Q1. Naive Bayes: Pacman or Ghost?

You are standing by an exit as either Pacmen or ghosts come out of it. Every time someone comes out, you get two observations: a visual one and an auditory one, denoted by the random variables X_v and X_a, respectively. The visual observation informs you that the individual is either a Pacman ($X_v = 1$) or a ghost ($X_v = 0$). The auditory observation X_a is defined analogously. Your observations are a noisy measurement of the individual's true type, which is denoted by Y. After the indiviual comes out, you find out what they really are: either a Pacman ($Y = 1$) or a ghost ($Y = 0$). You have logged your observations and the true types of the first 20 individuals:

individual i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
first observation $X_v^{(i)}$	0	0	1	0	1	0	0	1	1	1	0	1	1	0	1	1	1	0	0	0
second observation $X_a^{(i)}$	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
individual's type $Y^{(i)}$	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0

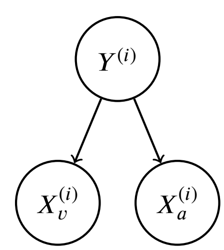
The superscript (i) denotes that the datum is the ith one. Now, the individual with $i = 20$ comes out, and you want to predict the individual's type $Y^{(20)}$ given that you observed $X_v^{(20)} = 1$ and $X_a^{(20)} = 1$.

(a) Assume that the types are independent, and that the observations are independent conditioned on the type. You can model this using naïve Bayes, with $X_v^{(i)}$ and $X_a^{(i)}$ as the features and $Y^{(i)}$ as the labels. Assume the probability distributions take on the following form:

$$P(X_v^{(i)} = x_v | Y^{(i)} = y) = \begin{cases} p_v & \text{if } x_v = y \\ 1 - p_v & \text{if } x_v \neq y \end{cases}$$

$$P(X_a^{(i)} = x_a | Y^{(i)} = y) = \begin{cases} p_a & \text{if } x_a = y \\ 1 - p_a & \text{if } x_a \neq y \end{cases}$$

$$P(Y^{(i)} = 1) = q$$



for $p_v, p_a, q \in [0, 1]$ and $i \in \mathbb{N}$.

(i) What's the maximum likelihood estimate of p_v, p_a and q?

$p_v =$ ____________________ $p_a =$ ____________________ $q =$ ____________________

(ii) What is the probability that the next individual is Pacman given your observations? Express your answer in terms of the parameters p_v, p_a and q (you might not need all of them).

$P(Y^{(20)} = 1 | X_v^{(20)} = 1, X_a^{(20)} = 1) =$ ______________________________

Now, assume that you are given additional information: you are told that the individuals are actually coming out of a bus that just arrived, and each bus carries *exactly* 9 individuals. Unlike before, the types of every 9 consecutive individuals are *conditionally* independent given the bus type, which is denoted by Z. Only after all of the 9 individuals have walked out, you find out the bus type: one that carries mostly Pacmans ($Z = 1$) or one that carries mostly ghosts ($Z = 0$). Thus, you only know the bus type in which the first 18 individuals came in:

individual i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
first observation $X_v^{(i)}$	0	0	1	0	1	0	0	1	1	1	0	1	1	0	1	1	1	0	0	0
second observation $X_a^{(i)}$	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
individual's type $Y^{(i)}$	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0

bus j	0	1	
bus type $Z^{(j)}$	0	1	

(b) You can model this using a variant of naïve bayes, where now 9 consecutive labels $Y^{(i)}, \dots, Y^{(i+8)}$ are *conditionally* independent given the bus type $Z^{(j)}$, for bus j and individual $i = 9j$. Assume the probability distributions take on the following form:

