2019

Assignment 3

COMP307

ZANE RAWSON

Table of Contents

Part 1: Reasoning Under Uncertainty Basics

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.

Using the Product Rule: P(X, Y) = P(X) * P(Y|X)

P(X=0, Y=0) = P(X=0) * P(Y=0|X=0) = 0.300*0.300 = 0.090
P(X=1, Y=0) = P(X=1) * P(Y=0|X=1) = 0.700*0.800 = 0.560
P(X=0, Y=1) = P(X=0) * P(Y=1|X=0) = 0.300*0.700 = 0.210
P(X=1, Y=1) = P(X=1) * P(Y=1|X=1) = 0.700*0.200 = 0.140

X	Y	P(X,Y)
0	0	0.090
1	0	0.560
0	1	0.210
1	1	0.140

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.

Using the Product rule as used above and the conditionally independent rule: P(Y, X|Z) = P(Y|Z)*P(X|Z) <- P(A|B,C) = P(A|C)

Using the product Rule P(Z,X,Y) = P(Z|X,Y)*P(X,Y)
Given Y, Z is independent from X and is conditionally independent so P(Z,X,Y) = P(Z|Y)*P(X,Y)

P(Z=0, X = 0, Y = 0) = P(Z=0|Y=0)*P(X=0,Y=0) = 0.6*0.09 = 0.054
P(Z=0, X = 0, Y = 1) = P(Z=0|Y=1)*P(X=0,Y=1) = 0.168
P(Z=0, X = 1, Y = 1) = P(Z=0|Y=1)*P(X=1,Y=1) =0.8*0.14 = 0.112
P(Z=0, X = 1, Y = 0) = P(Z=0|Y=0)*P(X=1,Y=0) =0.336
P(Z=1, X = 0, Y = 0) = P(Z=1|Y=0)*P(X=0,Y=0) =0.036
P(Z=1, X = 0, Y = 1) = P(Z=1|Y=1)*P(X=0,Y=1) =0.042
P(Z=1, X = 1, Y = 0) = P(Z=1|Y=0)*P(X=1,Y=0) =0.4*0.56 = 0.224
P(Z=1, X = 1, Y = 1) = P(Z=1|Y=1)*P(X=1,Y=1) =0.2*0.14 = 0.028

Z	X	Y	P(Z|Y)	P(X,Y)	P(Z,X,Y)
0	0	0	0.6	0.9	0.054
0	0	1			0.168
0	1	1	0.8	0.14	0.112
0	1	0			0.336
1	0	0			0.036
1	0	1			0.042
1	1	0	0.4	0.56	0.224
1	1	1	0.2	0.14	0.028

From the above joint probability table of X, Y, and Z:

calculate the probability of P(Z = 0) and P(X = 0, Z = 0),

Using the above table to find both P(Z=0) and P(X=0,Z=0)

P(Z=0) = P(Z=0,X=0,Y=0) + P(Z=0,X=0,Y=1) + P(Z=0,X=1,Y=0) + P(Z=0,X=1,Y=1)
= 0.054 + 0.168 + 0.336 + 0.112 = 0.67
P(X=0,Z=0) = P(Z=0,X=0,Y=0) + P(Z=0,X=0,Y=1) = 0.054 + 0.168 = 0.222

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

If Z and X are independent of each other then P(X,Z) = P(X) * P(Z)

Using what we looked at previously where X=0 and Z=0:

P(X,Z) = 0.222

P(Z=0) = 0.67

P(X=0) = P(X=0,Y=0,Z=0) + P(X=0,Y=1,Z=0) + P(X=0,Y=0,Z=1) + P(X=0,Y=1,Z=1)
= 0.054 + 0.168 + 0.036 + 0.042 = 0.300

P(X) * P(Z) = 0.300 * 0.670 = 0.201

0.201 =/= 0.222 therefore P(X,Z) =/= P(X) * P(Z) meaning that they are not independent of each other

From the above joint probability table of X, Y, and Z:

calculate the probability of P(X = 1, Y = 0|Z = 1),

Using Bayes rule

P(X=1, Y=0|Z=1) = P(X|Z,Y) * P(X,Y)/P(Z)

Given that Z is independent of X given Y P(X|Y, Z) = P(Z|Y)

So P(X=1, Y=0|Z=1) = P(Z=1|Y=0) * P(X=1,Y=0)/P(Z=1)
= 0.400 * 0.560/(1-P(Z=0)) = 0.400 * 0.560/0.330
= 0.679

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)
=0.161

Part 2: Naïve Bayes Method

the probabilities P(Fi |c) for each feature i.

