COMP 307 Assignment 3: Uncertainty and Probability

10% of Final Mark, Due: 23:59 Sunday 16 May 2021

Part 1: Reasoning Under Uncertainty Basics [30 marks]

Question 1 [10 marks]

1. **Compute the table of the joint distribution P(X, Y, Z). Show the rule(s) you used, and the steps of calculating each joint probability.**

   The probability P(X, Y, Z) can be transformed into P(X) * P(Y|X) * P(Z|X,Y) using the product rule. This can be simplified further as since Z is independent of X given Y, P(Z|X, Y) is equivalent to P(Z|Y), via one of the independence rules, hence the final equation is just P(X) * P(Y|X) * P(Z|Y).

X	Y	Z	Working	P(X, Y, Z)
0	0	0	0.35*0.1*0.7	0.0245
0	0	1	0.35*0.1*0.3	0.0105
0	1	0	0.35*0.9*0.2	0.063
0	1	1	0.35*0.9*0.8	0.252
1	0	0	0.65*0.6*0.7	0.273
1	0	1	0.65*0.6*0.3	0.117
1	1	0	0.65*0.4*0.2	0.052
1	1	1	0.65*0.4*0.8	0.208

2. **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). Show the rule(s) used, and the steps of calculating each joint probability.**

   P(X,Y) = P(X) * P(X|Y), this is the product rule.

X	Y	Working	P(X,Y)
0	0	0.35*0.1	0.035
0	1	0.35*0.9	0.315
1	0	0.65*0.6	0.39
1	1	0.65*0.4	0.26

3. **From the above joint probability table of X, Y, and Z, calculate the following probabilities. Show your working.**

   a. P(Z = 0)
   P(Z=0) is simply the equation $\sum_{x,y=0}^{1} P(Z = 0|X = x, Y = y)$, which means that we sum over the rows in the table from question 1 where Z is 0.
   This is 0.0245 + 0.063 + 0.273 + 0.052 = 0.4125, P(Z=0) = 0.4125

   b. P(X = 0, Z = 0)
   P(X=0, Z=0) is simply the equation $\sum_{y=0}^{1} P(X = 0, Z = 0 \,|Y = y)$, which means that we sum over the rows in the table from question 1 where X is 0 and Z is 0.
   This is 0.0245 + 0.063, therefore P(X=0, Z=0) = 0.0875

   c. P(X = 1, Y = 0|Z = 1)
   For this one, I added up all the entries in the conditional probability table where Z = 1, then divided the line where X=1, Y=0, and Z = 1. This is using both Bayes Rule and the Product rule one after another.

   P(X=1, Y=0 | Z=1) = (0.117)/(0.117+0.0105+0.252+0.208)
   = 0.19914...
   = 0.199

   d. P(X = 0|Y = 0, Z = 0)
   P(X=0 |Y=0,Z=0) = (P(X=0,Y=0,Z=0))/(P(Y=0,X=0)
   = (0.0245) / (0.0245+0.273)
   = 0.08235...
   = 0.082

Question 2 [10 marks]
Consider three Boolean variables A, B, and C, P(B) = 0.7, P(C) = 0.4, P(A|B) = 0.3, P(A|C) = 0.5, and P(B|C) = 0.2, we also know that A is independent from B given C. Calculate the following probabilities. Show your working.

i. **P(B, C)**
   P(B,C) = P(C) * P(B|C) (Via the product rule)
   = 0.4*0.2
   = 0.08

ii. **P(¬A|B)**
   P(¬A|B) = 1-P(A|B) (via them being mutually exclusive)
   = 1-0.3
   = 0.7

iii. **P(A, B|C)**
   P(A,B|C) = P(A|C) * P(B|C) (as A is independent from B given C)
   = P(A|C) * P(B|C)
   = 0.5*0.2
   = 0.1

iv. **P(A|B, C)**
   P(A|B,C) = P(A|C) (as A is independent from B given C)
   = 0.5

v. **P(A, B, C)**
   P(A,B,C) = P(A|B,C) * P(B,C) (via the product rule)
   = 0.5 * 0.08
   = 0.04

Question 3 [10 marks]

Dragonfly has a rare species, which always has an extra set of wings. However, common dragonflies can sometimes mutate and get an extra set of wings. A dragonfly either belongs to the common species or the rare species with the extra wings. There are 0.3% of dragonflies belonging to the rare species with the extra set of wings. For the common dragonflies, the probability of the extra-wing mutation is 0.1%. Now you see a dragonfly with an extra pair of wings. What is the probability that it belongs to the rare species? Show your working.

