COMP 307 — Introduction to AI Assignment 3: Uncertainty and Probability 10% of Final Mark — Due: 23:59 Monday 28 May 2018 1

Question Description Part 1: Reasoning Under Uncertainty Basics [10 marks]

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=x, Y=y)
1	1	0.14
1	0	0.56
0	1	0.21
0	0	0.09

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=x,Y=y,Z=z)
0	0	0	0.054
0	0	1	0.036
0	1	0	0.168
0	1	1	0.042
1	0	0	0.336
1	0	1	0.224
1	1	0	0.112
1	1	1	0.028

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

First we need to find P(Y), from the P(X=x,Y=y)table using the sum rule:

THE NORMALISATION RULE : INSERT EQUATION

Y	P(Y)
0	0.56 + 0.09 = 0.65
1	0.21 + 0.14 = 0.35

Then from here we need to construct a P(Z =z ,Y =y) table.

Use the product rule: P(Z, Y) = P(Y) * P(Z | Y)

Z	Y	P(Z=z,Y=y)
0	0	0.65 * 0.6 = 0.39
0	1	0.35 * 0.8 = 0.28
1	0	0.65 * 0.4 = 0.26
1	1	0.35 * 0.2 = 0.07

From here we use the Sum rule again to construct Z probability from the above table:

THE NORMALISATION RULE : INSERT EQUATION

Z	P(Z)
0	0.28 +0.39 = 0.67
1	0.26+0.07= 0.33

Now we can construct the P(X=x,Z=z) table:

THE NORMALISATION RULE : INSERT EQUATION

X	Z	P(X=x,Z=z)
0	0	0.054 + 0.168 = 0.222
0	1	0.042 + 0.036 =0.078
1	0	0.112 +0.336 = 0.448
1	1	0.224 +0.028 = 0.252

If independent: P(X=x, Z=z) = P(X=x) * P(Z=z)
However as illustrated below this is not the case:
From above table: P(X=0, Z=0) = 0.222.

And:
P(X=0) = 0.3
P(Z=0) = 0.67.
0.3 * 0.67 = 0.14874.

Thus they are not independent.

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(A,B) = P(B) * P(A|B)
let A = X,Y & B = Z
P(A|B) = P(A|B) / P(B)
P(A|B) = P(X,Y,Z)/P(Z)
plug in values from probability table for x=1, y=0, z=1
P(X = 1, Y = 0|Z = 1) = 0.224/0.33
P(X = 1, Y = 0|Z = 1) = 0.679

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

P(A,B) = P(B) * P(A|B)
let A = X & B = Z,Y
P(A|B) = P(A|B) / P(B)
P(A|B) = P(X,Y,Z)/P(Y, Z)
plug in values from probability table for x=0, y=0, z=0
P(X = 0|Y = 0, Z = 0) = 0.054/0.39
P(X = 0|Y = 0, Z = 0) = 0.138

Part 2: Naive Bayes Method [25 marks]

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

	Spam	Not spam
Total	51	149
P(Feature 1 = t)	0.6667	0.3557
P(Feature 1 = f)	0.3333	0.6443
P(Feature 2 = t)	0.5882	0.5772
P(Feature 2 = f)	0.4118	0.4228
P(Feature 3 = t)	0.451	0.3423
P(Feature 3 = f)	0.549	0.6577
P(Feature 4 = t)	0.6078	0.396
P(Feature 4 = f)	0.3922	0.604
P(Feature 5 = t)	0.4902	0.3356
P(Feature 5 = f)	0.5098	0.6644
P(Feature 6 = t)	0.3529	0.4698
P(Feature 6 = f)	0.6471	0.5302
P(Feature 7 = t)	0.7843	0.5034
P(Feature 7 = f)	0.2157	0.4966
P(Feature 8 = t)	0.7647	0.349
P(Feature 8= f)	0.2353	0.651
P(Feature 9 = t)	0.3333	0.2416
P(Feature 9 = f)	0.6667	0.7584
P(Feature 10 = t)	0.6667	0.2886
P(Feature 10 = f)	0.3333	0.7114
P(Feature 11 = t)	0.6667	0.5839
P(Feature 11 = f)	0.3333	0.4161
P(Feature 12 = t)	0.7843	0.3356
P(Feature 12 = f)	0.2157	0.6644

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*.

instance 1 : is not Spam
spam prob = 6.040489748774789E-4
not spam prob = 0.03162718597368903

instance 2: is spam
spam prob = 0.011028195352395692
not spam prob = 0.0027968287684020814

instance 3: is spam
spam prob = 0.03728891074351882
not spam prob = 0.008746516559680537

instance 4: is not Spam
spam prob = 0.0010470182231209631
not spam prob = 0.041333650052675405

instance 5: is spam
spam prob = 0.01172796386288092
not spam prob = 0.006253146952268239

instance 6: is spam
spam prob = 0.011186673223055645
not spam prob = 0.003101980890857802

instance 7: is not Spam
spam prob = 6.871057089231324E-4
not spam prob = 0.022497259785629678

instance 8: is not Spam
spam prob = 0.012380507914844192
not spam prob = 0.026974744168660303

instance 9: is spam
spam prob = 0.03728891074351882
not spam prob = 0.0025285418724905998

instance 10: is not Spam
spam prob = 0.004083371070171757
not spam prob = 0.04710694228517442

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.