P(Feature(F) n | Category(C)) =

P(F1=0 | non-spam) = 0.6423841059, P(F1=0 | spam) = 0.3396226415

P(F1=1 | non-spam) = 0.357615894, P(F1=1 |spam) = 0.660377359

P(F2=0 | non-spam) =0.4238410596, P(F2=0 |spam) = 0.4150943396

P(F2=1 | non-spam) =0.57615894, P(F2=1 |spam) =0.58490566

P(F3=0 | non-spam) =0.6556291390, P(F3=0 |spam) =0.5471698113

P(F3=1 | non-spam) =0.344370861, P(F3=1 |spam) =0.452830189

P(F4=0 | non-spam) =0.6026490066, P(F4=0 |spam) =0.3962264150

P(F4=1 | non-spam) =0.397350993, P(F4=1 |spam) =0.603773585

P(F5=0 | non-spam) =0.6622516556, P(F5=0 |spam) =0.5094339622

P(F5=1 | non-spam) =0.337748344, P(F5=1 |spam) =0.490566038

P(F6=0 | non-spam) =0.5298013245, P(F6=0 |spam) =0.6415094339

P(F6=1 | non-spam) =0.470198676, P(F6=1 |spam) =0.358490566

P(F7=0 | non-spam) =0.4966887417, P(F7=0 |spam) =0.2264150943

P(F7=1 | non-spam) =0.503311258, P(F7=1 |spam) =0.773584906

P(F8=0 | non-spam) =0.6490066225, P(F8=0 |spam) =0.2452830188

P(F8=1 | non-spam) =0.350993378, P(F8=1 |spam) =0.754716981

P(F9=0 | non-spam) =0.7549668874, P(F9=0 |spam) =0.6603773584

P(F9=1 | non-spam) =0.245033113, P(F9=1 |spam) =0.339622642

P(F10=0 | non-spam) =0.7086092715, P(F10=0 |spam) =0.3396226415

P(F10=1 | non-spam) =0.291390729, P(F10=1 |spam) =0.660377359

P(F11=0 | non-spam) =0.4172185430, P(F11=0 |spam) =0.3396226415

P(F11=1 | non-spam) =0.582781457, P(F11=1 |spam) =0.660377359

P(F12=0 | non-spam) =0.6622516556, P(F12=0 |spam) = 0.2264150943

P(F12=1 | non-spam) =0.337748344, P(F12=1 |spam) =0.773584906

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

```
Classifying email 1 (Vector F1)...
Probability of being spam: 3.0202448743873 97E-6
Probability of not being spam: 4.620049715764377E-4
Class: Non-Spam
-------------------------------------------------------
Classifying email 2 (Vector F2)...
Probability of being spam: 5.514097676197845E-5
Probability of not being spam: 4.085563593057941E-5
Class: Spam
-------------------------------------------------------
Classifying email 3 (Vector F3)...
Probability of being spam: 1.8644455371759412E-4
Probability of not being spam: 1.277677419012157E-4
Class: Spam
-------------------------------------------------------
Classifying email 4 (Vector F4)...
Probability of being spam: 5.235091115604818E-6
Probability of not being spam: 6.037954762596701E-4
Class: Non-Spam
-------------------------------------------------------
Classifying email 5 (Vector F5)...
Probability of being spam: 5.863981931440462E-5
Probability of not being spam: 9.134498979293798E-5
Class: Non-Spam
-------------------------------------------------------
Classifying email 6 (Vector F6)...
Probability of being spam: 5.5933366115278246E-5
Probability of not being spam: 4.531325026841299E-5
Class: Spam
-------------------------------------------------------
Classifying email 7 (Vector F7)...
Probability of being spam: 3.435528544615664E-6
Probability of not being spam: 3.286364419665509E-4
Class: Non-Spam
-------------------------------------------------------
Classifying email 8 (Vector F8)...
Probability of being spam: 6.1902539574221E-5
Probability of not being spam: 3.9404283148337113E-4
```

```
Class: Non-Spam
-------------------------------------------------------
Classifying email 9 (Vector F9)...
Probability of being spam: 1.8644455371759412E-4
Probability of not being spam: 3.693654303932347E-5
Class: Spam
-------------------------------------------------------
Classifying email 10 (Vector F10)...
Probability of being spam: 2.0416855350858795E-5
Probability of not being spam: 6.881308235775479E-4
Class: Non-Spam
-------------------------------------------------------
```

