# Assignment 3

## Part 1: Reasoning Under Uncertainty Basics

### Question 1

| X P(X) | P(X) |
|---|---|
| 0 | 0.35 |
| 1 | 0.65 |

| X | Y | P(Y\|X) |
|---|---|---|
| 0 | 0 | 0.10 |
| 0 | 1 | 0.90 |
| 1 | 0 | 0.60 |
| 1 | 1 | 0.40 |
| **Y** | **Z** | **P(Z\|Y)** |
| 0 | 0 | 0.70 |
| 0 | 1 | 0.30 |
| 1 | 0 | 0.20 |
| 1 | 1 | 0.80 |

#### 1. Table of the joint distribution P(X, Y,Z)

P(X,Y) = P(X)*P(Y|X)

P(Z|X,Y) = P(Z|Y)( as $Z \perp X|Y$)

P(X,Y,Z)=P(Z|X,Y)*P(X,Y)=**P(Z|Y)*P(Y|X)*P(X)**

| X | Y | Z | P(X,Y,Z) | Calculations |
|---|---|---|---|---|
| 0 | 0 | 0 | 0.245 | 0.7*0.1*0.35 |
| 0 | 0 | 1 | 0.0105 | 0.3*0.1*0.35 |
| 0 | 1 | 0 | 0.063 | 0.2*0.9*0.35 |
| 0 | 1 | 1 | 0.0252 | 0.8*0.9*0.35 |
| 1 | 0 | 0 | 0.273 | 0.7*0.6*0.65 |
| 1 | 0 | 1 | 0.117 | 0.3*0.6*0.65 |
| 1 | 1 | 0 | 0.052 | 0.2*0.4*0.65 |
| 1 | 1 | 1 | 0.208 | 0.8*0.4*0.65 |

#### 2. Create 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).

**P(X,Y) = P(X)*P(Y|X)**

| X | Y | P(X,Y) | Calculations |
|---|---|---|---|
| 0 | 0 | 0.035 | 0.10*0.35 |
| 0 | 1 | 0.315 | 0.90*0.35 |
| 1 | 0 | 0.39 | 0.60*0.65 |
| 1 | 1 | 0.26 | 0.40*0.65 |

## 3. From the above joint probability table of X, Y , and Z, calculate the following probabilities.

(a) P(Z = 0),
0.245+0.063+0.273+0.052 = 0.633
(b) P(X = 0;Z = 0),
0.245+0.063=0.308
(c) P(X = 1; Y = 0|Z = 1),
P(X,Y,Z)=P(X,Y|Z)*P(Z) ( it comes from P(X,Y) = P(Y|X)*P(X) and then we consider X,Y as X and Y as Z )
P(X,Y|Z)= P(X=1,Y=0,Z=1)/P(Z=1) = 0.117/(0.0105+0.0252+0.117+0.208)= 0.324

(d) P(X = 0|Y = 0;Z = 0).
P(X,Y,Z)=P(X|Y,Z)*P(Y,Z)
P(X|Y,Z) = P(X,Y,Z)/P(Y,Z)
P(X=0,Y=0,Z=0)/P(Y=0,Z=0) =0.245/(0.245+0.273)=0.47

## Question 2

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,
P(B|C) = 0.2

$A \perp B | C$

$A|B,C = A|C$

(i) P(B,C) = P(B|C)*P(C) = 0.2*0.4 = 0.08
(ii) P(⌐A|B)
P(A,B) + P(⌐A,B) =1
P(A|B)*P(B)+P(⌐A|B)*P(B)=1
P(⌐A|B)*P(B)=1- P(A|B)*P(B)
P(⌐A|B)=1- P(A|B)=1-0.3 =0.7
(iii) P(A,B|C)
If $X \perp Y| Z$ , $P( X, Y|Z )= P( X|Z) * P(Y|Z)$
P(A|C)*P(B|C) = 0.5*0.2 = 0.1
(iv) P(A|B,C)
If $X \perp Y|Z$, $P(X|Y, Z) = P(X|Z)$
$P(A|C)=0.5$

