Assignment 3: Uncertainty and Probability

Ritesh Patel

300362131

Part 1: Reasoning Under Uncertainty Basics

X	*P*(X)
0	0.300
1	0.700

Y	*X*	*P*(*Y*|*X*)
0	0	0.300
1	0	0.700
0	1	0.800
1	1	0.200

Z	*Y*	*P*(*Z*|*Y*)
0	0	0.600
1	0	0.400
0	1	0.800
1	1	0.200

1. Create the full joint probability table of *X* and *Y*, i.e. the table containing the following four joint probabilities *P*(X= 0, Y= 0), *P*(X= 0, Y= 1), *P*(X= 1, Y= 0), *P*(X= 1, Y= 1). Also explain which probability rules you used.

X	*Y*	*P(X,Y)*	=
0	0	*P*(X= 0, Y= 0) => *P*(X= 0) × *P*(Y= 0|X= 0)	0.300 × 0.300 = 0.09
0	1	*P*(X= 0, Y= 1) => *P*(X= 0) × *P*(Y= 1|X= 0)	0.300 × 0.700 = 0.21
1	0	*P*(X= 1, Y= 0) => *P*(X= 1) × *P*(Y= 0|X= 1)	0.700 × 0.800 = 0.56
1	1	*P*(X= 1, Y= 1) => *P*(X= 1) × *P*(Y= 1|X= 1)	0.700 × 0.200 = 0.14

I have used the Product Rule since we are using joint probabilities where:

P(X,Y) = P(X) × P(Y|X)

2. If given *P*(X= 1, Y= 0, Z= 0) = 0.336, *P*(X= 0, Y= 1, Z= 0) = 0.168, *P*(X= 0, Y=0, Z= 1) = 0.036, and *P*(X= 0, Y= 1, Z= 1) = 0.042, create the full joint probability table of the three variables X, Y, and Z. Also explain which probability rules you used.

X	*Y*	*Z*	*P(X,Y,Z)*	=
0	0	0	*P*(X= 0, Y= 0, Z= 0) => *P*(X= 0, Y= 0) × *P*(Z= 0|Y= 0)	0.09 × 0.600 = 0.054
0	0	1	*P*(X= 0, Y= 0, Z= 1) => *P*(X= 0, Y= 0) × *P*(Z= 1|Y= 0)	0.09 × 0.400 = 0.036
0	1	0	*P*(X= 0, Y= 1, Z= 0) => *P*(X= 0, Y= 1) × *P*(Z= 0|Y= 1)	0.21× 0.800 = 0.168
0	1	1	*P*(X= 0, Y= 1, Z= 1) => *P*(X= 0, Y= 1) × *P*(Z= 1|Y= 1)	0.21 × 0.200 = 0.042
1	0	0	*P*(X= 1, Y= 0, Z= 0) => *P*(X= 1, Y= 0) × *P*(Z= 0|Y= 0)	0.56 × 0.600 = 0.336
1	0	1	*P*(X= 1, Y= 0, Z= 1) => *P*(X= 1, Y= 0) × *P*(Z= 1|Y= 0)	0.56 × 0.400 = 0.224
1	1	0	*P*(X= 1, Y= 1, Z= 0) => *P*(X= 1, Y= 1) × *P*(Z= 0|Y= 1)	0.14 × 0.800 = 0.112
1	1	1	*P*(X= 1, Y= 1, Z= 1) => *P*(X= 1, Y= 1) × *P*(Z= 1|Y= 1)	0.14 × 0.200 = 0.028

I have used the Product Rule and the independence rule where:

P(X,Y,Z) = P(X,Y) × P(Z|(X,Y)) Product Rule

Since Z is independent from X|Y we use the conditionally independent rule.

=> P(X,Y,Z) = P(X,Y) × P(Z|Y)

3. From the above joint probability table of X, Y, and Z:
   (i) calculate the probability of *P*(Z= 0) and *P*(X= 0, Z= 0)

   P(Z = 0) = P(X= 0, Y= 0, Z= 0) + P(X= 0, Y= 1, Z= 0) + P(X= 1, Y= 0, Z= 0) + P(X= 1, Y= 1, Z= 0)
   Sum Rule => 0.054 + 0.168 + 0.336 + 0.112 = 0.67

   P(X= 0, Z= 0) = P(X= 0, Y= 0, Z= 0) + P(X= 0, Y=1, Z= 0)
   Sum Rule => 0.054 + 0.168 = 0.222

   (ii) judge whether X and Z are independent to each other and explain why.