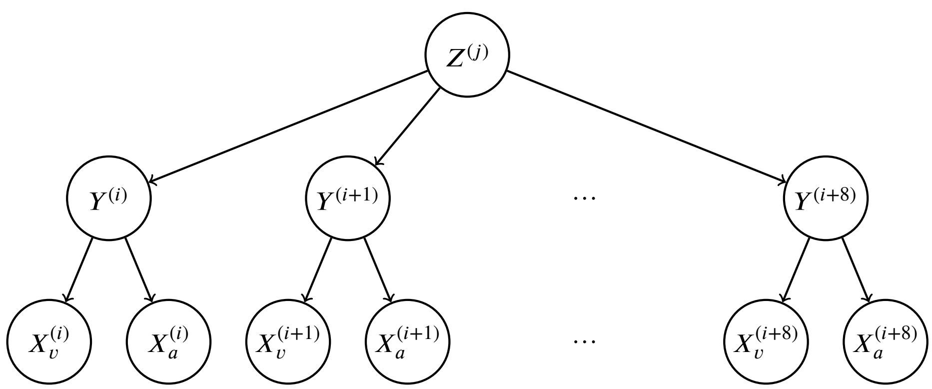
$$P(X_v^{(i)} = x_v | Y^{(i)} = y) = \begin{cases} p_v & \text{if } x_v = y \\ 1 - p_v & \text{if } x_v \neq y \end{cases}$$

$$P(X_a^{(i)} = x_a | Y^{(i)} = y) = \begin{cases} p_a & \text{if } x_a = y \\ 1 - p_a & \text{if } x_a \neq y \end{cases}$$

$$P(Y^{(i)} = 1 | Z^{(j)} = z) = \begin{cases} q_0 & \text{if } z = 0 \\ q_1 & \text{if } z = 1 \end{cases}$$

$$P(Z^{(j)} = 1) = r$$

for $p, q_0, q_1, r \in [0, 1]$ and $i, j \in \mathbb{N}$.



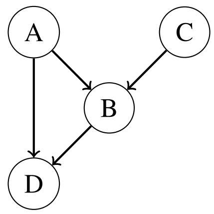
(i) What's the maximum likelihood estimate of q_0, q_1 and r?

$q_0 =$ ______________________ $q_1 =$ ______________________ $r =$ ______________________

(ii) Compute the following joint probability. Simplify your answer as much as possible and express it in terms of the parameters p_v, p_a, q_0, q_1 and r (you might not need all of them).

$P(Y^{(20)} = 1, X_v^{(20)} = 1, X_a^{(20)} = 1, Y^{(19)} = 1, Y^{(18)} = 1) =$ ______________________________________

Q2. Bayes Nets: Representation



Consider the Bayes net graph depicted above.

(a) **(i)** Select all conditional independences that are enforced by this Bayes net graph.

- ☐ $A \perp\!\!\!\perp B$
- ☐ $A \perp\!\!\!\perp C \mid B$
- ☐ $D \perp\!\!\!\perp C \mid A, B$
- ☐ $D \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp C$
- ☐ $A \perp\!\!\!\perp C \mid D$
- ☐ $A \perp\!\!\!\perp C \mid B, D$
- ☐ $D \perp\!\!\!\perp C \mid B$

(ii) Because of these conditional independences, there are some distributions that cannot be represented by this Bayes net. What is the minimum set of edges that would need to be added such that the resulting Bayes net could represent any distribution?

- ☐ $A \rightarrow C$
- ☐ $C \rightarrow A$
- ☐ $C \rightarrow D$
- ☐ $D \rightarrow C$
- ☐ $D \rightarrow A$
- ☐ $D \rightarrow B$
- ☐ $B \rightarrow C$
- ☐ $B \rightarrow A$

(b) For the rest of this Q2, we use the **original, unmodified** Bayes net depicted at the beginning of the problem statement. Here are some partially-filled conditional probability tables on A, B, C, and D. Note that these are not necessarily factors of the Bayes net. Fill in the six blank entries such that this distribution can be represented by the Bayes net.

