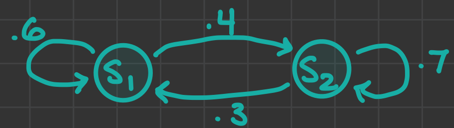
Hidden Markov Models

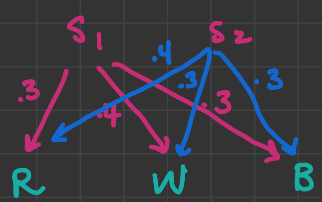
Source: James Allen's CSC 248/448 Speech Recognition course materials, Rochester 2003

Suppose someone is drawing balls from two buckets, which we'll call STATE 1 and STATE 2. Initially, they're more likely to draw from state 1 (w.p. 0.8) versus state 2 (w.p. 0.2). Then they switch between buckets following this transition matrix, or equivalently, this diagram:

$$\text{from}\;\begin{array}{c} \\ s_1 \\ s_2 \end{array}\!\!\begin{array}{c} \overset{\text{to}}{\begin{matrix} s_1 & s_2 \end{matrix}} \\ \begin{pmatrix} .6 & .4 \\ .3 & .7 \end{pmatrix} \end{array}$$



State 1 has a few more white balls than red and blue, and state 2 has a few more red balls than white and blue, according to this diagram:



The question: if the person tells us the colors of the balls he's drawn, what's our best guess about the trajectory of the states they've been in?

That is, if we let w_t stand for the unknown state at each time, we need to find w_1, w_2, w_3, and w_4 that maximize the probability of drawing a red ball, then a white ball, then a blue ball, and finally another blue ball:

$$\max_{\substack{w_1, w_2, \\ w_3, w_4 \in s_1, s_2}} P(R\text{-}W\text{-}B\text{-}B \mid w_1, w_2, w_3, w_4)$$

This is like MLE of discrete unobservables in a dynamic setting.

Solution #1: brute force

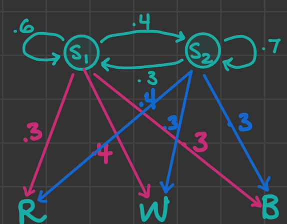
$P(R\text{-}W\text{-}B\text{-}B \mid \omega_1, \omega_2, \omega_3, \omega_4)$

Pick any trajectory, like staying in S_1 the entire time. We can calculate the probability of seeing R-W-B-B given that we stay in S_1:

$$P(R\text{-}W\text{-}B\text{-}B \mid S_1, S_1, S_1, S_1) =$$

$$\begin{aligned}
&P(\text{start in } S_1)\, P(R \mid S_1) \times \\
&P(\text{move from } S_1 \text{ to } S_1)\, P(W \mid S_1) \times \\
&P(\text{move from } S_1 \text{ to } S_1)\, P(B \mid S_1) \times \\
&P(\text{move from } S_1 \text{ to } S_1)\, P(B \mid S_1)
\end{aligned}$$

$$\begin{aligned}
&= (.8)(.3)(.6)(.4)(.6)(.3)(.6)(.3) \\
&= .8^1 .6^3 .4^1 .3^3 \\
&\approx .0019
\end{aligned}$$



If we calculated this probability for all possible trajectories (there are 16 of them in this example), our best guess is simply the trajectory with the highest probability!

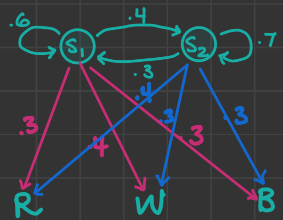
But as T and the number of possible states increase, doing all those computations becomes really time consuming. Luckily, there's a trick!

Solution #2 : Forward-Backward Algorithm

Let $\alpha_t(i)$ be the probability of seeing the observations 1 to t and ending up in state i in time t. So for example:

$\alpha_1(s_1) = P($see R and end in s_1 at time 1$)$
$\alpha_2(s_2) = P($see R-W and end in s_2 at time 2$)$
$\alpha_3(s_2) = P($see R-W-B and end in s_2 at time 3$)$
$\alpha_4(s_1) = P($see R-W-B-B and end in s_1 at time 4$)$

We can calculate these probabilities:



$\alpha_1(s_1) = P(\text{start at } s_1)\,P(R \mid s_1)$
$= (.8)(.3) = \boxed{.24}$

$\alpha_1(s_2) = P(\text{start at } s_2)\,P(R \mid s_2)$
$= (.2)(.4) = \boxed{.08}$

Right now it seems like it's more likely that we started in s_1, but we can't say for sure until we factor in the rest of the observations.