   Assuming X and Z are independent, => P(X,Z) = P(X) × P(Z)
   TEST: P(X= 0, Z= 0) = P(X= 0) × P(Z= 0)
   P(X= 0, Z= 0) = 0.222
   P(Z = 0) = 0.67

   0.3 × 0.67 = 0.201
   => 0.201 != 0.222

   This proves that Z and X are not independent to each other.

4. From the above joint probability table of X, Y, and Z:
   (i) calculate the probability of *P*(X= 1, Y= 0|Z= 1),

   P(X= 1, Y= 0|Z= 1) = P(X= 1, Y= 0, Z= 1) / P(X= 0, Y= 0, Z= 1) + P(X= 0, Y= 1, Z= 1) + P(X= 1, Y= 0, Z= 1) + P(X= 1, Y= 1, Z= 1)
   = (0.224 / (0.036 + 0.042 + 0.224 + 0.028)) = 0.679

   (ii) calculate the probability of *P*(X= 0|Y= 0, Z= 0).

   P(X= 0|Y= 0, Z= 0) = P(X= 0, Y= 0, Z= 0) / ((P(X= 0, Y= 0, Z= 0) + P(X= 1, Y= 0, Z= 0))
   =(0.054 / (0.054 + 0.336)) = 0.138

Part 2: Naive Bayes Method

1. the probabilities $P(F_i|c)$ for each feature $_i$.

Feature	Spam	Non-Spam
1 - TRUE	35/53	54/151
1 - FALSE	18/53	97/151
2 - TRUE	31/53	87/151
2 - FALSE	22/53	64/151
3 - TRUE	24/53	52/151
3 - FALSE	29/53	99/151
4 - TRUE	32/53	60/151
4 - FALSE	21/53	91/151
5 - TRUE	26/53	51/151
5 - FALSE	27/53	100/151
6 - TRUE	19/53	71/151
6 - FALSE	34/53	80/151
7 - TRUE	41/53	76/151
7 - FALSE	12/53	75/151
8 - TRUE	40/53	53/151
8 - FALSE	13/53	98/151
9 - TRUE	18/53	37/151
9 - FALSE	35/53	114/151
10 - TRUE	35/53	44/151
10 - FALSE	18/53	107/151
11 - TRUE	35/53	88/151
11 - FALSE	18/53	63/151
12 - TRUE	41/53	51/151
12 - FALSE	12/53	100/151

2. For each instance in the unlabelled set, given the input vector *F*, the probability *P*(*S*|*D*), the probability *P*($\bar{S}$|*D*), and the predicted class of the input vector. Here D is an email represented by *F, S* refers to class spam and $\bar{S}$ refers to class non-spam.

Vector	*P*(S|D)	*P*(>S|D)	Prediction
F1 – 1 1 0 0 1 1 0 0 0 0 0 0	3.6310682971325263E-6	4.5745793354799594E-4	False
F2 – 0 0 1 1 0 0 1 1 1 0 0 1	5.73701316149022E-5	4.1956427373119854E-5	True
F3 – 1 1 1 1 1 0 1 0 0 0 1 1	1.8599635002963176E-4	1.2886216893902397E-4	True
F4 – 0 1 0 0 0 0 1 0 1 0 0 0	6.097638620708581E-6	5.970900059679335E-4	False
F5 – 1 1 1 0 1 1 0 1 0 0 1 1	6.142744481444356E-5	9.257237491670748E-5	False
F6 – 1 1 1 1 1 1 0 0 0 1 1 1	5.915235426576045E-5	4.6409871775721404E-5	True
F7 – 0 0 0 0 1 1 0 1 0 0 0 0	4.077711579339186E-6	3.269186174049223E-4	False
F8 – 0 1 0 1 1 1 1 0 0 0 1 1	6.459074928550027E-5	3.9111423607025793E-4	False
F9 – 1 1 1 1 1 0 1 0 0 1 0 1	1.8599635002963176E-4	3.793605908017997E-5	True
F10 – 1 1 0 0 0 1 0 1 0 0 1 0	2.2559891786918056E-5	6.775992308249818E-4	False