A	C	$P(C \mid A)$
$+a$	$+c$	0.8
$+a$	$-c$	0.2
$-a$	$+c$	0.8
$-a$	$-c$	0.2

A	B	D	$P(D \mid A, B)$
$+a$	$+b$	$+d$	0.60
$+a$	$+b$	$-d$	0.40
$+a$	$-b$	$+d$	0.10
$+a$	$-b$	$-d$	0.90
$-a$	$+b$	$+d$	0.20
$-a$	$+b$	$-d$	0.80
$-a$	$-b$	$+d$	0.50
$-a$	$-b$	$-d$	0.50

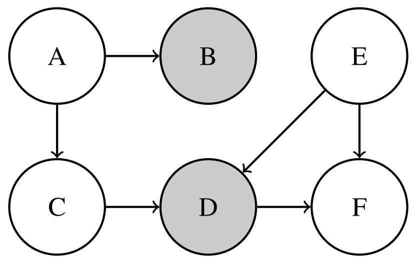
A	B	C	$P(C \mid A, B)$
$+a$	$+b$	$+c$	0.50
$+a$	$+b$	$-c$	0.50
$+a$	$-b$	$+c$	0.20
$+a$	$-b$	$-c$	0.80
$-a$	$+b$	$+c$	0.90
$-a$	$+b$	$-c$	0.10
$-a$	$-b$	$+c$	0.40
$-a$	$-b$	$-c$	0.60

C	$P(C)$
$+c$	**(i)**
$-c$	**(ii)**

A	B	C	D	$P(D, C \mid A, B)$
$+a$	$+b$	$+c$	$+d$	**(iii)**
$+a$	$+b$	$-c$	$-d$	**(iv)**
$+a$	$-b$	$+c$	$+d$	**(v)**
$+a$	$-b$	$-c$	$-d$	**(vi)**
⋮	⋮	⋮	⋮	⋮

Q3. Bayes Nets and Sampling

You are given a bayes net with the following probability tables:



E	D	F	$P(F\|E,D)$
0	0	0	0.6
0	0	1	0.4
0	1	0	0.7
0	1	1	0.3
1	0	0	0.2
1	0	1	0.8
1	1	0	0.7
1	1	1	0.3

A	$P(A)$
0	0.75
1	0.25

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

A	C	$P(C\|A)$
0	0	0.3
0	1	0.7
1	0	0.7
1	1	0.3

E	$P(E)$
0	0.1
1	0.9

E	C	D	$P(D\|E,C)$
0	0	0	0.5
0	0	1	0.5
0	1	0	0.2
0	1	1	0.8
1	0	0	0.5
1	0	1	0.5
1	1	0	0.2
1	1	1	0.8

You want to know $P(C = 0|B = 1, D = 0)$ and decide to use sampling to approximate it.

(a) With prior sampling, what would be the likelihood of obtaining the sample [A=1, B=0, C=0, D=0, E=1, F=0]?

- ○ 0.25*0.1*0.3*0.9*0.8*0.7
- ○ 0.75*0.1*0.3*0.9*0.5*0.8
- ○ 0.25*0.9*0.7*0.1*0.5*0.6
- ○ 0.25*0.5*0.7*0.5*0.9*0.2
- ○ 0.25*0.5*0.3*0.2*0.9*0.2
- ○ 0.75*0.1*0.3*0.9*0.5*0.2 + 0.25*0.5*0.7*0.5*0.9*0.2

○ Other ____________________

(b) Assume you obtained the sample [A = 1, B=1, C=0, D=0, E=1, F=1] through likelihood weighting. What is its weight?

- ○ 0.25*0.5*0.7*0.5*0.9*0.8
- ○ 0.25*0.7*0.9*0.8 + 0.75*0.3*0.9*0.8
- ○ 0.25*0.5*0.7*0.5*0.8
- ○ 0
- ○ 0.5*0.5
- ○ 0.9*0.5 + 0.1*0.5

○ Other ____________________

(c) You decide to use Gibb's sampling instead. Starting with the initialization [A = 1, B=1, C=0, D=0, E=0, F=0], suppose you resample F first, what is the probability that the next sample drawn is [A = 1, B=1, C=0, D=0, E=0, F=1]?

- ○ 0.4
- ○ 0.6*0.1*0.5
- ○ 0.25*0.5*0.7*0.5*0.1*0.3
- ○ 0.6
- ○ 0
- ○ 0.9*0.5 + 0.1*0.5

○ Other ____________________