Hidden Markov Model

Σ : Alphabet of symbols (Observed values)

Q : hidden states

transition probability : a_{kl} for each $k, l \in Q$

Emission probability : $e_k(b)$ for each $k \in Q$ and $b \in \Sigma$

loaded die problem

$\Sigma = \{1, 2, 3, 4, 5, 6\}$ observed states

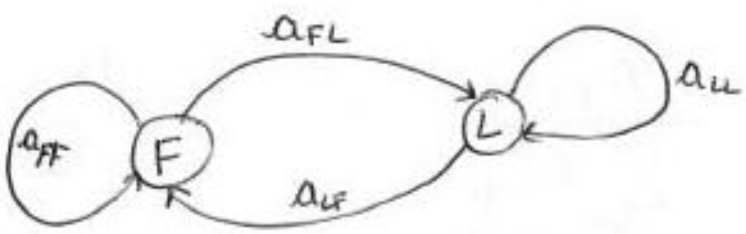
$Q = \{F, L\}$ hidden states

Emission → $e_F(i) = \frac{1}{6}$ for all $i = 1, \dots, 6$

$e_L(i) = 0.1$ for all $i = 1, \dots, 5$ and $e_L(6) = 0.5$

transition → $a_{FF} = 0.95$ $\quad a_{LL} = 0.9$

$a_{FL} = 0.05$ $\quad a_{LF} = 0.1$



Consider the following rolls

Observed (x) = 2 1 6 6 5 2 6 1

Underlying die (π) = F F L L F L L F

$$P(x|\pi) = a_{\pi_0 \pi_1} \prod_{i=1}^{L} e_{\pi_i}(x_i) \cdot a_{\pi_i \pi_{i+1}}$$

π_0 = begin

π_{L+1} = end

HMM computational problems

	Hidden sequence known	Hidden sequence unknown.
Transition and emission probabilities known	Model fully specified	Use ~~to~~ viterbi to determine optimal hidden sequence
Transition and emission probabilities unknown	Maximum likelihood	Expected maximization and also known as Baum Welch

I case : Probabilities unknown but hidden sequence known. (Maximum likelihood)

Hidden sequence ✓
Probability . not given ×

Observed (X) : 2 1 6 6 5 2 6 1
Hidden (π) : F F L L F L L F

$$a_{kl} = \frac{A_{kl}}{\sum_{q \in Q} A_{kq}}$$

state outcome

A_{KL} states

$A_{FF} = 1$ $\quad$ $A_{FL} = 2$

$A_{LF} = 2$

$A_{LL} = 2$

$a_{FF} = \frac{\cancel{2}\,1}{2+1} = \frac{1}{3}$ $\quad$ $a_{FL} = \frac{2}{2+1} = \frac{2}{3}$

$a_{LF} = \frac{2}{4} = \frac{1}{2}$ $\quad$ $a_{LL} = \frac{2}{4} = \frac{1}{2}$

Calculating the emission probability

	1	2	3	4	5	6
$E_F(i)$	2	1	0	0	1	0
$E_L(i)$	0	1	0	0	0	3

observed (x): 2 1 6 6 5 2 6 1
Hidden(π): F F L L F L L F

$$e_k(b) = \frac{E_k(b)}{\sum_{\sigma \in \Sigma} E_k(\sigma)}$$

$$e_F(1) = \frac{2}{2+1+0+0+1+0} = \frac{2}{4}$$

$$e_F(2) = \frac{1}{2+1+0+0+1+0} = \frac{1}{4}$$

$$e_F(3) = 0$$

$$e_F(4) = 0$$

$$e_F(5) = \frac{1}{4}$$

$$e_F(6) = 0$$

$$e_L(1) = 0$$

$$e_L(2) = \frac{1}{3+1} = \frac{1}{4}$$

$$e_L(3) = 0$$

$$e_L(4) = 0$$

$$e_L(5) = 0$$

$$e_L(6) = \frac{3}{4}$$

Substituting the above calculated probability in equation

$$P(x|\pi) = a_{\pi_0 \pi_1} \prod_{i=1}^{L} e_{\pi_i}(x_i)\,(x_i) \cdot a_{\pi_i \pi_{i+1}}$$