I will be using the terms R for Rare, ¬R for common, EW for extra wings and ¬EW for no extra wings. We can then produce a table with various probabilities in it

	¬EW	EW	Total
¬R	0.996003	0.000977	0.997
R	0	0.003	0.003
Total	0.996003	0.003997	1

From the question, we know that we are trying to find P(R|EW), which in terms of the table would be [R, EW]/[Total, EW].
Also, from the question we can deduce the values in the yellow cells [R, ¬EW] and [R, Total], and we know that both totals must add to 1 so we can fill in that cell ([Total, Total]) as well. From knowing the yellow values, we can then work out by maths the values in the pink cells.
The last piece of information in the question is that only 0.1% of common (¬R) dragonflies have EW, so we can multiply the total of common dragonflies, 0.997, by 0.1% to get the value in the [¬R, EW] green cell, and the grey cells can then be worked out using the row and column totals.

Finally, we can divide the two relevant boxes, to give us our probability P(R|EW) of 0.003/0.003997 which is 0.7505629... or 0.751 rounded to 3 decimal places.

Part 2: Naive Bayes Method [25 marks]

1. The conditional probabilities P(Fi |C) for each feature i and each class label.

	Not Spam	Spam
P(Class)	0.740196	0.259804
P(Feature 0 = true | Class)	0.357616	0.660377
P(Feature 0 = false | Class)	0.642384	0.339623
P(Feature 1 = true | Class)	0.576159	0.584906
P(Feature 1 = false | Class)	0.423841	0.415094
P(Feature 2 = true | Class)	0.344371	0.452830
P(Feature 2 = false | Class)	0.655629	0.547170
P(Feature 3 = true | Class)	0.397351	0.603774
P(Feature 3 = false | Class)	0.602649	0.396226
P(Feature 4 = true | Class)	0.337748	0.490566
P(Feature 4 = false | Class)	0.662252	0.509434
P(Feature 5 = true | Class)	0.470199	0.358491
P(Feature 5 = false | Class)	0.529801	0.641509
P(Feature 6 = true | Class)	0.503311	0.773585
P(Feature 6 = false | Class)	0.496689	0.226415
P(Feature 7 = true | Class)	0.350993	0.754717
P(Feature 7 = false | Class)	0.649007	0.245283
P(Feature 8 = true | Class)	0.245033	0.339623
P(Feature 8 = false | Class)	0.754967	0.660377
P(Feature 9 = true | Class)	0.291391	0.660377
P(Feature 9 = false | Class)	0.708609	0.339623
P(Feature 10 = true | Class)	0.582781	0.660377
P(Feature 10 = false | Class)	0.417219	0.339623
P(Feature 11 = true | Class)	0.337748	0.773585
P(Feature 11 = false | Class)	0.662252	0.226415

2. **For each instance in the unlabelled set, given the input vector F = (f1, f2, . . . , f12), the probability P(C = 1, F) (enumerator of P(C = 1|F), score of spam), the probability P(C = 0, F) (enumerator of P(C = 0|F), score of non-spam), and the predicted class of the input vector.**

