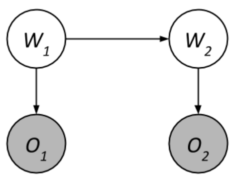
CS188 Fall 2017 Section 10: HMMs and Naive Bayes

1 HMMs

Consider the following Hidden Markov Model.



W_1	$P(W_1)$
0	0.3
1	0.7

W_t	W_{t+1}	$P(W_{t+1}\|W_t)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

W_t	O_t	$P(O_t\|W_t)$
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

Suppose that we observe $O_1 = A$ and $O_2 = B$.
Using the forward algorithm, compute the probability distribution $P(W_2|O_1 = A, O_2 = B)$ one step at a time.

1. Compute $P(W_1, O_1 = A)$.

   $P(W_1, O_1 = A) = P(W_1)P(O_1 = A|W_1)$
   $P(W_1 = 0, O_1 = A) = (0.3)(0.9) = 0.27$
   $P(W_1 = 1, O_1 = A) = (0.7)(0.5) = 0.35$

2. Using the previous calculation, compute $P(W_2, O_1 = A)$.

   $P(W_2, O_1 = A) = \sum_{x_1} P(x_1, O_1 = A)P(W_2|x_1)$
   $P(W_2 = 0, O_1 = A) = (0.27)(0.4) + (0.35)(0.8) = 0.388$
   $P(W_2 = 1, O_1 = A) = (0.27)(0.6) + (0.35)(0.2) = 0.232$

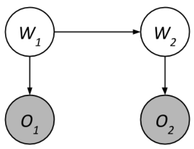
3. Using the previous calculation, compute $P(W_2, O_1 = A, O_2 = B)$.

   $P(W_2, O_1 = A, O_2 = B) = P(W_2, O_1 = A)P(O_2 = B|W_2)$
   $P(W_2 = 0, O_1 = A, O_2 = B) = (0.388)(0.1) = 0.0388$
   $P(W_2 = 1, O_1 = A, O_2 = B) = (0.232)(0.5) = 0.116$

4. Finally, compute $P(W_2|O_1 = A, O_2 = B)$.

   Renormalizing the distribution above, we have
   $P(W_2 = 0|O_1 = A, O_2 = B) = 0.0388/(0.0388 + 0.116) \approx 0.25$
   $P(W_2 = 1|O_1 = A, O_2 = B) = 0.116/(0.0388 + 0.116) \approx 0.75$

Let's use Particle Filtering to estimate the distribution of $P(W_2|O_1 = A, O_2 = B)$. Here's the HMM again:



W_1	$P(W_1)$
0	0.3
1	0.7

W_t	W_{t+1}	$P(W_{t+1}\|W_t)$
0	0	0.4
0	1	0.6
1	0	0.8
1	1	0.2

W_t	O_t	$P(O_t\|W_t)$
0	A	0.9
0	B	0.1
1	A	0.5
1	B	0.5

We start with two particles representing our distribution for W_1.
$P_1 : W_1 = 0$
$P_2 : W_1 = 1$
Use the following random numbers to run particle filtering:

$$[0.22, 0.05, 0.33, 0.20, 0.84, 0.54, 0.79, 0.66, 0.14, 0.96]$$

1. **Observe**: Compute the weight of the two particles after evidence $O_1 = A$.

   $w(P_1) = P(O_t = A|W_t = 0) = 0.9$
   $w(P_2) = P(O_t = A|W_t = 1) = 0.5$

2. **Resample**: Using the random numbers, resample P_1 and P_2 based on the weights.

   We now sample from the weighted distribution we found above. After normalizing the weights, we find that P_1 maps to range [0, 0.643), and P_2 maps to range [0.643, 1). Using the first two random samples, we find:
   $P_1 = sample(weights, 0.22) = 0$
   $P_2 = sample(weights, 0.05) = 0$

3. **Elapse Time**: Now let's compute the elapse time particle update. Sample P_1 and P_2 from applying the time update.

   $P_1 = sample(P(W_{t+1}|W_t = 0), 0.33) = 0$
   $P_2 = sample(P(W_{t+1}|W_t = 0), 0.20) = 0$

4. **Observe**: Compute the weight of the two particles after evidence $O_2 = B$.

   $w(P_1) = P(O_t = B|W_t = 0) = 0.1$
   $w(P_2) = P(O_t = B|W_t = 0) = 0.1$

5. **Resample**: Using the random numbers, resample P_1 and P_2 based on the weights.

   Because both of our particles have $X = 0$, resampling will still leave us with two particles with $X = 0$.
   $P_1 = 0$
   $P_2 = 0$

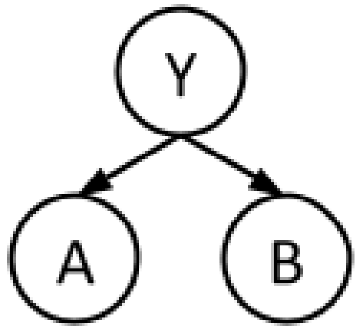
6. What is our estimated distribution for $P(W_2|O_1 = A, O_2 = B)$?

   $P(W_2 = 0|O_1 = A, O_2 = B) = 2/2 = 1$
   $P(W_2 = 1|O_1 = A, O_2 = B) = 0/2 = 0$

2 Naive Bayes

In this question, we will train a Naive Bayes classifier to predict class labels Y as a function of input features A and B. Y, A, and B are all binary variables, with domains 0 and 1. We are given 10 training points from which we will estimate our distribution.

A	1	1	1	1	0	1	0	1	1	1
B	1	0	0	1	1	1	1	0	1	1
Y	1	1	0	0	0	1	1	0	0	0



1. What are the maximum likelihood estimates for the tables $P(Y)$, $P(A|Y)$, and $P(B|Y)$?

Y	$P(Y)$
0	3/5
1	2/5

A	Y	$P(A\|Y)$
0	0	1/6
1	0	5/6
0	1	1/4
1	1	3/4

B	Y	$P(B\|Y)$
0	0	1/3
1	0	2/3
0	1	1/4
1	1	3/4

2. Consider a new data point ($A = 1$, $B = 1$). What label would this classifier assign to this sample?

$$\begin{align}
P(Y=0, A=1, B=1) &= P(Y=0)P(A=1|Y=0)P(B=1|Y=0) \tag{1}\\
&= (3/5)(5/6)(2/3) \tag{2}\\
&= 1/3 \tag{3}\\
P(Y=1, A=1, B=1) &= P(Y=1)P(A=1|Y=1)P(B=1|Y=1) \tag{4}\\
&= (2/5)(3/4)(3/4) \tag{5}\\
&= 9/40 \tag{6}\\
& \tag{7}
\end{align}$$

Our classifier will predict label 0.

3. Let's use Laplace Smoothing to smooth out our distribution. Compute the new distribution for $P(A|Y)$ given Laplace Smoothing with $k = 2$.

A	Y	$P(A\|Y)$
0	0	3/10
1	0	7/10
0	1	3/8
1	1	5/8