π_0 = begin
π_{L+1} = end

2) Probability known but hidden sequence unknown:

<u>Viterbi Algorithm</u>

No hidden states ×

Given the probability ✓

$\dot{V}_k(i)$ x_1 x_2 x_3 ..... $\underset{\substack{\uparrow \\ K\ \text{probability}}}{x_K}$

number of states $\boxed{\text{begin, fair, loaded}}$

for the first state

$$V_B(0) = 1$$
$$V_F(0) = V_L(0) = 0$$

Recurrence: for $i = 0, \dots, n-1$

$$V_L(i+1) = e_L(x_{i+1}) \max \begin{Bmatrix} V_L(i)\, a_{LL} \\ V_F(i)\, a_{FL} \\ V_B(i)\, a_{BL} \end{Bmatrix}$$

$$V_F(i+1) = e_F(x_{i+1}) \max \begin{Bmatrix} V_L(i)\, a_{LF} \\ V_F(i)\, a_{FF} \\ V_B(i)\, a_{BF} \end{Bmatrix}$$

$$V(i+1) = \max\{V_F(i+1), V_L(i+1)\}$$

$$T(i+1) = \begin{Bmatrix} 0 \text{ if max from } V_F \\ 1 \text{ if max from } V_L \end{Bmatrix}$$

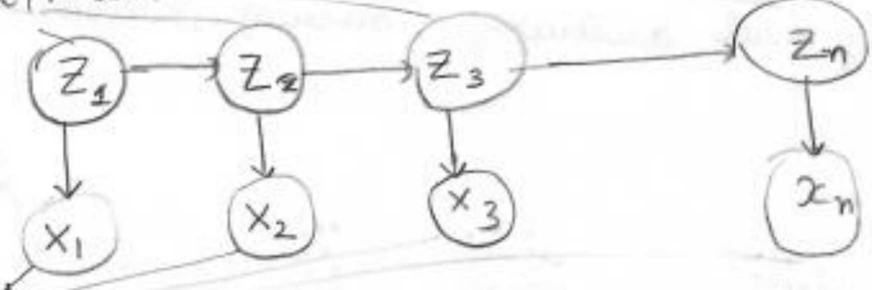
Hidden Markov Model (HMM)

$Z_1, \ldots Z_n \in \{1, \ldots, n\}$ → Hidden model / variables

$x_1, \ldots, x_n \in \{$ discreate, finite, Real valued, $\mathbb{R}$, $\mathbb{R}^d \}$

latent / Hidden

"Trellis Diagram"



Observed

X_i : They are also called observed random variable

$D = (x_1, \ldots, x_n)$

Z = Hidden / Latent variables

$$\boxed{P(x_1, \ldots, x_n, Z_1, \ldots, Z_n) = P(Z_1)\, P(x_1 \mid Z_1) \prod_{k=2}^{n} P(Z_k \mid Z_{k-1})\, P(x_k \mid Z_k)}$$

⇓

Joint distribution

Application →
Hand writing Recognization
Observed — Hidden Markov → (handwritten digit)
HIDDEN MARKOV
Hidden must be what letter it could be

1-26

Parameter

- Transition probabilities T_{ij} : (Transition matrix)

$$T(i,j) = P(Z_{k+1} = j \mid Z_k = i) \neq 1 \leftarrow \text{density (pdf)}$$

- Emission probabilities : $\mathcal{E}_i(x) = P(x_k \mid Z_k = i)$ → for pdf
$\forall\, i \in \{1, \ldots, m\}\ x \in \mathcal{X}$ continuous values

(i.e. $\mathcal{E}_i$ is probability dist on $\mathcal{X}$)