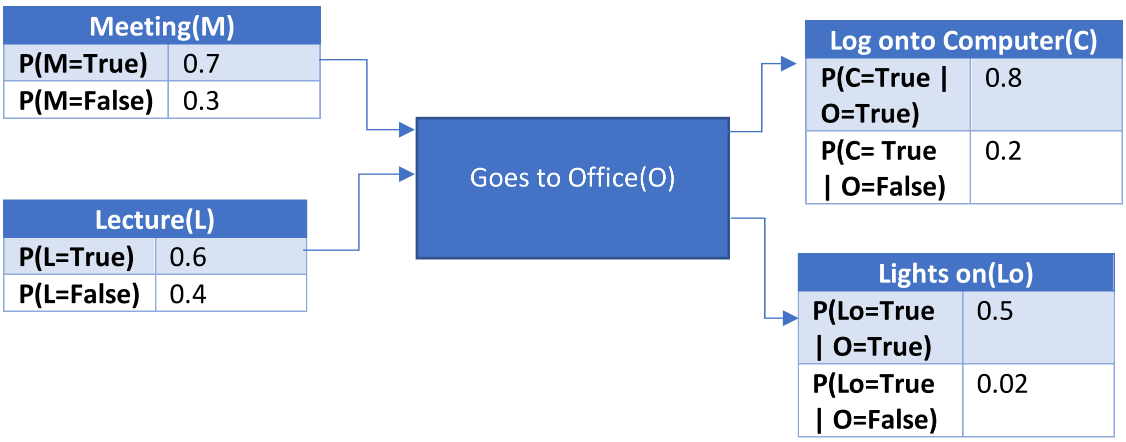
The derivation of the Naive Bayes algorithm assumes that the attributes are conditionally independent. Why is this like to be an invalid assumption for the spam data? Discuss the possible effect of two attributes not being independent

This is because in Naive Bayes we will us P(X,Y) = P(X) * P(Y) but if the two attributes that we use are not independent then P(X,Y) = P(X) * P(Y) would be false

Part 3: Bayesian Networks

Construct a Bayesian network to represent the above scenario

Meeting(M)	Lecture(L)	P(Office = True | M,L)
T	T	0.95
T	F	0.75
F	T	0.80
F	F	0.06



Calculate how many free parameters in your Bayesian network

There are 10 free parameters in the Bayesian Network I have Made

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 = T, M = F, O =T, C = T, Lo =F) = P(L=T) * P(M=F) * P(O=T | M=F,L=T) * P(C=T|O=T) + P(Lo=F|O=T)

=0.6 * 0.3 * 0.8 * 0.8 *0.5 = 0.0576

Calculate the probability that Rachel is in the office.

P(L=T, M =T, O=T) + P(L=T, M=F, O=T) + P(L=F, M=T, O=T) + P(L=F, M=F, O=T)

Because L and M are independent of each other

P(O=T | L=T, M=T) * P(M=T) * P(L=T) + P(O=T | L=T, M=F) * P(M=F) * P(L=T) + P(O=T | L=F, M=T) * P(M=T) * P(L=F) + P(O=T | L=F, M=F) * P(M=F) * P(L=F)

=0.95 * 0.7 * 0.6 + 0.8 * 0.3 * 0.6 + 0.75 * 0.7 * 0.4 + 0.06 * 0.3 * 0.4

= 0.399 + 0.144 +0.21 + 0.0072

=0.7602

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

P(C=T, Lo=F | O=T)

= P(C=T, Lo=F, O=T) / P(O=T)

= P(C = T | O = T) * P(Lo=F | O=T)

=0.8 * 0.5 = 0.4

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?

