Assignment #3

Uncertainty and Probability

COMP 307 Introduction to Artificial Intelligence

Qiangqiang Li(Aaron)

ID: 300422249

Part 1: Reasoning Under Uncertainty Basics

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	$P(X)$
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

Reply: According to Product Rule P(X,Y)=P(X)*P(Y|X)=P(Y)*P(X|Y)

P(X = 0, Y = 0)= P(X = 0)* P(Y = 0|X = 0)=0.300*0.300=0.090

P(X = 0, Y = 1)= P(X = 0)* P(Y = 1|X = 0)=0.300*0.700=0.210

P(X = 1, Y = 0)= P(X = 1)* P(Y = 0|X = 1)=0.700*0.800=0.560

P(X = 1, Y = 1)= P(X = 1)* P(Y = 1|X = 1)=0.700*0.200=0.140

So product rules are used here.

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	$P(X)$
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

Reply:

According to the tables, it is clear that Z is independent from X given Y, So ***P(Z|X, Y) = P(Z|Y) and P(X, Y, Z) = P(X, Y) * P(Z| X, Y) = P(X, Y) * P(Z|Y).***

P(X = 0, Y = 0, Z = 0) = P(X=0, Y=0) * P(Z=0|Y=0)=0.090*0.600=0.054

P(X =1, Y = 0, Z = 1) = P(X=1, Y=0) * P(Z=1|Y=0)= 0.560*0.400=0.224

P(X = 1, Y = 1, Z = 0) = P(X=1, Y=1) * P(Z=0|Y=1)=0.140*0.800=0.112

P(X = 1, Y = 1, Z = 1) = P(X=1, Y=1) * P(Z=1|Y=1)=0.140*0.200=0.028

X	Y	Z	P(X, Y, 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

So product rule and Independence are used here.

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

Reply:

(i) According to Sum Rule:

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)=0.054+0.168+0.336+0.112=0.670

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

(ii) According to the result of above question, P(X = 0, Z = 0)= 0.222, If X and Z are independent, P(Z = 0, X = 0) = P(Z = 0) * P(X = 0) need correct. But P(Z = 0) * P(X = 0) =0.670*0.300=0.201, It is clear that P(Z = 0, X = 0)!= P(Z = 0) * P(X = 0).

So X and Z are dependent.

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

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

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

Reply:

(i) According to the result of the above question, P(Z=0)=0.670, Using Normalization Rule,

P(Z=1)=1- P(Z=0)=1-0.670=0.330

P(X = 1, Y = 0|Z = 1)= P(X = 1, Y = 0,Z = 1)/P(Z)=0.224/0.330=0.679

(ii) P(X = 0|Y = 0, Z = 0)= P(X = 0, Y = 0,Z = 0)/P(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)=0.13846

Part 2: Genetic Programming for Symbolic Regression

1. A report in PDF, text or DOC format. The report should include:

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

Reply:

P(F0 = 1| C = 1) = 0.6666666666666666

P(F0 = 0| C = 1) = 0.3333333333333333

P(F0 = 1| C = 0) = 0.35570469798657717

P(F0 = 0| C = 0) = 0.6442953020134228

P(F1 = 1| C = 1) = 0.5882352941176471

P(F1 = 0| C = 1) = 0.4117647058823529

P(F1 = 1| C = 0) = 0.5771812080536913

P(F1 = 0| C = 0) = 0.4228187919463087

P(F2 = 1| C = 1) = 0.45098039215686275

P(F2 = 0| C = 1) = 0.5490196078431373

P(F2 = 1| C = 0) = 0.3422818791946309

P(F2 = 0| C = 0) = 0.6577181208053692

P(F3 = 1| C = 1) = 0.6078431372549019

P(F3 = 0| C = 1) = 0.39215686274509803

P(F3 = 1| C = 0) = 0.3959731543624161

P(F3 = 0| C = 0) = 0.6040268456375839

P(F4 = 1| C = 1) = 0.49019607843137253

P(F4 = 0| C = 1) = 0.5098039215686274

P(F4 = 1| C = 0) = 0.33557046979865773

P(F4 = 0| C = 0) = 0.6644295302013423

P(F5 = 1| C = 1) = 0.35294117647058826

P(F5 = 0| C = 1) = 0.6470588235294118

P(F5 = 1| C = 0) = 0.4697986577181208

P(F5 = 0| C = 0) = 0.5302013422818792

P(F6 = 1| C = 1) = 0.7843137254901961

P(F6 = 0| C = 1) = 0.21568627450980393

P(F6 = 1| C = 0) = 0.5033557046979866

P(F6 = 0| C = 0) = 0.4966442953020134

P(F7 = 1| C = 1) = 0.7647058823529411

P(F7 = 0| C = 1) = 0.23529411764705882

P(F7 = 1| C = 0) = 0.348993288590604

P(F7 = 0| C = 0) = 0.6510067114093959

P(F8 = 1| C = 1) = 0.3333333333333333

P(F8 = 0| C = 1) = 0.6666666666666666

P(F8 = 1| C = 0) = 0.24161073825503357

P(F8 = 0| C = 0) = 0.7583892617449665

P(F9 = 1| C = 1) = 0.6666666666666666

P(F9 = 0| C = 1) = 0.3333333333333333

P(F9 = 1| C = 0) = 0.28859060402684567

P(F9 = 0| C = 0) = 0.7114093959731543

P(F10 = 1| C = 1) = 0.6666666666666666

P(F10 = 0| C = 1) = 0.3333333333333333

P(F10 = 1| C = 0) = 0.5838926174496645

P(F10 = 0| C = 0) = 0.4161073825503356

P(F11 = 1| C = 1) = 0.7843137254901961

P(F11 = 0| C = 1) = 0.21568627450980393

P(F11 = 1| C = 0) = 0.33557046979865773

P(F11 = 0| C = 0) = 0.6644295302013423

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.

Reply: Look at the “sampleoutput.txt” file ,include Score(spam), Score(non-spam), Instance, The class of the input vector.

Score(spam) is 3.020244874387394e-06, Score(non-spam) is 0.0004620049715764379.

['1', '1', '0', '0', '1', '1', '0', '0', '0', '0', '0', '0'] nospam

Score(spam) is 5.5140976761978446e-05, Score(non-spam) is 4.0855635930579417e-05.

['0', '0', '1', '1', '0', '0', '1', '1', '1', '0', '0', '1'] spam

Score(spam) is 0.0001864445537175941, Score(non-spam) is 0.00012776774190121569.

['1', '1', '1', '1', '1', '0', '1', '0', '0', '0', '1', '1'] spam

Score(spam) is 5.2350911156048155e-06, Score(non-spam) is 0.0006037954762596702.

['0', '1', '0', '0', '0', '0', '1', '0', '1', '0', '0', '0'] nospam

Score(spam) is 5.863981931440459e-05, Score(non-spam) is 9.134498979293801e-05.

['1', '1', '1', '0', '1', '1', '0', '1', '0', '0', '1', '1'] nospam

Score(spam) is 5.5933366115278225e-05, Score(non-spam) is 4.531325026841299e-05.

['1', '1', '1', '1', '1', '1', '0', '0', '0', '1', '1', '1'] spam

Score(spam) is 3.435528544615666e-06, Score(non-spam) is 0.000328636441966551.

['0', '0', '0', '0', '1', '1', '0', '1', '0', '0', '0', '0'] nospam

Score(spam) is 6.190253957422096e-05, Score(non-spam) is 0.00039404283148337113.

['0', '1', '0', '1', '1', '1', '1', '0', '0', '0', '1', '1'] nospam

Score(spam) is 0.0001864445537175941, Score(non-spam) is 3.6936543039323476e-05.

['1', '1', '1', '1', '1', '0', '1', '0', '0', '1', '0', '1'] spam

Score(spam) is 2.0416855350858785e-05, Score(non-spam) is 0.000688130823577548.

['1', '1', '0', '0', '0', '1', '0', '1', '0', '0', '1', '0'] nospam

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.