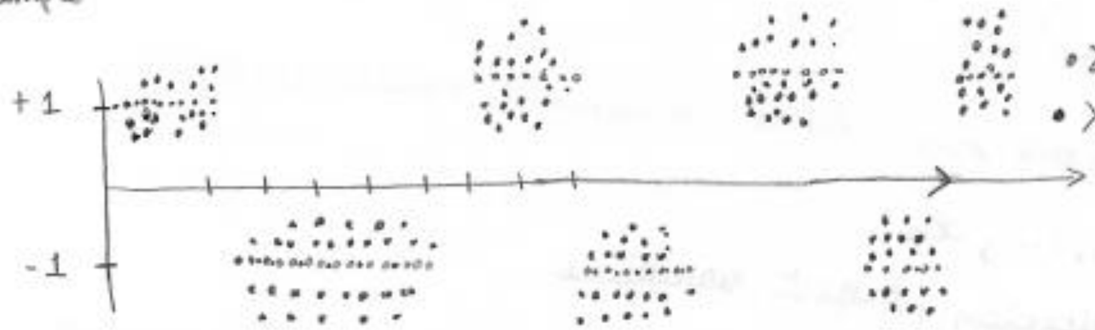
$$\mathcal{E}_i(x) = P(X_k = x \mid Z_k = i) \leftarrow \text{P.m.f} \rightarrow \text{discreate values}$$

- Initial distribution: $\pi(i) = P(Z_1 = i)$, $i \in \{1, \ldots, m\}$

$$P(x_1, \ldots, x_n, z_1, \ldots, z_n) = \pi(1)\, \underbrace{\mathcal{E}_{z_1}(x_1)}_{\text{emission}} \prod_{k=2}^{n} \underbrace{T(z_{k-1}, z_k)}_{\text{transition}} \times \underbrace{\mathcal{E}_{z_k}(x_k)}_{\text{emission}}$$

Remark: $\mathcal{E}_i$'s pretty much arbitrary (discreate, Real valued)

Example

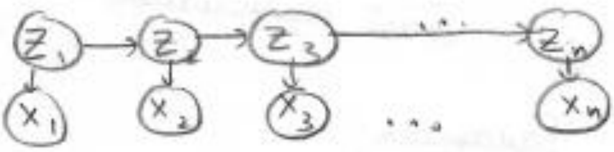


- $Z_k \in \{-1, 1\}$ → hidden
- $X \in \mathbb{R}$ → observed

$$T = \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & .99 & .01 \\ 2 & .01 & .99 \end{array}$$

✧ Forward - Backward Algorithm

"Dynamic Programming" perfect example



Assumes $P(x_k \mid Z_k)$, $P(Z_k \mid Z_{k-1})$, $P(Z_1)$ these all are known

Forward Backward Algorithm: Compute $P(Z_k \mid x_{1:n})$

Forward algorithm: Compute $P(Z_k, x_{1:k})$ $\forall k = 1, \ldots, n$

Backward algorithm: $P(x_{k+1:n} \mid Z_k)$ $\forall k = 1, \ldots, n$

Here, $x = (x_1, \ldots, x_n)$

$x_{i:j} = (x_i, x_{i+1}, \ldots, x_{i+2}, \ldots, x_j)$

$x = x_{1:n}$

$x_{k+1:n} = (x_{k+1}, \ldots, x_n)$

conditionally independent

$$P(z_k|x) \underset{z_k}{\propto} P(z_k, x) = P(x_{k+1:n} \mid z_k, x_{1:k})\, P(z_k, x_{1:k})$$

$$= \overbrace{P(x_{k+1:n} \mid z_k)}^{\text{Backward}} \times \overbrace{P(z_k, x_{1:k})}^{\text{Forward}}$$

Q What you can do from above:

→ Inference : $P(z_k \neq z_{k+1}|x)$ "Change detection"

→ Estimate parameters of HMM using ("Baum-Welch").

→ Sampling from posterior list : $z|x$

☆ Forward Algorithm (for HMM)

Goal / Objective → Compute $P(z_k, x_{1:k})$

Fix : $x_1, \dots x_n$

$$\alpha_k(z_k) = P(z_k, x_{1:k}) = \sum_{z_{k-1}=1}^{m} P(z_k, z_{k-1}, x_{1:k})$$

$$= \sum_{z_{k-1}} P(x_k|z_k, z_{k-1}, x_{1:k-1})\, P(z_k|z_{k-1}, x_{1:k-1})\, P(z_{k-1}, x_{1:k-1})$$