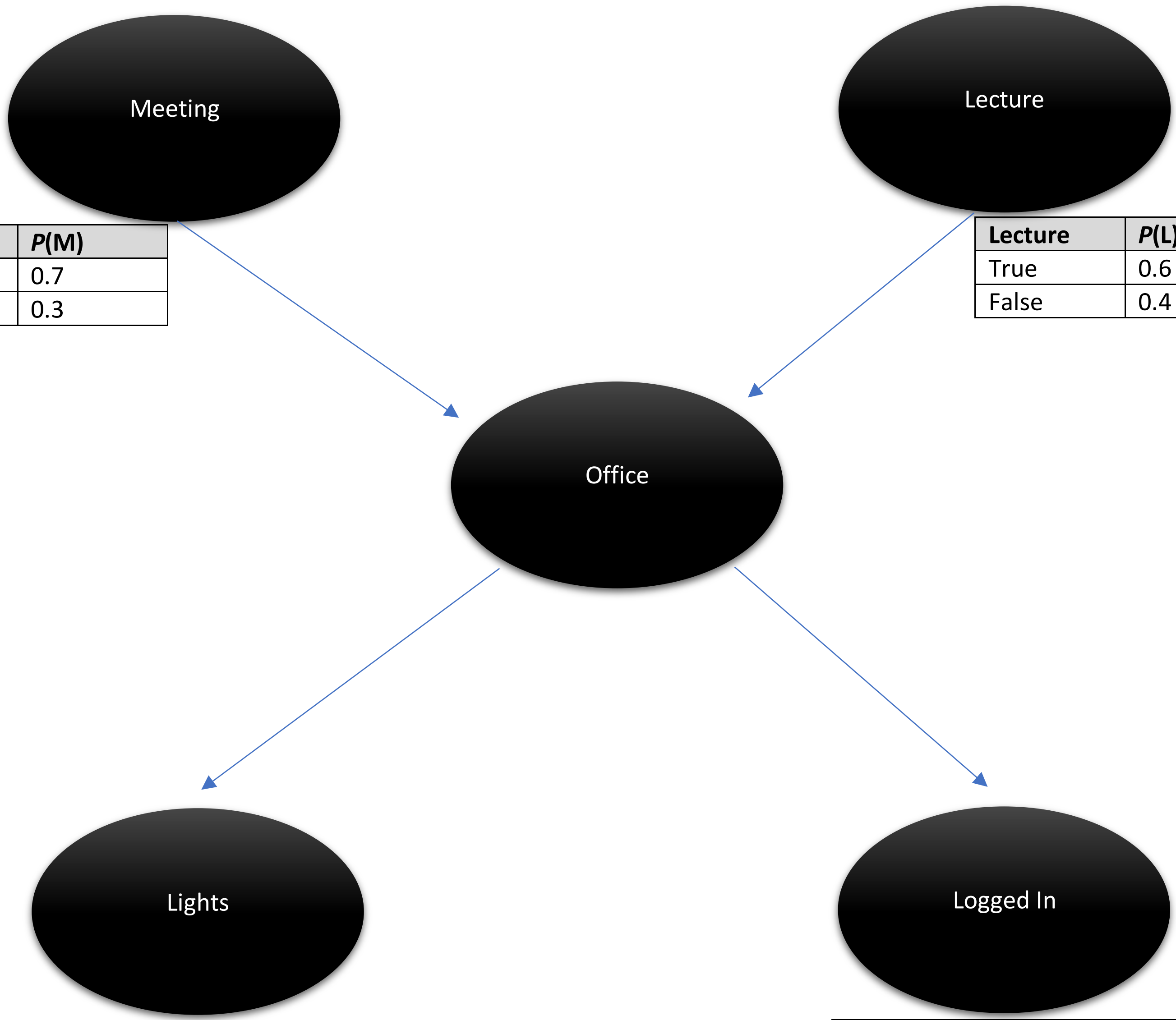
3. The derivation of the Naive Bayes algorithm assumes that the attributes are conditionally independent. Why is this like to be an invalid assumption for the spam data? Discuss the possible effect of two attributes not being independent.