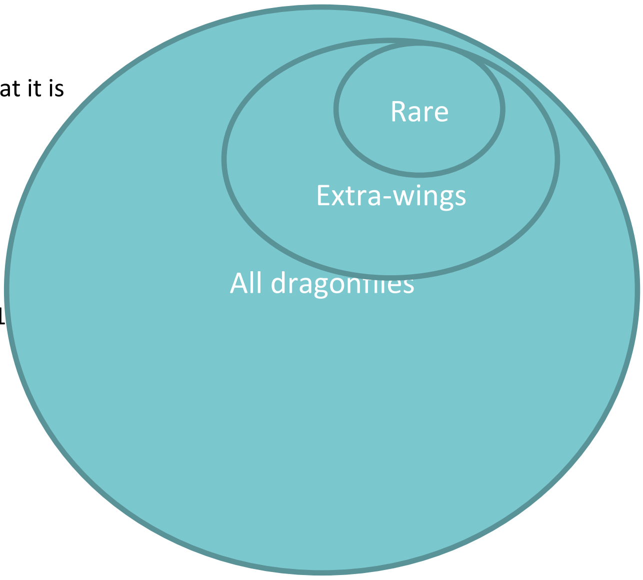
(v) P(A,B,C)
P(A,B)=P(A|B)*P(B)
P(A,B,C)=P(A|B,C)*P(B,C)=P(A|B,C)*P(B|C)*P(C)=0.5*0.2*0.4 = 0.04

## Question 3

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% 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?

P(R) = 0.003, probability of dragonfly belongs to rare spices
P(⌐R) = 1 – 0.003=0.997

$P(W|R) = 1$ – all the rare species have extra wings
$P(W|\neg R) = 0.001$ -probability extra wings for non rare
$P(\neg R\ |W) = P(\neg R)*P(W/\neg R)=0.001*0.997 = 0.000997$ – probability that it is Not rare given extra wings
$P(R|W)$-? – probability that the dragonfly belongs to rare if
Solution:
$P(W) = P(R) + P(\neg R)*P(W|\neg R) = 0.003+0.997*0.001= 0.003997$
$P(R,W) = P(W)*P(R|W)=P(R)*P(W|R)$
$P(W)*P(R|W)=P(R)*P(W|R)$
$P(R|W) = P(R)*P(W|R)/(P(W) = 0.003*1/0.003997 = 0.750562922191$




Answer:
If dragonfly has extra wings 75.05% that it belongs to rare species

## Question 4

(i) **If $P(A|B,C) = P(B|A,C)$, then $P(A|C) = P(B|C)$**

$P(A,B,C)=P(A|B,C)*P(B,C) = P\ (B|A,C)*P(A,C) = P(A|B,C)*P(B|C)*P(C) = P\ (B|A,C)*P(A|C)*A(C)$
$P(A|B,C)*P(B|C)*P(C) = P\ (B|A,C)*P(A|C)*P(C)$
The last equation shows:
$P(C)$ are equal anyway we could get rid of them
$P(A|B,C)*P(B|C) = P\ (B|A,C)*P(A|C)$
The highlighted pieces of the equation are equal to the condition above. So it is proved

(ii) If $P(A|B,C) = P(A)$, then $P(B,C|A) = P(B,C)$
$P(A,B,C)=P(A|B,C)*P(B,C) = P(B,C|A)*P(A)$
$P(A|B,C)*P(B,C) = P(B,C|A)*P(A)$
Proved.