   Note that for my answer, as per the assignment, 1 is Spam and 0 is not spam, so P(C=1, F) is equivalent to P(C=Spam, F)

```
Instance 0
P(C=1, F) = 0.00000366
P(C=0, F) = 0.00045599
The predicted class for this label is NOT Spam (Class 0)
Instance 1
P(C=1, F) = 0.0000579
P(C=0, F) = 0.00004182
The predicted class for this label is Spam (Class 1)
Instance 2
P(C=1, F) = 0.00018771
P(C=0, F) = 0.00012845
The predicted class for this label is Spam (Class 1)
Instance 3
P(C=1, F) = 0.00000615
P(C=0, F) = 0.00059518
The predicted class for this label is NOT Spam (Class 0)
Instance 4
P(C=1, F) = 0.00006199
P(C=0, F) = 0.00009228
The predicted class for this label is NOT Spam (Class 0)
Instance 5
P(C=1, F) = 0.0000597
P(C=0, F) = 0.00004626
The predicted class for this label is Spam (Class 1)
Instance 6
P(C=1, F) = 0.00000412
P(C=0, F) = 0.00032587
The predicted class for this label is NOT Spam (Class 0)
Instance 7
P(C=1, F) = 0.00006519
P(C=0, F) = 0.00038986
The predicted class for this label is NOT Spam (Class 0)
Instance 8
P(C=1, F) = 0.00018771
P(C=0, F) = 0.00003781
The predicted class for this label is Spam (Class 1)
Instance 9
P(C=1, F) = 0.00002277
P(C=0, F) = 0.00067543
The predicted class for this label is NOT Spam (Class 0)
```

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 conditionally independent.**

   This is likely to be an invalid assumption for the spam data as the features are likely to be correlated with each other. For instance, if one of the features is the presence of "MILLION DOLLARS", as suggested by the example, and another is a significant amount of text in caps, these are not conditionally independent as the first phrase uses a high number of caps, so is correlated with whether lots of caps were used. There are more examples like this, but the reality is that we are likely to find correlated features in most datasets.

   The possible effect of two attributes not being conditionally independent is that we would have to include far more terms in the calculation of our probability as if two features are not conditionally independent, then we need to find the probabilities of the features being true given that other features were true, and this makes the calculations much longer, so the algorithm would take a lot longer to run if we didn't make this assumption.

Part 3: Building Bayesian Network [30 marks]

This part is to build a Bayesian Network for the problem described below.

Dr. Rachel Nicholson is a Professor, who lives far away from her university. So, she prefers to work at home, and she only comes to her office if she has research meetings with her postgraduate students, or teaching lectures for undergraduate students, or she has both meetings and teaching:

- The probability for Rachel to have meetings is 70%, the probability of Rachel has lectures is 60%.
- If Rachel has both meetings and lectures, the probability of Rachel comes to her office is 95%.
- If Rachel only has meetings (without lectures), the probability of Rachel comes to her office is 75% because she can Skype with her students.
- If Rachel only has lectures (without meetings), the probability of Rachel comes to her office is 80%.
- If Rachel has neither meetings nor lectures, there is a only 6% chance that she comes to the office.
- When Rachel is in her office, half the time her light is off (when she is trying to hide from others to get work done quickly).
- When she is not in her office, she leaves her light on only 2% of the time since the cleaners come for cleaning.
- When Rachel is in her office, 80% of the time she logged onto the computer.
- Because she sometimes works from home, 20% of the time she is not in her office, she is still logged onto the computer.

Note regarding the calculation, you should show your working process of the calculation to demonstrate that you know how to calculate them.

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 effect. Finally, you have to add the prior probabilities for nodes without parents, and the conditional probabilities for nodes that have parents.)**

   My approach to this section is broken down into the following steps

   a. Domain Variables
   From analysis of the above data I have determined that there are five domain variables in the question:
      - Has Meeting (HM)
      - Has Lecture (HL)
      - In Office (IO)
      - Lights On (LO)
      - Logged Into Computer (LIC)