This is an invalid assumption for spam data since the data contains attributes that are not always conditionally independent. Words in an email can be conditionally dependent on each other. E.g. A spam email containing the term "Viagra" would be likely to have the term "pills" as well, but it would be unlikely for it to have the term "MILLION DOLLARS". The effect of two attributes not being independent is that a dependent attribute is affected by the outcome of the attribute it depends on. This can result in different probability results as to the results that are not dependant.

Part 3: Bayesian Networks

1. Construct a Bayesian network to represent the above scenario. *(Hint: First decide what your domain variables are; these will be your network nodes. Then decide what the causal relationships are between the domain variables and add directed arcs in the network from cause to affect. Finally, you have to add the prior probabilities for nodes without parents, and the conditional probabilities for nodes that have parents.)*

Ritesh Patel 300362131 patelrite



Meeting	***P*(M)**
True	0.7
False	0.3

Lecture	***P*(L)**
True	0.6
False	0.4

Lights	**O**	***P*(Li)**
True	True	0.5
False	True	0.5
True	False	0.02
False	False	0.98

Logged In	**O**	***P*(Lo)**
True	True	0.8
False	True	0.2
True	False	0.2
False	False	0.8

Office	**M**	**L**	***P*(O|M,L)**
True	True	True	0.95
True	True	False	0.75
True	False	True	0.8
True	False	False	0.06
False	True	True	0.05
False	True	False	0.25
False	False	True	0.2
False	False	False	0.94

2. Calculate how many free parameters in your Bayesian network?

Total Conditional Dependency Table = (L - 1) + (M - 1) + ((M × L × (O - 1)) + (O × (Li – 1)) + (O × (Lo – 1)) =

  (2 – 1) = 1
+ (2 – 1) = 1
+ (2 × 2 × (2 - 1)) = 4
+ (2 × (2 – 1)) = 2
+ (2 × (2 – 1)) = 2

1 + 1 + 4 + 2 + 2 = 10 free parameters

3. What is the joint probability that Rachel has lectures, has no meetings, she is in her office and logged on her computer but with lights off?

P(L = true, M = false, O = true, Lo = true, Li = false)

= P(O =|M,L) × P(Lo, Li | O)

= 0.8 × 0.8 × 0.5

= 0.32

I got a different result calculating in an expanded way show below.

P(L = true, M = false, O = true, Lo = true, Li = false)

'¬' = False, else true.

P(L) × P(¬M) × P(O | L, ¬M) × P(Lo | O) × P(¬Li | O)

= 0.6 × 0.3 × 0.8 × 0.8 × 0.5 = 0.0576

4. Calculate the probability that Rachel is in the office.

P(O = true)

'¬' = False, else true.

= P(O, M, L) + P(O, M, ¬L) + P(O, ¬M, L) + P(O, ¬M, ¬L)

P(O | M, L) × P(M, L)	= 0.95 × 0.42 = 0.399
+ P(O, M, ¬L) × P(M , ¬L)	= 0.75 × 0.28 = 0.21
+ P(O, ¬M, L) × P(¬M, L)	= 0.8 × 0.18 = 0.144
+ P(O, ¬M, ¬L) × P(¬M, ¬L)	= 0.06 × 0.12 = 0.0072
	= 0.7602

5. If Rachel is in the office, what is the probability that she is logged on, but her light is off.

P(Li = false, Lo = true | O = true)

'¬' = False, else true.

= P(¬Li | O) × P(Lo | O)

= 0.5 × 0.8 = 0.4

6. Suppose a student checks Rachel's login status and sees that she is logged on. What effect does this have on the student's belief that Rachels light is on?

"Common cause: Effect becomes independent once common cause is known."
Rachel's login status has no effect on the student's belief that Rachel's light is on since the probability of when Rachel is in her office is known. This means that the probability of Rachel's light and Login status becomes independent.