These are conditionally independent

$$= \sum_{z_{k-1}} \underbrace{P(x_k|z_k)}_{\text{Emission probability}}\ \underbrace{P(z_k|z_{k-1})}_{\text{Transition probability}}\ \underbrace{P(z_{k-1}, x_{1:k-1})}_{\alpha_{k-1}(z_{k-1})}$$

$$\boxed{\alpha_k(z_k) = \sum_{z_{k-1}=1}^{m} P(x_k|z_k)\, P(z_k|z_{k-1})\, \alpha_{k-1}(z_{k-1})}$$

$$\forall\, k = 2, \dots, n$$

$$\alpha_1(z_1) = P(z_1, x_1) = \underbrace{P(z_1)}_{\text{Initial dis}} \underbrace{P(x_1|z_1)}_{\text{Emission distribution}}$$

$\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$

$\Theta(m)$ for each z_k

$\Theta(m^2)$ for each k

$\Theta(nm^2)$ complexity of this algorithm

$O(f(n))$ means $\exists\, c\ \exists N$ s.t $\forall n > N$

$\text{time}_n \leq c\, f(n)$

$\Theta(f(n))$ means

$\exists\, c, c'\ \exists N$ s.t $\forall n > N$

$$\underset{\text{lower}}{c f(n)} \leq \text{time}_n \leq \underset{\text{upper found}}{c\, f(n)}$$

<u>Comparing it to naive approach:</u>

$$P(z_k | x) = \frac{P(z_k, x)}{P(x)}$$

$$P(z_k, x) = \sum_{\substack{z_1, z_2, \ldots, z_{k-1} \\ z_{k+1}, \ldots, z_n}} P(x, z) \quad \& \quad P(x) = \sum_{z_{1:n}} P(x, z)$$

$$= \underbrace{\sum_{\substack{z_1 \\ m}} \sum_{\substack{z_2 \\ m}} \cdots \sum_{\substack{z_n \\ m}}}_{m^n} P(x, z)$$

Suppose $m = 10$, $n = 100$

Time $= m^n$

Time it would take 10^{100}

On the other hand Forward Backward approach will take

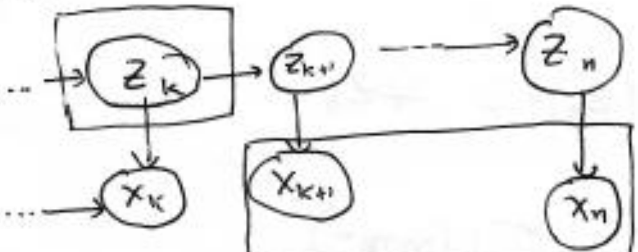
$$nm^2 = 100 \times 10^2 = 10{,}000.$$

> Remark:
>
> Dynamic Programming → Reuse earlier computations

In forward algorithm we were reusing α numbers and avoid the calculations

Backward Algorithm

Given ($x_1, \dots, x_n$ assume dist)



Goal: compute the .

$$P(x_{k+1:n} \mid z_k) \quad (\forall k = 1, \dots, n-1) \quad (\forall z_k = 1, \dots, m)$$

$$\beta_k(z_k) = P(x_{k+1:n} \mid z_k) = \sum_{z_{k+1}=1}^{m} P(x_{k+1:n}, z_{k+1} \mid z_k)$$

$$= \cancel{\sum_{z_{k+1}} P(z_{k+1} \mid z_k, x_{k+1:n})}$$

$$= \sum_{z_{k+1}} P(X_{k+2:n} \mid Z_{k+1}, \underbrace{Z_k, X_{k+1}) P(X_{k+1} \mid Z_{k+1}, Z_k}_{\text{conditional independent so ignore it}}) \times P(Z_{k+1} \mid Z_k)$$

$$= \sum \underbrace{P(X_{k+2:n} \mid Z_{k+1})}_{\beta_{k+1}(Z_{k+1})} \; \underbrace{P(x_{k+1} \mid Z_{k+1})}_{\text{known emission probability}} \; \underbrace{P(Z_{k+1} \mid Z_k)}_{\text{known transition probability}}$$

$$\boxed{\begin{aligned} &\beta_k(Z_k) = \sum_{Z_{k+1}=1}^{m} \beta_{k+1}(Z_{k+1}) P(X_{k+1} \mid Z_{k+1}) P(Z_{k+1} \mid Z_k) \quad \forall\, k = 1, \dots, n-1 \\ &\beta_n(Z_n) = 1 \quad \forall Z_n \end{aligned}}$$