$\alpha_2(s_1) = P($see R-W and end in s_1 at time 2$)$

Note: there are 2 ways of ending in s_1 at time 2: (1) $s_1 \to s_1$ (2) $s_2 \to s_1$, so we'll ADD the probabilities of these ways.

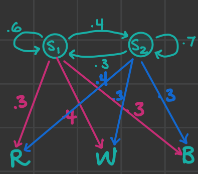
$= P($see R-W and traj was $s_1 \to s_1) +$
$P($see R-W and traj was $s_2 \to s_1)$

$= P(W \mid s_1)\,P(\text{move from } s_1 \text{ to } s_1)\,\underbrace{P(R \mid s_1)\,P(\text{start in } s_1)}_{\alpha_1(s_1) = .24}$

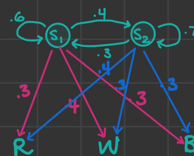
$+$

$P(W \mid s_1)\,P(\text{move from } s_2 \text{ to } s_1)\,\underbrace{P(R \mid s_2)\,P(\text{start in } s_2)}_{\alpha_1(s_2) = .08}$

$= (.4)(.6)(.24) + (.4)(.3)(.08)$
$= \boxed{.0672}$



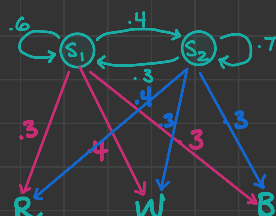
$$\alpha_2(s_2) = P(\text{see R-W, end in } s_2 \text{ at time } 2) =$$
$$= P(\text{R-W}, s_1 \to s_2) + P(\text{R-W}, s_2 \to s_2)$$
$$= P(W|s_2)\,P(s_1 \to s_2)\,\underbrace{P(R|s_1)\,P(\text{start in } s_1)}_{\alpha_1(s_1) = .24}$$
$$+$$
$$P(W|s_2)\,P(s_2 \to s_2)\,\underbrace{P(R|s_2)\,P(\text{start in } s_2)}_{\alpha_1(s_2) = .08}$$



$$= (.3)(.4)(.24) + (.3)(.7)(.08)$$
$$= \boxed{.0456}$$

$$\alpha_3(s_1) = P(\text{see R-W-B and end in } s_1 \text{ in } t=3)$$
$$= P(\text{R-W-B and go from } s_1 \text{ in } t=2 \text{ to } s_1 \text{ in } t=3)$$
$$+$$
$$P(\text{R-W-B and go from } s_2 \text{ in } t=2 \text{ to } s_1 \text{ in } t=3)$$
$$= P(B|s_1)\,P(s_1 \to s_1)\,\underbrace{P(\text{see R-W and end up in } s_1 \text{ in } t=2)}_{\alpha_2(s_1) = .0672}$$
$$+$$
$$P(B|s_1)\,P(s_2 \to s_1)\,\underbrace{P(\text{see R-W and end up in } s_2 \text{ in } t=2)}_{\alpha_2(s_2) = .0456}$$

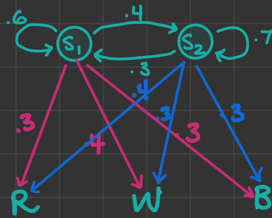
$$= (.3)(.6)(.0672) + (.3)(.3)(.0456)$$
$$= \boxed{.0162}$$



$$\alpha_3(s_2) = P(B|s_2)\,P(s_1 \to s_2)\,\alpha_2(s_1)$$
$$+$$
$$P(B|s_2)\,P(s_2 \to s_2)\,\alpha_2(s_2)$$

$$= (.3)(.4)(.0672) + (.3)(.7)(.0456)$$
$$= \boxed{.01764}$$

$$\alpha_4(s_1) = P(B|s_1)P(s_1 \to s_1)\,\alpha_3(s_1) \; + $$
$$P(B|s_1)\,P(s_2 \to s_1)\,\alpha_3(s_2)$$
$$= (.3)(.6)(.0162) + (.3)(.3)(.01764)$$
$$= \boxed{.0045036}$$



$$\alpha_4(s_2) = P(B|s_2)\,P(s_1 \to s_2)\,\alpha_3(s_1) \; +$$
$$P(B|s_2)\,P(s_2 \to s_2)\,\alpha_3(s_2)$$
$$= (.3)(.4)(.0162) + (.3)(.7)(.01764)$$
$$= \boxed{.0056484}$$