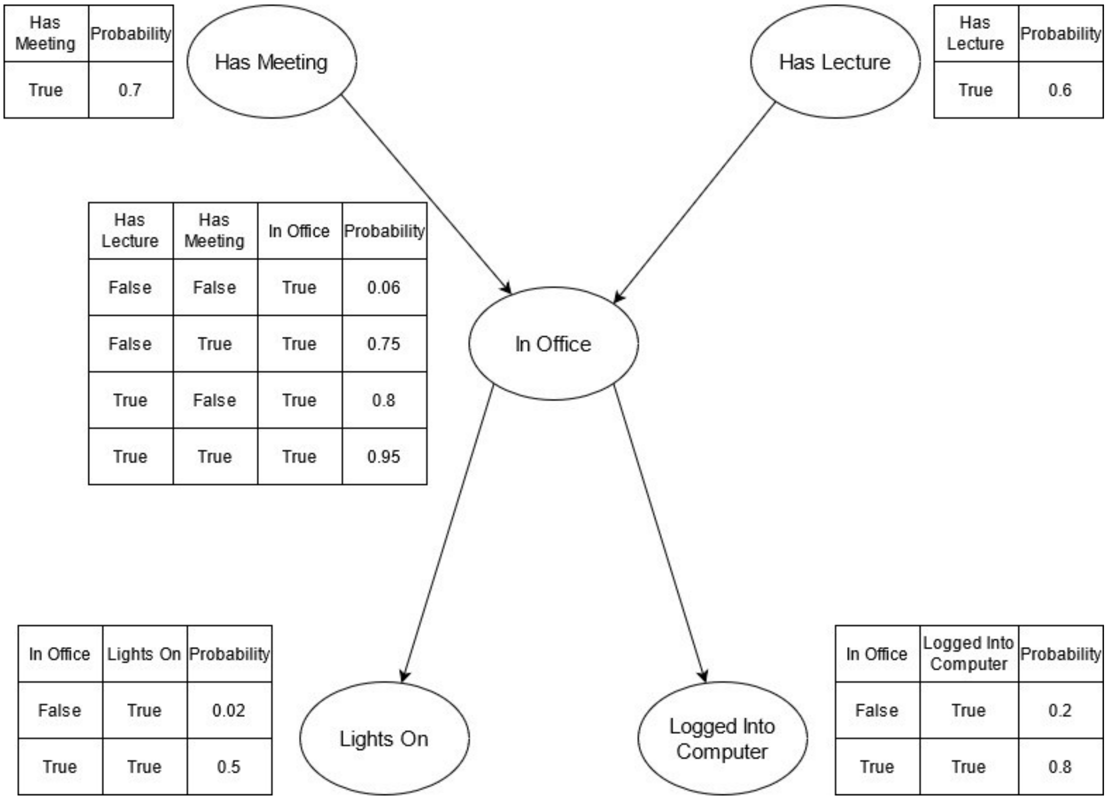
   b. Probabilities
   These are the probabilities I have deduced from the above information

Event	Probability
P(Has Meeting)	0.7
P(Has Lecture)	0.6
P(In Office | Has Meeting, Has Lecture)	0.95
P(In Office | Has Meeting, ¬ Has Lecture	0.75
P(In Office | ¬ Has Meeting, Has Lecture)	0.8
P(In Office | ¬ Has Meeting, ¬ Has Lecture)	0.06
P(Lights On | In Office)	0.5
P(Lights On | ¬ In Office)	0.02
P(Logged Into Computer | In Office)	0.8
P(Logged Into Computer | ¬ In Office)	0.2

   c. Variable Relationships
   The next step is to identify the relationships that exist between the variables. We can see a rough relationship in the above table. Has Meeting and Has Lecture do not seem to have any causes, or other factors influencing their probability, so I would select them as the two main causes. We can then see

that they influence whether she is in the office, which in turn influences whether or not the lights are on and whether or not she is logged into her computer.

d. The final step is to put it all together and draw it out, which I have done here. Note that I have only included the free parameters in the tables, i.e. half of the possible outcomes, as the others can be worked out



Has Meeting	Probability
True	0.7

Has Lecture	Probability
True	0.6

Has Lecture	Has Meeting	In Office	Probability
False	False	True	0.06
False	True	True	0.75
True	False	True	0.8
True	True	True	0.95

In Office	Lights On	Probability
False	True	0.02
True	True	0.5

In Office	Logged Into Computer	Probability
False	True	0.2
True	True	0.8

2. **Calculate how many free parameters in your Bayesian network?**
There are 10 free parameters in the network, 1 for each initial cause, four for the middle node and two per final node, 1+1+4+2+2=10.

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?**

I have rearranged it to be in the same order as the network.