(iii) If $P(A,B|C) = P(A|C) * P(B|C)$, then $P(A|B,C) = P(A|C)$
if $P(A|B,C) = P(A|C)$ it means that $A\perp B|C$
then
$P(A,B|C)= P(A|C)*P(B|C)$

# Part 2: Naive Bayes Method

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

Feature 0 : Probabilities for spam, if true = 0.67 , if false = 0.35 , for non spam, if true = 0.36 , for non spam, if false = 0.65
Feature 1 : Probabilities for spam, if true = 0.6 , if false = 0.42 , for non spam, if true = 0.58 , for non spam, if false = 0.43
Feature 2 : Probabilities for spam, if true = 0.46 , if false = 0.56 , for non spam, if true = 0.35 , for non spam, if false = 0.66
Feature 3 : Probabilities for spam, if true = 0.62 , if false = 0.4 , for non spam, if true = 0.4 , for non spam, if false = 0.61
Feature 4 : Probabilities for spam, if true = 0.5 , if false = 0.52 , for non spam, if true = 0.34 , for non spam, if false = 0.67
Feature 5 : Probabilities for spam, if true = 0.37 , if false = 0.65 , for non spam, if true = 0.47 , for non spam, if false = 0.53
Feature 6 : Probabilities for spam, if true = 0.79 , if false = 0.23 , for non spam, if true = 0.51 , for non spam, if false = 0.5
Feature 7 : Probabilities for spam, if true = 0.77 , if false = 0.25 , for non spam, if true = 0.35 , for non spam, if false = 0.65
Feature 8 : Probabilities for spam, if true = 0.35 , if false = 0.67 , for non spam, if true = 0.25 , for non spam, if false = 0.76
Feature 9 : Probabilities for spam, if true = 0.67 , if false = 0.35 , for non spam, if true = 0.29 , for non spam, if false = 0.71
Feature 10 : Probabilities for spam, if true = 0.67 , if false = 0.35 , for non spam, if true = 0.59 , for non spam, if false = 0.42
Feature 11 : Probabilities for spam, if true = 0.79 , if false = 0.23 , for non spam, if true = 0.34 , for non spam, if false = 0.67

### 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.*

##################################################
Instance 0
Probability of spam for [1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0] is 4.563570134160372e-06
Probability of non spam for [1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0] is 0.0004954267164176627
Result: NON SPAM
##################################################
Instance 1
Probability of spam for [0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1] is 7.210346867817725e-05
Probability of non spam for [0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1] is 4.5438790152493946e-05
Result: SPAM
##################################################
Instance 2
Probability of spam for [1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1] is 0.0002337624408575223
Probability of non spam for [1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1] is 0.0001395576654071092
Result: SPAM
##################################################
Instance 3
Probability of spam for [0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0] is 7.663585265070244e-06
Probability of non spam for [0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0] is 0.0006466481819829653
Result: NON SPAM
##################################################
Instance 4
Probability of spam for [1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1] is 7.720274851187897e-05
Probability of non spam for [1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1] is 0.00010025583637879426
Result: NON SPAM
##################################################
Instance 5
Probability of spam for [1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1] is 7.434338745588349e-05
Probability of non spam for [1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1] is 5.026186824410613e-05
Result: SPAM
##################################################
Instance 6
Probability of spam for [0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0] is 5.124916761793713e-06
Probability of non spam for [0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0] is 0.0003540527014157242
Result: NON SPAM
##################################################
Instance 7
Probability of spam for [0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1] is 8.117842746585807e-05
Probability of non spam for [0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1] is 0.00042357652476948584
Result: NON SPAM
##################################################
Instance 8
Probability of spam for [1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1] is 0.0002337624408575223
Probability of non spam for [1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1] is 4.10847332740555e-05

Result: SPAM
####################################################
Instance 9
Probability of spam for [1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0] is 2.8353542253659128e-05
Probability of non spam for [1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0] is 0.0007338396327966039
Result: NON SPAM

### 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.*

The words identified spam normally not separate and it could be the better effect if look at them in groups. So probability of one word indicated spam might be under the influence of probability of another. So they are more likely to be clustered in SPAM emails.