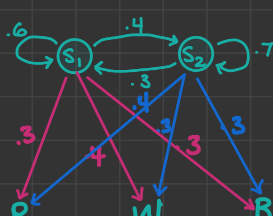
There's a higher probability of ending up in s_1 after the entire trajectory, so we can be confident our best guess for w_4 is s_1. But what about w_3, w_2, and w_1? There's a little more work we need to do (Backward algorithm).

Backward Probability:

Let $\beta_t(i)$ be the probability of starting in state s_i at time t and generating the rest of the observation sequence $o_{t+1}, \dots, o_T$.

$$\beta_t(i) = P(\text{see } o_{t+1}, \dots, o_T \text{ given you're in state } s_i \text{ in time } t)$$

$$\beta_3(s_1) = P(\text{see B given you're in state } s_1 \text{ in time } 3)$$
$$= P(B|s_1)P(s_1 \to s_1) + P(B|s_2)P(s_1 \to s_2)$$
$$= (.3)(.6) + (.3)(.4)$$
$$= \boxed{.3}$$



$$\beta_3(s_2) = P(\text{see B given } s_2 \text{ in } t3)$$
$$= P(B|s_1)P(s_2 \to s_1) + P(B|s_2)\,P(s_2 \to s_2)$$
$$= (.3)(.3) + (.3)(.7)$$

$= \boxed{.3}$

$\beta_2(s_1) = P(\text{see B-B given } s_1 \text{ in } t2)$

$= P(B|s_1)P(s_1 \to s_1)\underbrace{P(\text{see B given } s_1 \text{ in } t3)}_{\beta_3(s_1)}$

$+$

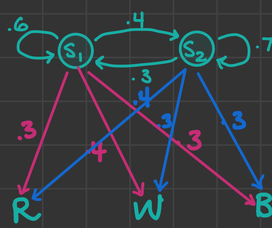
$P(B|s_2)P(s_1 \to s_2)\underbrace{P(\text{see B given } s_2 \text{ in } t3)}_{\beta_3(s_2)}$

$= (.3)(.6)(.3) +$

$(.3)(.4)(.3)$

$= \boxed{.09}$

t_1	t_2	t_3	t_4
R	W	B	B
	$s_1 \to$	$? \to$	?
		s_1	
		s_2	



$\beta_2(s_2) = P(\text{see B-B given } s_2 \text{ in } t2)$

$= P(B|s_1)P(s_2 \to s_1)\, P(\text{see B given } s_1 \text{ in } t3)$

$+$

$P(B|s_2)P(s_2 \to s_2)\, P(\text{see B given } s_2 \text{ in } t3)$

t_1	t_2	t_3	t_4
R	W	B	B
	$s_2 \to$	$? \to$	?
		s_1	
		s_2	

$= (.3)(.3)\beta_3(s_1)$

$+ (.3)(.7)\,\beta_3(s_2)$

$= (.3)(.3)(.3) + (.3)(.7)(.3)$

$= \boxed{.09}$

$\beta_1(s_1) = P(\text{see W-B-B given } s_1 \text{ in } t1)$

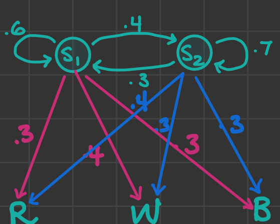
$= P(W|s_1)P(s_1 \to s_1)\, P(\text{BB and } s_1 \text{ in } t2)$