Part 4: Inference in Bayesian Networks

1. Using inference by enumeration to calculate the probability $P(P{=}t|X{=}t)$
   (i) describe what are the evidence, hidden and query variables in this inference,

Evidence Variables: Xray = True (Evidence variable = Observed)
Hidden: Dyspnoea, Cancer, Smoker (Hidden = Unobserved)
Query: Pollution = true. (P= t) (Query = Observation – what we want to find out)

   (i) describe how would you use variable elimination in this inference, i.e. to perform the join operation and the elimination operation on which variables and in what order and report the probability.

1. Joint Distribution

$P(P{=}\ t \mid X{=}\ t) = P(X, t) / P(t) = \alpha P(X, t)$
$= \sum_{S,C,D} P(X, t, S, C, D)$

2. Joint probability using Bayes net factors
   $= \sum_{S,C,D} P(S)\ P(C \mid P, S)\ P(X \mid C)\ P(D \mid C)$
3. Variable order and Summations

$= P(P) \sum_{S} P(S) \sum_{C} P(C \mid P, S)\ P(X \mid C) \sum_{D} P(D \mid C)$

4. Factor

$= P(\mathrm{P}) \sum_{s} P(\mathrm{S}) \sum_{c} P(\mathrm{C} \mid \mathrm{P}, \mathrm{S})\ P(\mathrm{X} \mid \mathrm{C})\ f_1(\mathrm{C})$

$f_1(\mathrm{C})$

C	P(D | C)
True	0.65
False	0.3

$= P(\mathrm{P}) \sum_{s} P(\mathrm{S})\ f_2(\mathrm{P}, \mathrm{S})$

$f_2(\mathrm{P}, \mathrm{S})$

P	S	P(C | P, S) P(X | C)
True	True	$((0.05 \times 0.9) + (0.95 \times 0.2)) = 0.235$
True	False	$((0.02 \times 0.9) + (0.98 \times 0.2)) = 0.214$
False	True	$((0.03 \times 0.9) + (0.97 \times 0.2)) = 0.221$
False	False	$((0.001 \times 0.9) + (0.999 \times 0.2)) = 0.2007$

$= \alpha\ P(\mathrm{P})\ f_3 P(\mathrm{P})$

$f_3(\mathrm{P})$

P	P(S)
True	$((0.9 \times 0.235) + (0.1 \times 0.214)) = 0.2329$
False	$((0.9 \times 0.221) + (0.1 \times 0.2007)) = 0.21897$

$\alpha = (1 / ((P(P = t)\ f_3(P= t)) + (P(P= f)\ f_3(P= f))))$

$\Rightarrow \alpha = (1 / ((0.9 * 0.2329)) + (0.1 * 0.21897)))) = 4.31952381569$

