(x_1 x_2 \cdots x_{t-1}\, z_{t-1}) \cdot a_{z_t z_{t-1}}$$

recursion

$$P(x_1 \cdots x_t , z_t^k=1) = \alpha_t^k = b_{z_t x_t} \sum_i \alpha_{t-1}^i \, a_{z_t z_{t-1}}$$

<Forward>

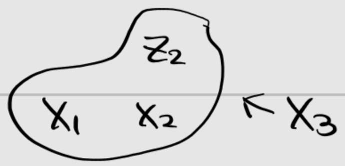
Initialize $\alpha_1^k = b_{k x_1} \pi_k$

Iterate $\alpha_t^k = b_{k x_t} \sum_i \alpha_{t-1}^i \cdot a_{ik}$

return $\sum_i \alpha_T^i$

$\alpha_t^k \to$ DP Memoization 사용하기

Forward 만으로 설명이 가능한가?



x_3가 영향이 없을까?

<Backward>

$P(Z_t^k=1 \mid X)$ joint probability 사용

$P(A,B) = P(A)\cdot P(B|A)$

↙ 활용

$$P(Z_t^k=1, X) = P(\underbrace{x_1 x_2 \cdots x_t, Z_t^k=1}, / x_{t+1} x_{t+2} \cdots x_T)$$

$$= P(x_1 x_2 \cdots x_t, Z_t^k=1) \cdot P(x_{t+1} x_{t+2} \cdots x_T \mid x_1 x_2 \cdots x_t, Z_t^k=1)$$

$$= P(x_1 x_2 \cdots x_t, Z_t^k=1) \cdot P(x_{t+1} x_{t+2} \cdots x_T \mid Z_t^k=1)$$

↳ forward 에서

$$\begin{tikzcd}
Z_1 \arrow[r] & Z_2 \arrow[r] & Z_3 \arrow[r] \arrow[d] & Z_4 \arrow[d] \\
x_1 & x_2 & x_3 & x_4
\end{tikzcd}$$

Z_2만 알면 $x_1 x_2 \cdots$ 필요 X

기존 $\alpha_t^k = P(x_1 x_2 \cdots x_t, Z_t^k=1)$

new) $\beta_t^k = P(x_{t+1} x_{t+2} \cdots x_T \mid Z_t^k=1)$

$$\beta_t^k = P(x_{t+1} x_{t+2} \cdots x_T \mid Z_t^k=1)$$

$$= \sum_{Z_{t+1}} P(Z_{t+1}, x_{t+1}, x_{t+2} \cdots x_T \mid Z_t^k=1)$$

→ Marginalize

$$\begin{tikzcd}
Z_t \arrow[r] & \boxed{Z_{t+1} \downarrow x}
\end{tikzcd}$$

Ball 생각

$$= \sum_i \underbrace{P(Z_{t+1}^i=1 \mid Z_t^k=1)} \cdot \underbrace{P(x_{t+1} \mid Z_{t+1}^i=1, \cancel{Z_t^k=1})} \cdot P(x_{t+2} \cdots x_T \mid Z_{t+1}^i=1, \cancel{Z_t^k=1})$$

$$= \sum_i a_{Z_t Z_{t+1}} \cdot b_{Z_{t+1} x_{t+1}} \cdot \beta_{t+1}^{(i)}$$

$$\therefore \beta_t^k = \sum_i a_{ki}\, b_{i x_t}\, \beta_{t+1}^i$$

<Decoding> Viterbi Decoding

P(X)로 나눠줘도 normalize 된거라 같음

$$k^* = \text{argmax}_k P(z^k=1 \mid X) = \text{argmax}_k P(z^k=1, X)$$

$$V_t^k = \max_{z_1 z_2 \cdots z_{t-1}} P(x_1, x_2 \cdots x_{t-1}, z_1 z_2 \cdots z_{t-1}, x_t, z_t^k = 1)$$

$$= \max_{z_1 z_2 \cdots z_{t-1}} P(x_t, z_t^k=1 \mid x_1 x_2 \cdots x_{t-1}, z_1 z_2 \cdots z_{t-1}) \cdot P(x_1 x_2 \cdots x_{t-2}, z_1 z_2 \cdots \underbrace{x_{t-1} z_{t-1}}_{\text{repeat}})$$

$$= \max_{z_{t-1}} P(x_t, z_t^k=1 \mid z_{t-1}) \cdot \underbrace{\max_{z_1 z_2 \cdots z_{t-2}} P(x_1 x_2 \cdots x_{t-2}, z_1 z_2 \cdots x_{t-1}, z_{t-1})}_{\text{반복}}$$

$$= \max_{z_{t-1}} P(x_t, z_t^k=1 \mid z_{t-1}) \cdot V_{t-1}^k$$

Algorithm

Initialize → $V_1^k = b_{k x_1} \pi_k$

Iterate → $V_t^k = b_{k, idx(x_t)} \max_{i \in z_{t-1}} a_{ik} V_{t-1}^i$

$\text{trace}_t^k = \text{argmax}_{i \in z_{t-1}} a_{ik} V_{t-1}^i$

return $P(X, z^*) = \max_k V_T^k, z_T^* = \text{argmax}_k V_T^k, z_{t-1}^* = \text{trace}_t^{z_t^*}$

단점 → Underflow 발생

0.1 × 0.1 × .. 0.3 너무 작은 수

→ log로 극복

<Only X is given>

only X

don't have Z, π, a, b

① optimized $\hat{\pi}$, $\hat{a}$, $\hat{b}$ with X

② most probable Z with X, $\hat{\pi}$, $\hat{a}$, $\hat{b}$

EM 알고리즘 활용

Expectation step $q^{t+1}(z) = P(Z \mid X, \pi^t, a^t, b^t)$

Maximization step $\pi^{t+1}, a^{t+1}, b^{t+1} = \arg\max Q(\pi, a, b, q^{t+1})$