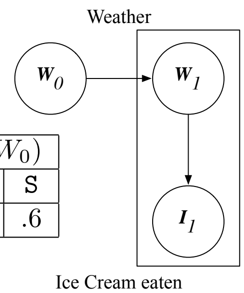
CS 188 Spring 2011	Introduction to Artificial Intelligence	Section 10 Sampling and Naive Bayes

Q1. Sampling and DBNs

The Dynamic Bayes Net (really just a Hidden Markov Model, or HMM) in the diagram to the right describes a person's ice-cream eating habits based on the weather. The nodes W_i stand for the weather on a day i, which can either be rainy R or sunny S. The nodes I_i represent whether or not the person ate ice-cream on day i, and the node takes values T (for truly eating ice cream) or F. The conditional probability distributions relevant to the graphical model are also given to you.



$\mathbb{P}(W_0)$	
R	S
.4	.6

W_0	$\mathbb{P}(W_1 \mid W_0)$	
	S	R
S	.7	.3
R	.5	.5

W_0	$\mathbb{P}(I \mid W)$	
	T	F
S	.9	.1
R	.2	.8

We will consider performing inference using sampling in the unrolled DBN over two time steps.

(a) Draw the unrolled DBN that we will use to perform inference.

Suppose we sample from the prior to produce the following samples of (W_1, I_1, W_2, I_2) from the ice-cream model:

R, F, R, F	R, F, R, F	S, F, S, T	S, T, S, T	S, T, R, F
R, F, R, T	S, T, S, T	S, T, S, T	S, T, R, F	R, F, S, T

(b) What is $\widehat{P}(W_2 = \texttt{R})$, the probability sampling assigns the event $W_2 = \texttt{R}$?

(c) Cross off samples rejected by rejection sampling if we're computing $\mathbb{P}(W_2 | I_1 = \texttt{T}, I_2 = \texttt{F})$

(d) Rejection sampling seems to be wasting a lot of effort, so we decide to switch to likelihood weighting. Assume we generate the following six samples given the evidence $I_1 = \texttt{T}$ and $I_2 = \texttt{F}$:

$$(W_1, I_1, W_2, I_2) = \Big\{\langle \texttt{S},\texttt{T},\texttt{R},\texttt{F}\rangle, \langle \texttt{R},\texttt{T},\texttt{R},\texttt{F}\rangle, \langle \texttt{S},\texttt{T},\texttt{R},\texttt{F}\rangle, \langle \texttt{S},\texttt{T},\texttt{S},\texttt{F}\rangle, \langle \texttt{S},\texttt{T},\texttt{S},\texttt{F}\rangle, \langle \texttt{R},\texttt{T},\texttt{S},\texttt{F}\rangle\Big\}$$

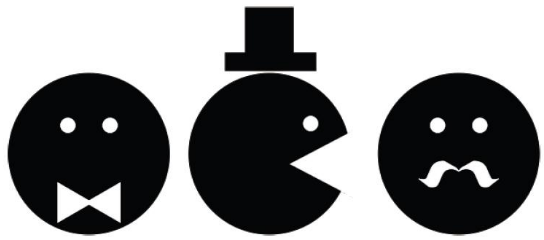
(i) What is the weight of the first sample (S, T, R, F) above?

(ii) Use likelihood weighting to estimate $\widehat{P}(W_2 | I_1 = \texttt{T}, I_2 = \texttt{F})$

Q2. Naive Pacbaby

Pacman and Ms. Pacman have been searching for each other in the Maze. Ms. Pacman has been pregnant with a baby, and just this morning (yesterday morning if you are in Lubomir's section, sorry for not telling you sooner) she gave birth to Pacbaby (Congratulations, Ms. Pacman!).

Because Pacbaby was born before Pacman and Ms. Pacman reunited in the Maze, Pacbaby has never met his father. Naturally, Ms. Pacman wants to teach Pacbaby to recognize his father, using a set of Polaroids of Pacman. She also has several pictures of ghosts to use as negative examples. Because the polaroids are black and white, and were taken from strange angles, Ms. Pacman has decided to teach Pacbaby to identify Pacman based on more salient features: the presence of a bowtie (b), hat (h), or mustache (m).



The table to the right summarizes the content of the Polaroids. Each binary feature is represented as 1 (meaning the feature is present) or 0 (it is absent). The subject y of the photo is encoded as +1 for Pacman or −1 for ghost.
For the remainder, suppose Pacbaby has a Naive Bayes based brain (poor Pacbaby).

(m)	(b)	(h)	Subject (y)
0	0	0	+1
1	0	0	+1
1	1	0	+1
0	1	1	+1
1	0	1	-1
1	1	1	-1

(a) Write the Naive Bayes classification rule for this problem (i.e. write a formula which given a data point $x = (m, b, h)$ returns the most likely subject y). Write the formula in terms of conditional and prior probabilities. Be explicit about which parameters are involved, but you do not need to estimate them yet.

(b) Assuming no smoothing, give estimates for the parameters of the classification rule based on the Polaroids.

	$y = +1$	$y = -1$
$P(y)$		

P	$y = +1$	$y = -1$
$P(m = 1\|y)$		
$P(b = 1\|y)$		
$P(h = 1\|y)$		

(c) A character comes by wearing a hat but without a mustache or bowtie. What will Pacbaby guess as the identity of the character?

(d) Pacbaby is actually born with a bit of prior information, so let's suppose that Pacbaby performs Laplace smoothing with strength $k = 1$ (on both the prior and class-conditional parameters). Reestimate the parameters. Now how will Pacbaby classify this new character with the hat and without a mustache or bowtie?

	$y = +1$	$y = -1$
$P(y)$		

P	$y = +1$	$y = -1$
$P(m = 1\|y)$		
$P(b = 1\|y)$		
$P(h = 1\|y)$		