$$\Theta(nm^2)$$

☆ Problem with HMM → Underflow and the "log sum exp" trick

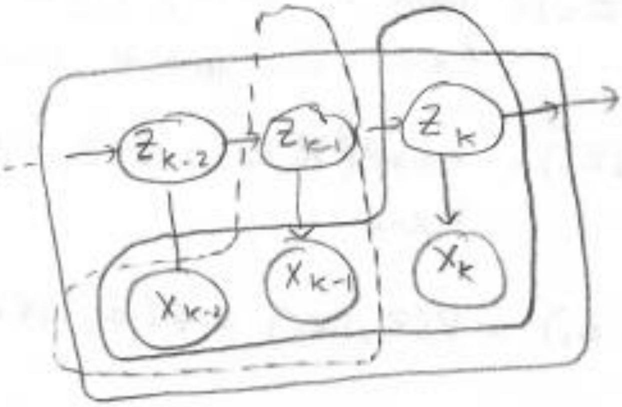
$$\underbrace{P(Z_k, X_{1:k})}_{\alpha_k(Z_k)} = \sum_{Z_{k-1}} P(\cancel{X} X_k \mid Z_k) P(Z_k \mid Z_{k-1}) \alpha_{k-1}(Z_{k-1})$$

$\boxed{\text{Note} \rightarrow \text{To represent very small number we take log.}}$

$$\log \alpha_k(Z_k) = \log \sum_{Z_{k-1}} \exp\Big(\underbrace{\log P(\pi_k \mid Z_k) + \log P(Z_k \mid Z_{k-1}) + \log \alpha_{k-1}(Z_{k-1})}_{x_i}\Big)$$

$$= b + \log \sum_{i=1}^{n} e^{a_i - b}$$

Viterbi Algorithm



Given: $x_1, \dots, x_n$

Assume distributions

Goal: Compute the most likely sequence

$$z^* = \underset{z}{\operatorname{argmax}}\ P(z|x)$$

$x = x_{1:n}$

$z = z_{1:n}$

R.K: If $f(a) \geq 0 \ \forall a$ and $g(a,b) \geq 0 \ \forall a, b$.

then $\max\limits_{a,b} f(a) g(a,b) = \max\limits_{a} \left\{ f(a) \max\limits_{b} g(a,b) \right]$

NOTE:
$$\underset{z}{\operatorname{argmax}}\ P(z|x) = \underset{z_{1:n}}{\operatorname{argmax}}\ P(z, x)$$

$$\max_{z_{1:k-1}} p(z, x)$$

We want some quantity on which we can recurse on.

$$M_k(z_k) = \max_{z_{1:k-1}} p(z, x) = \max_{z_{1:k-1}} P(z_{1:k}, x_{1:k})$$

$$= \max_{z_{1:k-1}} \overbrace{P(x_k | z_k) p(z_k | z_{k-1})}^{f(a)} \overbrace{P(z_{1:k-1}, x_{1:k-1})}$$

Substitute

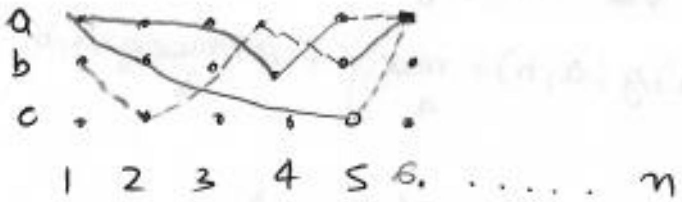
$$\boxed{\begin{aligned} a &= z_{k-1} \\ b &= z_{1:k-2} \end{aligned}}$$

$$= \max_{z_{k-1}} \left[\underbrace{p(x_k | z_k)}_{\text{known}} \underbrace{p(z_k | z_{k-1})}_{\text{known}} \underbrace{\max_{z_{1:k-2}} p(z_{1:k-1}, x_{1:k-1})}_{M_{k-1}(z_{k-1})} \right]$$

$$M_k(Z_k) = \max_{Z_{k-1}} \underbrace{P(x_k|Z_k)}_{\text{known (emission)}} \underbrace{P(Z_k|Z_{k-1})}_{\text{known (transition)}} \underbrace{M_{k-1}(Z_{k-1})} \quad \text{for } k = 2, 3, \ldots, n$$