There is a 80% chance that Rachel is logged on if she is in her office

The probability she is logged on is:

P(C=T | O=T) * P(O=T) + P(C=T | O=F) * P(O=F) = 0.8 * 0.7602 + 0.2*(1-0.7602)

=0.65612

The probability that she is in her office when she is logged on is:

P(C=T | O=T) * P(O=T) / P(C=T)

= 0.8 * 0.7602 / 0.65612

= 0.927

Therefore, the probability that she is logged on and not in her office is:

1 – 0.927 = 0.073

So, the probability that her office light is on when she is logged in is:

P(Lo=T | O=T) * P(O=T | C=T) + P(Lo=T | O=F) * P(O=F | C=T)

= 0.5 * 0.927 + 0.02 * 0.073

= 0.46496

Part 4: Inference in Bayesian Networks

Using inference by enumeration to calculate the probability P(P = t | X = t)

describe what are the evidence, hidden and query variables in this inference

- Evidence Variables = X
- Hidden Variables = S, C, D
- Query Variables = P

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

1. Join P(X | C) to P(C | P, S) to get P(X, C | P, S)
2. Eliminate C to get P(X | P, S)
3. Join P(S) to P(X | P, S) to get P(X, S | P)
4. Eliminate S to get P(X|P)
5. Join P(X|P) to P(P) to get P(X,P)
6. Eliminate P to get P(X)
7. Find P(P | X) from P(X,P) and P(X)

report the probability

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.

The groups that I found are

- Independent:
  - P(Pollution) and S(Smokers)
- Conditionally independent (given cancer)
  - X(X-ray) and D(Dyspnoea)
  - X(X-ray) and P(Pollution)

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.

1. Join P(X | C) and P(C | P, S)

P(X | C):

X	C	Probability
+	+	**0.9**
-	+	**0.1**
+	-	**0.2**
-	-	**0.8**

P(C | P, S):

C	P	S	Probability
+	+	+	0.05
-	+	+	0.95
+	+	-	0.98
-	+	-	0.02
+	-	-	0.001
-	-	-	0.999
+	-	+	0.03
-	-	+	0.97

P(X, C | P, S)

P	S	C	X	Probability
+	+	-	+	0.19
+	-	+	+	0.18
+	-	-	+	0.196
-	+	+	+	0.027
-	+	-	+	0.194
-	-	+	+	0.0009
-	-	-	+	0.1998
+	+	+	-	0.005
+	+	-	-	0.76
+	-	+	-	0.002
+	-	-	+	0.784
-	+	+	-	0.0003
-	+	-	-	0.776
-	-	+	-	0.0001
-	-	-	-	0.7992

2. Eliminate C to get P(X | P, S)

P(X | P, S):

P	S	X	Probability
+	+	+	**0.235**
+	-	+	**0.214**
-	+	+	**0.221**
-	-	+	**0.2007**
+	+	-	**0.765**
+	-	-	**0.786**
-	+	-	**0.779**
-	-	-	**0.7993**

3. Join P(S) and P(X | P, S) to get P(X, S | P) and P(S)

P(S):

S	Probability
+	0.3
-	0.7

P(X, S | P):

P	S	X	Probability
+	+	+	**0.0705**
+	-	+	**0.1498**
-	+	+	**0.0663**
-	-	+	**0.14049**
+	+	-	**0.2295**
+	-	-	**0.5502**
-	+	-	**0.2337**
-	-	-	**0.55951**

4. Eliminate S to get P(X | P)

P(X | P):