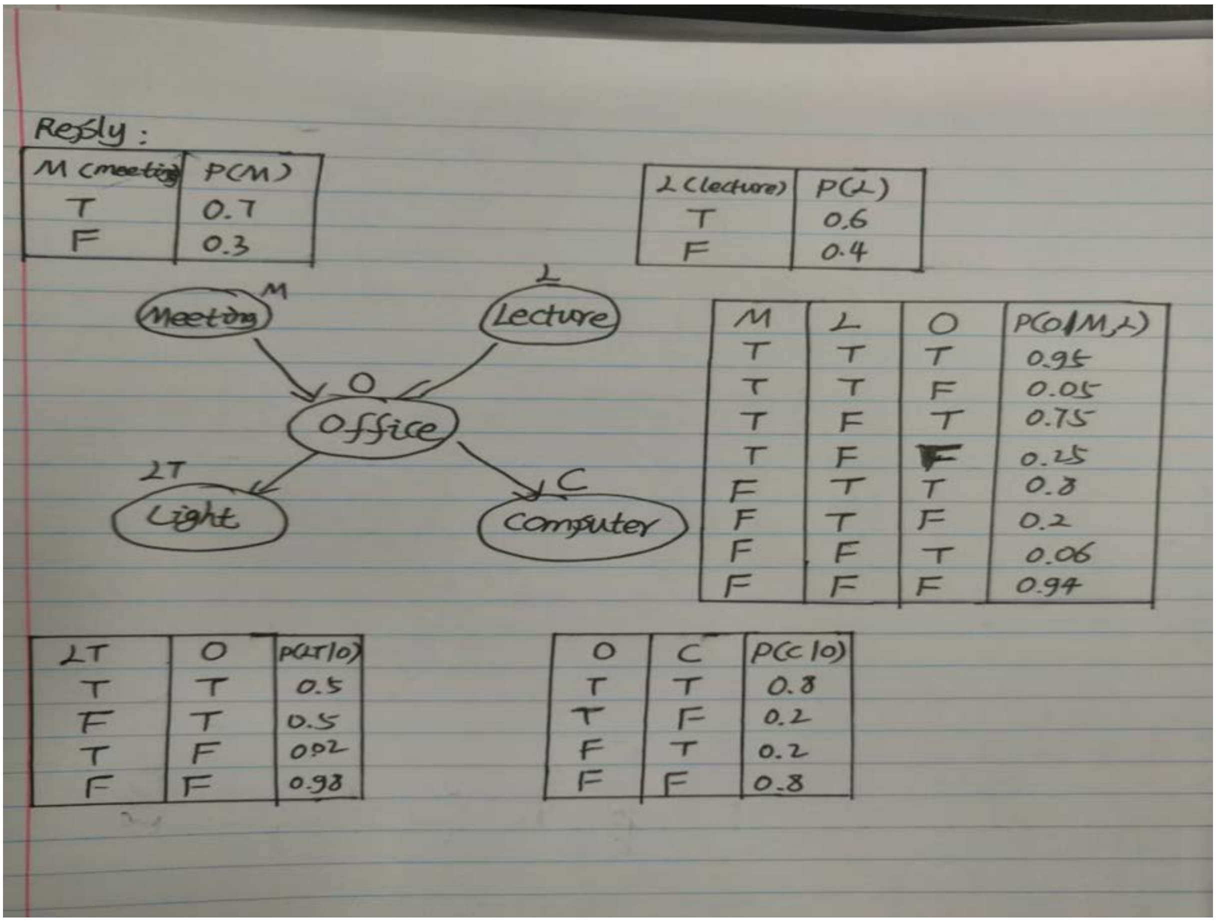
Reply:

Because the spam data includes 200 emails, each email have many attributes, and these attributes can affect others attributes. For example, the email from an invalid reply-to address, that may be contain "Viagra" or "MILLION DOLLARS", which means these attributes are not independent.

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

Reply:



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

Reply: According to Normalization Rule, using the half of the parameters. Like this: P(-X)=1-P(X), So the number of free parameters is 10:

(2+2+8+4+4)/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.

Reply: Using the result of the question 1:

P(L=+L,M=-M,O=+O,C=+C,LT=-LT)

= P(L = +L) * P(M =-M) *P(O = +O| M = -M, L = +L) * P(LT = -LT| O = +O) * P(C = +C| O = +O) = 0.6 * 0.3 * 0.8 * 0.5* 0.8 = 0.0576

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

Reply: Using the result of the question 1:

P(O=+O)=P(O=+O,M,L)= P(O=+O,M=+M,L=+L)+ P(O=+O, M=+M,L=-L)+ P(O=+O, M=-M,L=+L)+ P(O=+O, M=-M,L=-L)=P(O=+O|M=+M,L=+L)*P(M=+M,L=+L)+P(O=+O|M=+M,L=-L) *P(M=+M,L=-L)+P(O=+O|M=-M,L=+L) *P(M=-M,L=+L)+P(O=+O|M=-M,L=-L) *P(M=-M,L=-L)

= P(O=+O|M=+M,L=+L)*P(M=+M)*P(L=+L)+ P(O=+O|M=+M,L=-L) *P(M=+M)*P(L=-L)+P(O=+O|M=-M,L=+L) *P(M=-M) *P（L=+L)+P(O=+O|M=-M,L=-L) *P(M=-M) *P（L=-L)

=0.95*0.7*0.6+0.75*0.7*0.4+0.8*0.3*0.6+0.06*0.3*0.4=0.7602

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

Reply: According to the result of the question 1

P(C=+C,LT=-LT|O=+O)