$$\boxed{\begin{aligned} M_k(Z_k) &= \max_{Z_{K-1}} P(\cancel{x}_k|Z_k)\, P(Z_k|Z_{k-1})\, M_{k-1}|Z_{k-1}) \quad \text{for } k = 2, 3, \ldots, n \\ M_1(Z_1) &= P(Z_1|x_1) = P(Z_1)\,P(X_1|Z_1) \end{aligned}}$$

$$\max_{Z_n} M_n(Z_n) = \max_{Z_{1:n}\cancel{}} P(x_{1:n}, Z_{1:n})$$



$Z_5 = a$ $\quad$ $Z_6 = a$

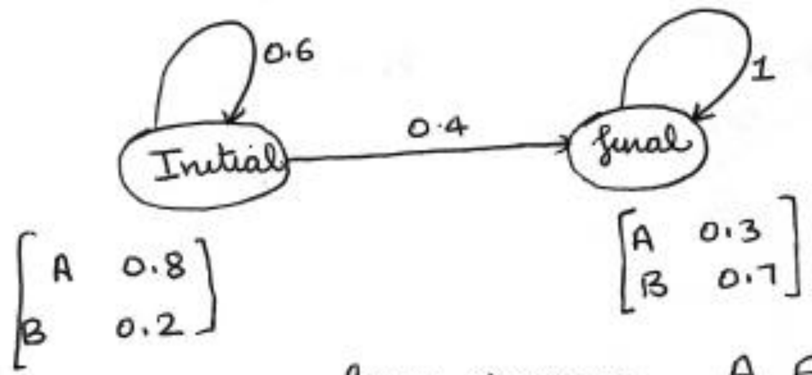
$Z_5 = b$

$Z_5 = c$

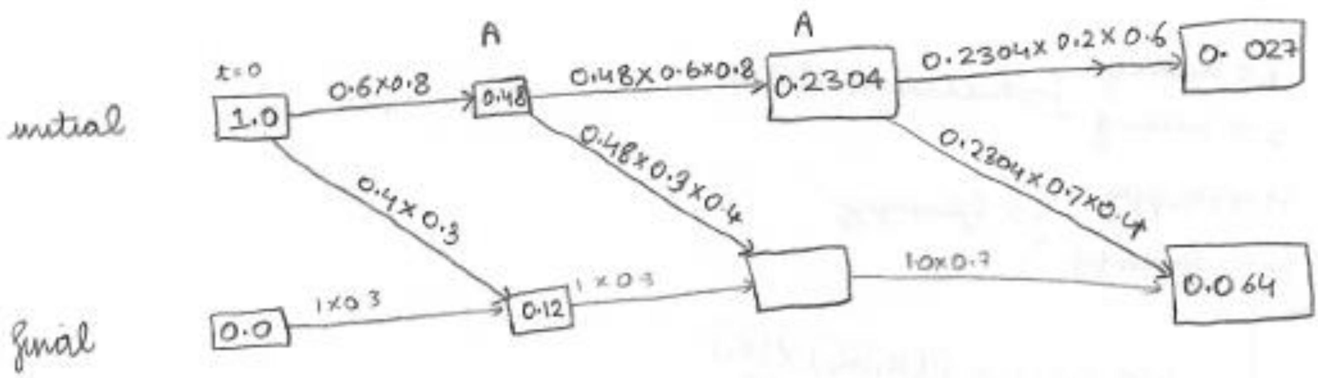
Use of Veterbi Algorithm → One thing which you can't do using forward backward algorithm is find the ~~p~~ most probable sequence of hidden state for a given sequence of observed states so ~~we~~ Veterbi Algorithm do that

$$\max_{Z_n} M_n(Z_m) = \max_{Z_{1:n}\cancel{}} P(X_{1:m}, Z_{1:n})$$

Hidden states



Given sequence A A B



forward Algorithm (for HMM)

$$\alpha_k(Z_k) = \sum_{Z_{k-1}=1}^{m} P(x_k|Z_k)\, P(Z_k|Z_{k-1})\, \alpha_{k-1}(Z_{k-1})$$