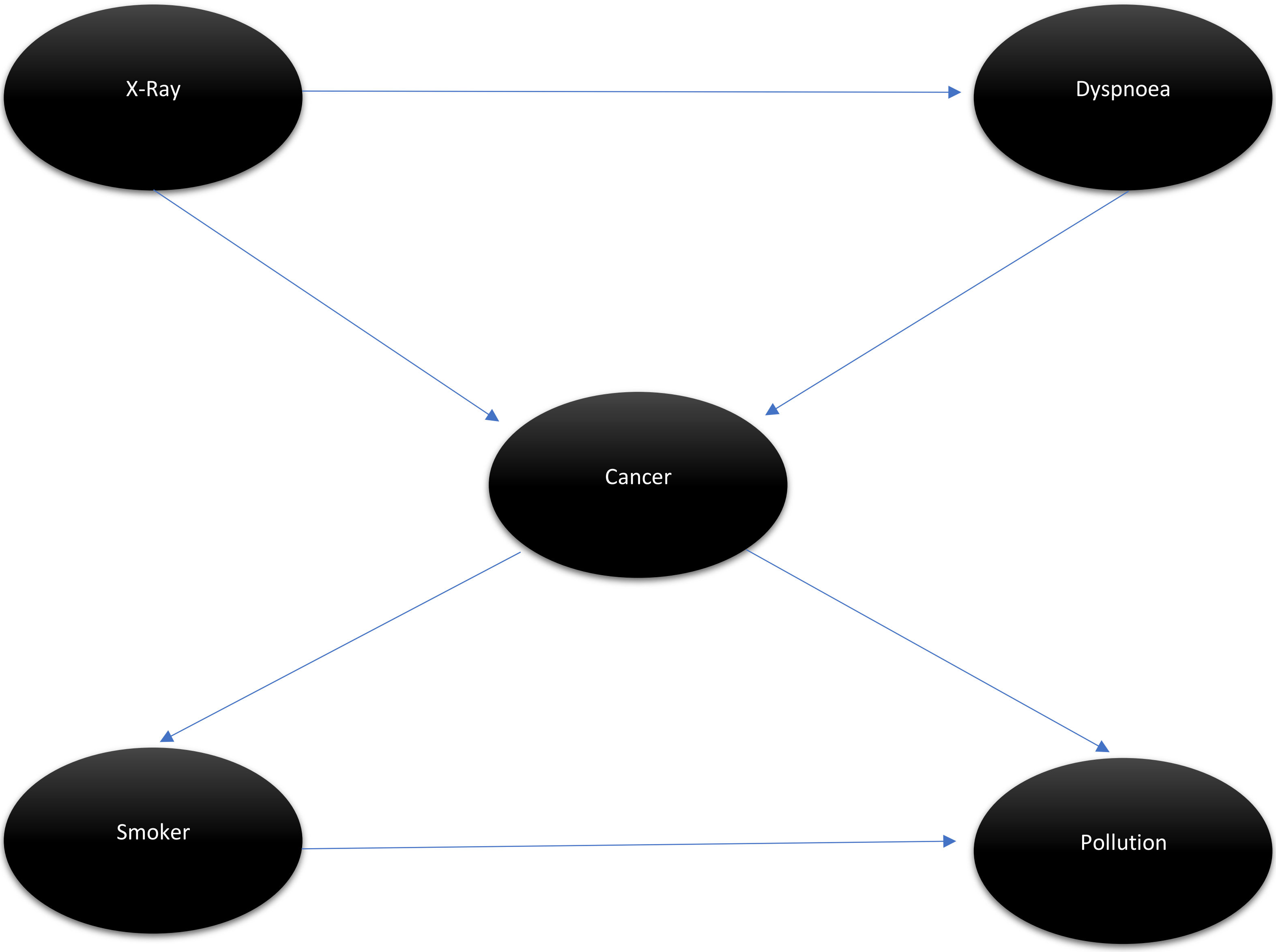
$P(P= t \mid X= t) = \alpha * 0.9 * 0.2329$

$\alpha = 4.31952381569 * 0.9 * 0.2329 = 0.905415387$

2. Given the Bayesian Network, find the variables that are independent of each other or conditionally independent given another variable. Find at least three pairs or groups of such variables.

a. Pollution & Smoker are independent of each other.
b. X-Ray & Dyspnoea are conditionally independent given Cancer.
c. X-Ray & Pollution are conditionally independent given Cancer.
d. X-Ray & Dyspnoea are conditionally independent of Pollution & Smoker given Cancer.

3. If given the variable order as <Xray, Dyspnoea, Cancer, Smoker, Pollution>, draw a new Bayesian Network structure (nodes and connections only) to describe the same problem/domain as shown in the above given Bayesian Network. *[hint: considering the above (conditionally)independent variables, the network should keep the original dependence between variables, which are that (conditionally) independent variables should remain being independent of each other, and dependent variables remain being dependent]*. For each connection, explain why it is needed.



<X-Ray, Dyspnoea, Cancer, Smoker, Pollution>

1. Add first node X-Ray
2. Add node Dyspnoea
3. P(D | X) = P(D)? No X -> D

4. Add node Cancer
5. P(C | X, D) = P(C)? No
   P(C | X, D) = P(C | D)? No
   P(C | X, D) = P (C | X)? No, X -> C & D -> C
6. Add node Smoker
7. P(S | X, D, C) = P(S)? No
   P(S | X, D ,C) = P(S)? Yes, link from C -> S, no other link
8. Add node Pollution
9. P(P | X, D ,C, S) = P(P)?
   P(P | X, D ,C, S) = P(P | C)? No
   P(P | X, D ,C, S) = P(P | S)? No
   P(P | X, D ,C, S) = P(P | C, S)? Yes, link from C -> P, S -> P, no other link