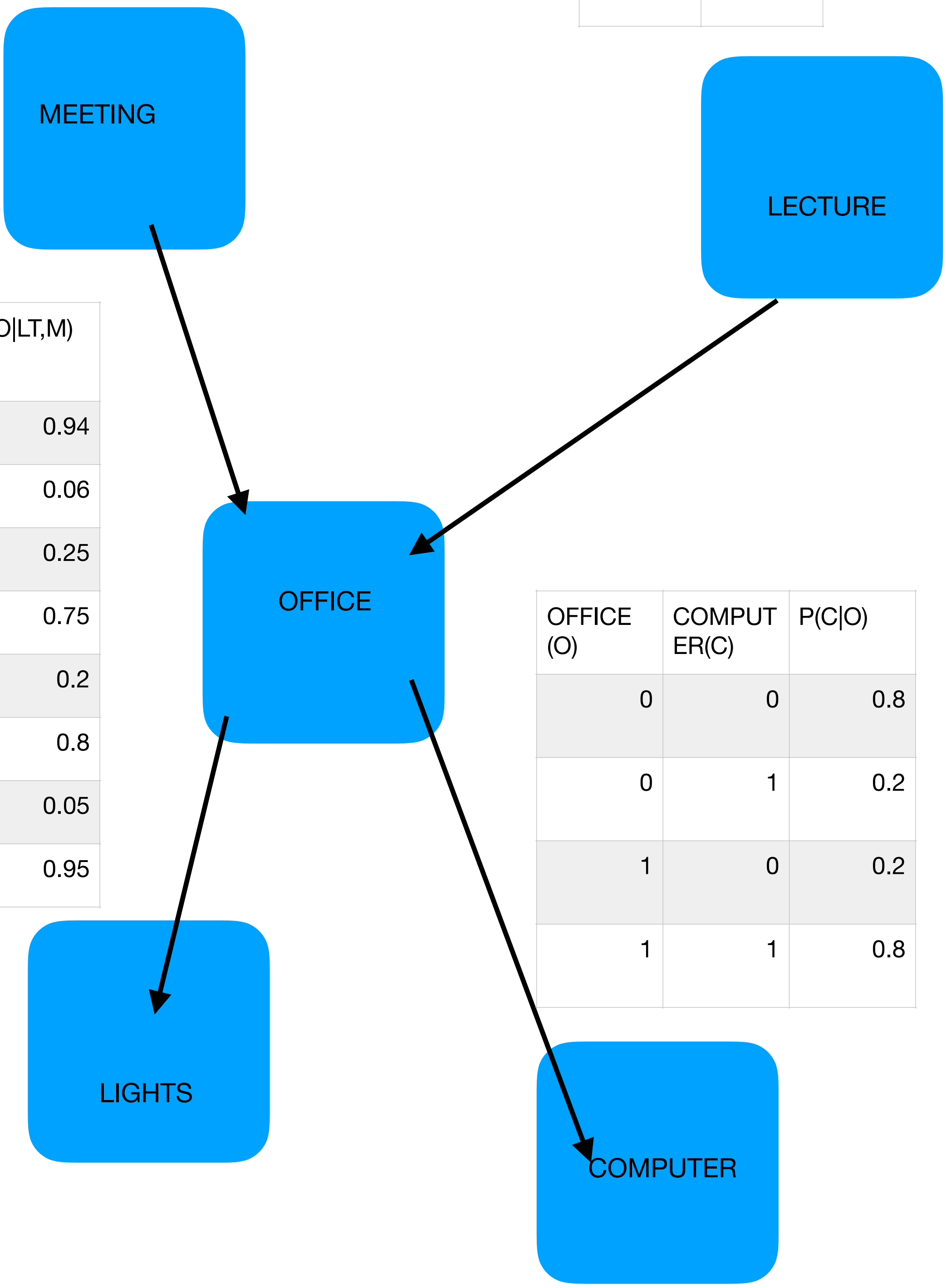
In reality, if one of the features indicates spam, it is more likely that the other features would too. Thus they are not independent. This causes problems for the Bayes algorithm as it needs independence of features to use this calculation: P(A, B | C) = P(A | C) * P(B | C

Part 3: Bayesian Networks [30 marks]