t_1	t_2	t_3	t_4
R	W	B	B
$s_1 \to$	$s_1 \to$	$? \to$	?
	s_2		

$+$

$P(W|s_2)P(s_1 \to s_2)\, P(\text{BB, } s_2 \text{ in } t2)$

$= (.4)(.6)(.09) + (.3)(.4)(.09)$



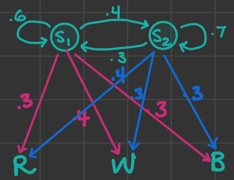
$= \boxed{.0324}$

$$\beta_1(s_2) = P(\text{see W-B-B given } S_2 \text{ in } t_1)$$

$$= P(W \mid s_1) P(s_2 \to s_1) P(BB \text{ and } s_1 \text{ in } t_3)$$

$$+$$

$$P(W \mid s_2) P(s_2 \to s_2) P(BB \text{ and } S_2 \text{ in } t_3)$$



t_1	t_2	t_3	t_4
R	W	B	B
$S_2 \to$	$S_1 \to$	$? \to$	?
	S_2		

$$= (.3)(.3)\beta_3(s_1) + (.3)(.7)\beta_3(s_2)$$

$$= (.3)(.3)(.3) + (.3)(.7)(.3)$$

$$= \boxed{.09}$$

Using the forward and backward probs:

Bayes: $P(x \mid y) = \dfrac{P(x \text{ and } y)}{P(y)}$

$$P(w_1 = s_1 \mid o_1, o_2, o_3, o_4) =$$

$$\frac{P(w_1 = s_1 \text{ and } o_1, o_2, o_3, o_4)}{P(o_1, o_2, o_3, o_4)}$$

$$= \frac{P(\text{see } o_1 \text{ and end in } s_1 \text{ at } t_1)\, P(\text{see } o_2, o_3, o_4 \mid s_1 \text{ at } t_1)}{P(\text{see } o_1, o_2, o_3, o_4 \text{ and end in } s_1 \text{ at } t_4) + P(\text{see } o_1, o_2, o_3, o_4 \text{ and end in } s_2 \text{ at } t_4)}$$

$$= \frac{\alpha_1(s_1)\, \beta_1(s_1)}{\alpha_4(s_1) + \alpha_4(s_2)}$$

$$= \frac{(.24)(.0324)}{.0045 + .0056}$$

$$= .77$$

$\alpha_1(s_1) = .24$ $\quad\quad$ $\beta_1(s_1) = .0324$

$\alpha_1(s_2) = .08$ $\quad\quad$ $\beta_1(s_2) = .0297$

$\alpha_2(s_1) = .067$ $\quad\quad$ $\beta_2(s_1) = .09$

$\alpha_2(s_2) = .046$ $\quad\quad$ $\beta_2(s_2) = .09$

$\alpha_3(s_1) = .016$ $\quad\quad$ $\beta_3(s_1) = .3$

$\alpha_3(s_2) = .017$ $\quad\quad$ $\beta_3(s_2) = .3$

$\alpha_4(s_1) = .0045$ $\quad\quad$ $\beta_4(s_1) = 1$

$\alpha_4(s_2) = .0056$ $\quad\quad$ $\beta_4(s_2) = 1$

$$P(w_1 = s_2 \mid o_1, o_2, o_3, o_4) = \frac{\alpha_1(s_2)\,\beta_1(s_2)}{\alpha_4(s_1) + \alpha_4(s_2)} = \frac{(.08)(.0297)}{.0045 + .0056} = .23$$

It's more likely we started in s_1.

$$P(w_2 = s_1 \mid o_1, o_2, o_3, o_4) = \frac{\alpha_2(s_1)\,\beta_2(s_1)}{\alpha_4(s_1) + \alpha_4(s_2)} = \frac{(.067)(.09)}{.0045 + .0056} = .58$$

It's more likely we then stayed in s_1.

$$P(w_3 = s_1 \mid o_1, o_2, o_3, o_4) = \frac{\alpha_3(s_1)\,\beta_3(s_1)}{\alpha_4(s_1) + \alpha_4(s_2)} = \frac{(.016)(.3)}{.0045 + .0056} = .48$$

→ Then we predict we switch to s_2.

$$P(w_4 = s_1 \mid o_1, o_2, o_3, o_4) = \frac{\alpha_4(s_1)\,\beta_4(s_1)}{\alpha_4(s_1) + \alpha_4(s_2)} = \frac{(.0045)(1)}{.0045 + .0056} = .45$$

We stay in s_2.