= P(C=+C |O=+O)* P(LT=-LT|O=+O)

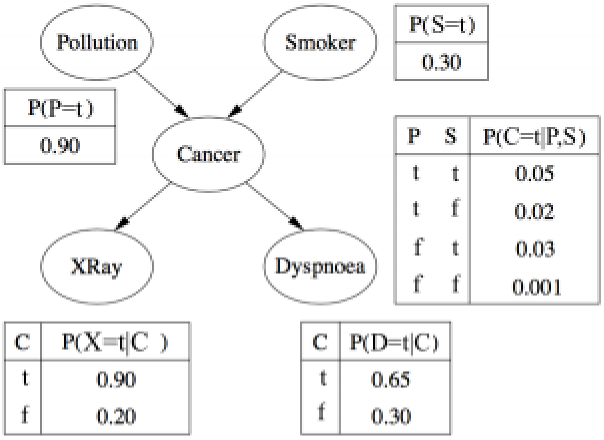
=0.8*0.5=0.4

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

Reply: Suppose the login status is true, but the office status will difficult to check true or false, So the login statue and the light are denpendt.

Bel(LT=+LT)=P(LT=+LT| C=+C)=P(LT=+LT, C=+C)/P(C=+C)

Part 4:Inference in Bayesian Networks



P(P=t)
0.90

P(S=t)
0.30

P	S	P(C=t|P,S)
t	t	0.05
t	f	0.02
f	t	0.03
f	f	0.001

C	P(X=t|C)
t	0.90
f	0.20

C	P(D=t|C)
t	0.65
f	0.30

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:

Reply: P(P = t|X = t)

Evidence variables: X(XRay)

Hidden variables: S(Smoker),C(Cancer),D(Dyspnoea).

Query variables: P(Pollution)

(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,

Reply:

Join P, S and C first, then eliminate S, then join X, then eliminate C, only P and X left.

(iii) report the probability,.

Reply: P(P = t|X = t)

$P(P = t \mid X = t) = \propto P(P = t, X = t, S, C) = \propto P(P = t) * P(S) * P(C|P = t, S) * P(X = t| C)$

$= \propto \sum_{S,C} P(P = t) * P(S) * P(C|P = t, S) * P(X = t| C)$

$= \propto (0.9 * 0.3 * 0.05 * 0.9 + 0.9 * 0.7 * 0.02 * 0.9 + 0.9 * 0.3 * 0.95 * 0.2 + 0.9 * 0.7 * 0.98 * 0.2)$

$= \propto 0.19827$

$P(P = f \mid X = t) = \propto P(P = f, X = t, S, C) = \propto P(P = t) * P(S) * P(C|P = f, S) * P(X = t| C)$

$= \propto \sum_{S,C} P(P = f) * P(S) * P(C|P = f, S) * P(X = t| C)$

$= \propto (0.1 * 0.3 * 0.03 * 0.9 + 0.1 * 0.7 * 0.001 * 0.9 + 0.1 * 0.3 * 0.97 * 0.2 + 0.1 * 0.7 * 0.999 * 0.2)$

$= \propto 0.020678$

$P(P = t \mid X = t) = 0.19827 / (0.19827 + 0.020678) = 0.9055574839688$

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.

Reply:

X is independent on (P, S) given C, because P and S are indirect cause of X.
X is independent on D given C, because C is common cause of X and D.
D is independent on (P, S) given C, because P and S are indirect cause of D.

3. 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.

Reply:

Reply: order as <Xray, Dyspnoea, Cancer, smoker, Pollution>

$$\begin{tikzcd} X \arrow[rr] \arrow[dr] & & D \arrow[dl] \\ & C \arrow[dl] \arrow[dr] & \\ S \arrow[rr] & & P \end{tikzcd}$$

1. step1: Add node X

step2: Add node D

$P(D|X) \neq P(D)$ so $X \to D$

step3: Add node C

$P(C|D,X) \neq P(C)$

$P(C|D,X) \neq P(C|D)$

$P(C|D,X) \neq P(C|X)$ so $D \to C$, $X \to C$

step4: Add node S

$P(S|C,D,X) \neq P(S)$

$P(S|C,D,X) = P(S|C)$ so $C \to S$

step5: Add node P

$P(P|S,C,D,X) \neq P(P)$

$P(P|S,C,D,X) \neq P(P|S)$

$P(P|S,C,D,X) \neq P(P|C)$

$P(P|S,C,D,X) = P(P|S,C)$ so, $S \to P$, $C \to P$