1. Construct a Bayesian network to represent the above scenario.

Meeting (M)	P(M)
0	0.3
1	0.7

Lecture	P(LT)
0	0.4
1	0.6



LECCTURE(LT)	MEETING(M)	OFFICE(O)	P(O|LT,M)
0	0	0	0.94
0	0	1	0.06
0	1	0	0.25
0	1	1	0.75
1	0	0	0.2
1	0	1	0.8
1	1	0	0.05
1	1	1	0.95

OFFICE (O)	COMPUTER(C)	P(C|O)
0	0	0.8
0	1	0.2
1	0	0.2
1	1	0.8

OFFICE (O)	LIGHT (L)	P(L|O)
0	0	0.98
0	1	0.02
1	0	0.5
1	1	0.5

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

P(M=0) = 0.3
P(L=0) = 0.4
P(O = 0 | M=1, L=1) = 0.05
P(O = 0 | M=0, L=1) = 0.2
P(O = 0 | M=0, L=0) = 0.94
P(L = 1|O = 1) = 0.5
P(L = 0|O = 0) = 0.98
P(C =0 |O =1) = 0.2
P(C =0 |O =0) = 0.8

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(Lt=1, M=0, O=1, C=1, L= 0):

= P(Lt=1) * P(M=0) * P(O =1| Lt =1, M=0) * P(C=1| 0=1) * P(L=0|O=1)
= 0.6 * 0.3 *0.8*0.8*0.5
= 0.0576

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

P(O) = P(O = 1, M=1, Lt =1)+ P(O =1, M=0, Lt =1) + P(O =1, M=1, Lt =0) + P(O =1, M=0, Lt =0)

P(O) = P(O| M=1, Lt =1)*P(M=1, Lt =1) + P(O| M=0, Lt =1)*P(M=0, Lt =1) + P(O| M=1, Lt =0)*P(M=1, Lt =0) * P(O| M=0, Lt =0)*P(M=0, Lt =0)

P(O) = P(O| M=1, Lt =1)*P(M=1) * P(Lt =1) + P(O| M=0, Lt =1)*P(M=0) * P(Lt =1) + P(O| M=1, Lt =0)*P(M=1) * P(Lt =0) + P(O| M=0, Lt =0)*P(M=0) * P(Lt =0)

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

P(O) = 0.7602

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

P(L=0, C=1 | O =1)

= P(L = 0 | O =1) * P(C = 1 | O =1)
= 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 students belief that Rachels light is on ?

Light, Logged-on and Office variables have a common cause relationship, where Lights and Logged on are effected by whether Rachel is or isn't in her office. This causes Light and logged on to be dependent on each other. If the student knows the relationship, the student can infer that there is a higher probability that Rachel is in her office if she is logged on. Given Rachel is in her office theres a higher probability that her lights are on, than if she wasn't.

Part 4: Inference in Bayesian Networks [35 marks]

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, (ii) 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 (iii) report the probability.

i) Evidence variables is X-ray. Hidden variables are smoking, Dyspnoea and cancer. The query variable is Pollution.
ii) The first merge will be of the Pollution, Smoker and Cancer classes to create a P(P, C, S) table From here I can make a P(P,C) table by removing the Smoker class . After this removal I will merge this table with X-ray creating P(P,C,X). After this we can eliminate Cancer to create P(P,X) table. From this probability table and the P(X) table we can infer a P(P|X) table.

Stage 1:
P(P,C,S)

Pollution	Cancer	Smoker	P(P=p, C=c, S=s)
0	0	0	0.07
0	0	1	0
0	1	0	0.029
0	1	1	0.001
1	0	0	0.0617
1	0	1	0.013
1	1	0	0.257
1	1	1	0.014

Stage 2:
P(P,C)

Pollution	Cancer	P(P=p, C=c)
0	0	0.099
0	1	0.001
1	0	0.874
1	1	0.027

Stag 3:
P(P,C,X) = merge P(X) and P(P,C) table.

Pollution	Cancer	Xrays	P(P=p, C=c, X=x)
0	0	0	0.0792
0	0	1	0.0198
0	1	0	0.0001
0	1	1	0.0009

1	0	0	0.6992
1	0	1	0.1748
1	1	0	0.0027
1	1	1	0.0243

Stage 4: Create a P(X,C) by removing Pollution:

Cancer	Xray	P(C=c, X=x)
0	0	0.0793
0	1	0.0207
1	0	0.7019
1	1	0.1991

Stage 5: Create P(X)

X-ray	P(X=x)
0	0.7812
1	0.2188

Finally we can create the P(P|X) table:

Pollution	Xray	P(P=p|X=x)
0	0	0.102
0	1	0.094
1	0	0.898
1	1	0.910

iii) P(P=1|X=1) = 0.910

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.

Group 1: Indirect cause. Smoker, Cancer, Dyspnonea are conditionally independent.
Group 2: Indirect cause. Pollution, Cancer, Dyspnonea are conditionally independent.
Group 3: Indirect cause. Pollution, Cancer, Xray are conditionally independent.