Backward

$$B_k(Z_k) = \sum_{Z_{k+1}=1}^{m} B_{k+1}(Z_{k+1})\, P(X_{k+1}|Z_{k+1})\, P(Z_{k+1}|Z_k) \quad \forall k=1,2,\ldots n-1$$

$$B_n(Z_n) = 1$$

$$B_3(S_1) = B_3(S_3) = 1.0$$

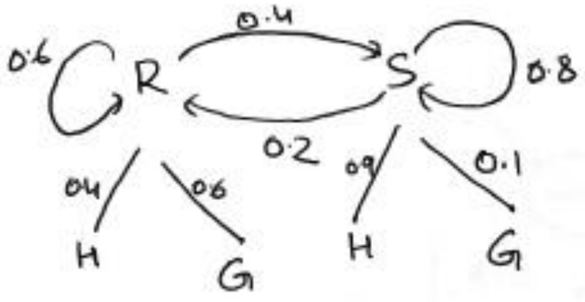
$$B_2(S_1) = \sum B_3(Z_3)\, P(X_3|Z_3)\, P(Z_3|Z_2)$$

$$= (1)\times 0.2\times 0.6 + (1)\times 0.7\times 0.4 = 0.40$$

$$B_1(S_1) = \sum B_2(Z_2)\, P(X_2|Z_2)\, P(Z_2|Z_1) = (0.40)\times(0.3\times 0.4) + (0.40)\times(0.6)\times 0.8$$

$$= 0.048 + 0.192$$

Hidden Markov Models Example 2



$P(R_0) = \frac{1}{2}$

$P(S_0) = \frac{1}{2}$

Sequence

R → Rainy
S → Sunny } → hidden states

H → Happy
G → Grumpy } → Grumpy

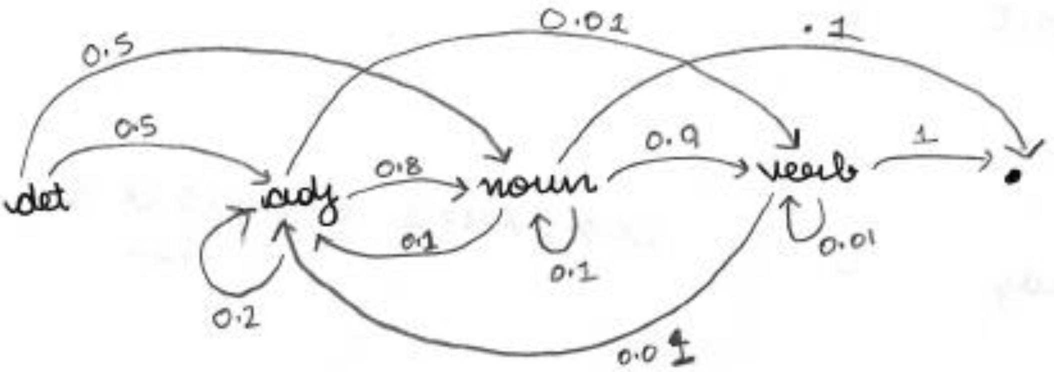
$$P(R_1|H_1) = \frac{P(H_1|R_1)\,P(R_1)}{P(H_1)}$$

$$P(R_1) = P(R_1|R_0)\,P(R_0) + P(R_1|S_0)\,P(S_0)$$

$$= 0.6 \times \frac{1}{2} + 0.2 \times \frac{1}{2}$$

$$= 0.4$$

Example 3



The	blue	bank	closed
det	adj	n	v
	or	adj	adj
	n	v	

P(e/s)

det → 1 The
adj → blue 0.4, blank 0.3, closed 0.3

noun → blue 0.4, bank 0.6

verb → "blue" .2
bank .2
closed .6

	the $t=0$	blue $t=1$	bank $t=2$	Closed $t=3$
det	1	0	0.2×0	0
adj	0	$1 \times 0.5 \times .4 = .2$	$0.2 \times 0.2 \times .3 = .012$	$0.96 \times 0.01 \times 0.3$
noun	0	$1 \times 0.5 \times .4 = .2$	$0.2 \times 0.8 \times 0.6 = 0.096$ $0.01 \times 0.2 \times 0.6$	$0.96 \times 0.01 \times 0$
verb	0	0	$0.2 \times 0.01 \times .2 = \epsilon$	$0.096 \times 0.9 \times$