P(¬LO, LIC, IO, HL, ¬HM)
= P(¬LO | LIC, IO, HL, ¬HM) * P(LIC, IO, HL, ¬HM)
Via conditional independence we can remove LIC, HL and ¬HM form the first argument as they have no impact on ¬LO, and we can do this for other variables too, which I will denote by '(via CI)'.
= P(¬LO | IO)$^{\text{(via CI)}}$ * P(LIC | IO, HL, ¬HM) * P(IO, HL, ¬HM)
= P(¬LO | IO) * P(LIC | IO)$^{\text{(via CI)}}$ * P(IO | HL, ¬HM) * P(HL, ¬HM)
= P(¬LO | IO) * P(LIC | IO) * P(IO | HL, ¬HM) * P(HL) * P(¬HM)$^{\text{(via CI)}}$
= (1-0.5) * 0.8 * 0.8 * 0.6 * (1-0.7)
= 0.0576

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

We need to use the sum rule for this: $P(X) = \sum_y P(X,Y)$

P(IO) = P(IO, HL, HM) + P(IO, HL, ¬HM) + P(IO, ¬HL, HM) + P(IO, ¬HL, ¬HM)

Then we use the product rule

P(IO) = P(IO | HL, HM) * P(HL, HM) + P(IO | HL, ¬HM) * P(HL, ¬HM)
+ P(IO | ¬HL, HM) * P(¬HL, HM) + P(IO | ¬HL, ¬HM) * P(¬HL, ¬HM)

Then conditional independence

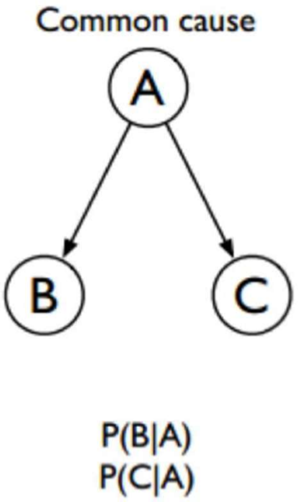
P(IO) = P(IO | HL, HM) * P(HL) * P(HM) + P(IO | HL, ¬HM) * P(HL) * P(¬HM)
+ P(IO | ¬HL, HM) * P(¬HL) * P(HM) + P(IO | ¬HL, ¬HM) * P(¬HL) * P(¬HM)

P(IO) = 0.95 * 0.6 * 0.7 + 0.8 * 0.6 * 0.3 + 0.75 * 0.4 * 0.7 + 0.06 * 0.4 * 0.3
P(IO) = 0.7602

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

P(LIC, ¬LO | IO) = P(LIC | IO) * P(¬LO | IO) (Via CI)
= 0.8 * (1-0.5)
= 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 Rachel's light is on.**



The two domain variables in this question, Logged Into Computer and Lights On, share a common cause, In Office. The structure is as in the diagram, with A being In Office, B being Logged Into Computer, and C being Lights On.

This kind of relationship means that given A, B and C are conditionally independent. However, in this question, we are not given A, that is we do not have information about whether Rachel is in the office, therefore we cannot assume conditional independence of Lights On and Logged Into Computer. Therefore, it does affect her belief. We can do the calculations as below:

P(LO) = P(LO, IO) + P(LO, ¬IO)
= P(LO | IO) * P(IO) + P(LO | ¬IO) * P(¬IO)
= 0.5 * 0.7602 + 0.02 * (1-0.7602)
= 0.384896

P(LIC) = P(LIC, IO) + P(LIC, ¬IO)
= P(LIC | IO) * P(IO) + P(LIC | ¬IO) * P(¬IO)
= 0.8 * 0.7602 + 0.2 * (1-0.7602)
= 0.65612

P(LO | LIC)
= P(LO, LIC)/P(LIC)
= (P(LO, LIC, IO) + P(LO, LIC, ¬IO))/P(LIC)
= (P(LO, LIC | IO) * P(IO) + P(LO, LIC | ¬IO) * P(¬IO)) /P(LIC)
= (P(LO | IO) * P(LIC | IO) * P(IO) + P(LO | ¬IO) * P(LIC | ¬IO) * P(¬IO)) /P(LIC)
= (0.5 * 0.8 * 0.7602 + 0.02 * 0.2 * (1-0.7602))/0.65612
= 0.464913735...
= 0.465

The knowledge that Rachel is logged in should increase the probability that the lights will be on, which makes sense, as being logged on means there is a higher chance that she is in the office, meaning there is a higher chance that the lights are on.