P	X	Probability
+	+	**0.2203**
-	+	**0.20679**
+	-	**0.7797**
-	-	**0.79321**

5. Join P(P) and P(X | P) to get P(X, P) and then eliminate P to get P(X) and find P(P | X) from P(X,P) and P(X)

P(X):

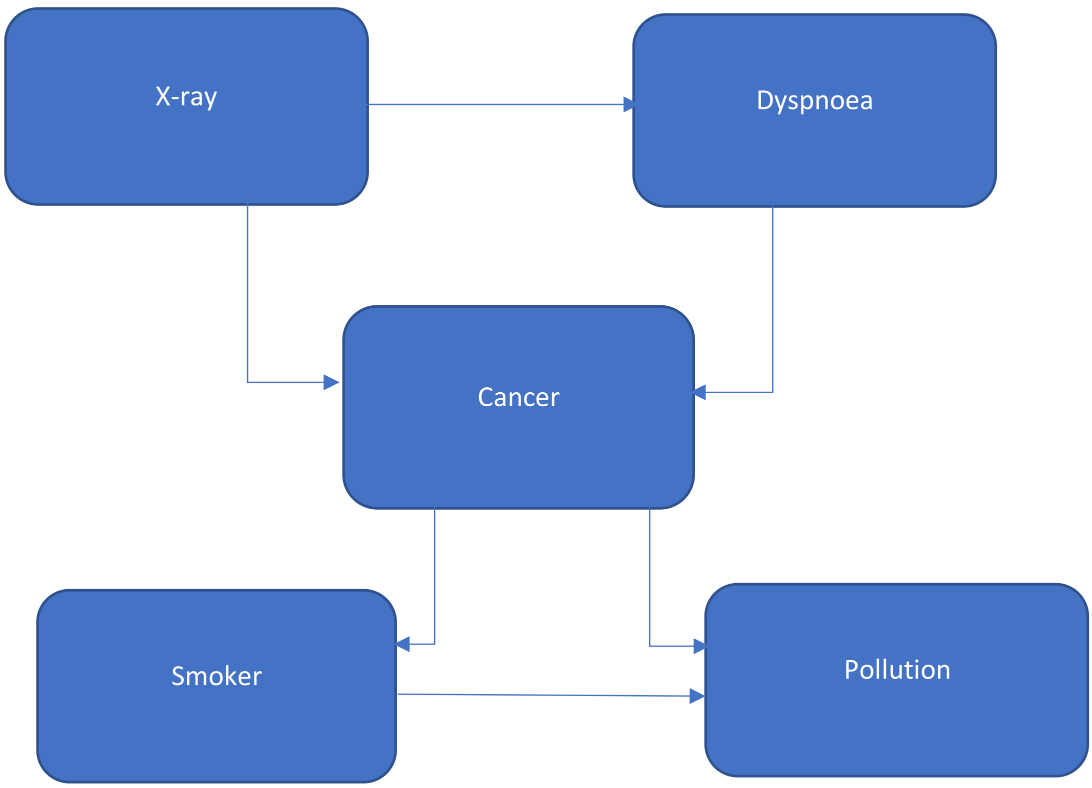
X	Probability
+	**0.218949**
-	**0.781051**

P(P | X):

P	X	Probability
+	+	**0.905553**
-	+	**0.094447**
+	-	**0.898443**
-	-	**0.101557**

Therefore, P(P=T | X=T) = 0.905553

If given the variable order as , 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



Reasoning:

X-ray to Dyspnoea: This connection is here as they are both a common cause of cancer. E.g. if someone has dyspnoea then the probability that they will have an X-ray will increase and if someone has an X-ray then the chance that they have dyspnoea will increase as it is a common reason to get an X-ray

X-ray to Cancer, Dyspnoea to Cancer: both X-ray and Dyspnoea have links to cancer as cancer is a common reason people will have either of the two

Cancer to Smoker: if the person smokes then this may lead to them having cancer

Cancer to Pollution: this may be another reason as to why a person has cancer

Smoker to Pollution: they both have a possibility to impact the likelihood of someone having cancer. E.g. if a person smokes and has cancer then this will decrease the likelihood of pollution being the main cause for them having cancer