To calculate probabilities in Naïve Bayes for previous case the following formula was used:

P(Spam|f1 ,f2…)= P(f1,f2..|Slam)*P(Spam)=P(f1|Spam)*P(f2|Spam)…*P(Spam)

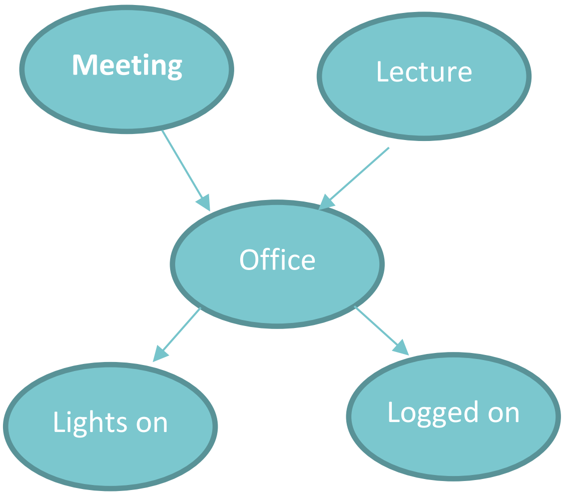
This transformation could be done only on the assumption that all the features are independent.(P(A,B) = P(A)*P(B))

If they are not independent this formula will not be valid for this case.

# Part 3: Building Bayesian Network [30 marks]

1.

Nodes: Meeting, Lecture, In Office, Lights on, Log out
Domains: True and False for each




| Meeting | P(M) |
|---|---|
| M | 0.7 |
| ⌐M | 0.3 |

| Lecture | P(L) |
|---|---|
| L | 0.6 |
| ⌐L | 0.4 |

| Meeting | Lecture | P(O\|M,L) |
|---|---|---|
| M | L | 0.95 |
| M | ⌐L | 0.75 |
| ⌐M | L | 0.8 |
| ⌐M | ⌐L | 0.06 |

| Light on | Office | P(Li\|O) |
|---|---|---|
| Li | O | 0.5 |
| Li | ⌐O | 0.02 |
| ⌐Li | O | 0.5 |
| ⌐Li | ⌐O | 0.98 |

| Log on | Office | P(Lo\|O) |
|---|---|---|
| Lo | O | 0.8 |
| Lo | ⌐O | 0.2 |
| ⌐Lo | O | 0.2 |
| ⌐Lo | ⌐O | 0.8 |

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

|M|-1+|L|-1+|M|*|L|*(|O|-1)+|O|*(|M|-1)+|O|*(|L|-1) = 2-1+2-1+2*2*1+2*1+2*1 = 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.*

L⊥M
Lo⊥Li
Li⊥M,L|O
Lo⊥M,L|O

P(L, ⌐M,O,Lo, ⌐Li) = P(L=1,M=0,O=1,Lo=1,Li=0) = P(L)*P(⌐M)*P(O|L, ⌐M)*P(⌐Li)*P(Lo)

P(O) = P(O|ML)*P(ML)+ P(O|⌐ML)*P(⌐ML)+ P(O|M⌐L)*P(M⌐L)+
P(O|⌐M⌐L)*P(⌐M⌐L)=0.95*0.7*0.6+0.8*0.3*0.6+0.75*0.7*0.4+0.06*0.4*0.3=0.7602
P(⌐O) = 1-P(O) = 1- 0.7602=0.2398
P(Li) = P(Li|O)*P(O)+P(Li|⌐O)*P(⌐O) = 0.5*0.7602+0.02*0.2398= 0.384896
P(⌐Li) = 1-P(Li)= 1=0.615104
P(Lo) = P(Lo|O)*P(O)+P(Lo|⌐O)*P(⌐O) = 0.8*0.7602+0.2*0.2398=0.65612

P(L, ⌐M,O,Lo, ⌐Li) = P(L)*P(⌐M)*P(O|L, ⌐M)*P(⌐Li)*P(Lo)=0.6*0.3*0.8*0.615*0.656=0.05809536

Answer 0.05809536

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

Calculated for previous question
P(O) = P(O|ML)*P(ML)+ P(O|⌐ML)*P(⌐ML)+ P(O|M⌐L)*P(M⌐L)+
P(O|⌐M⌐L)*P(⌐M⌐L)=0.95*0.7*0.6+0.8*0.3*0.6+0.75*0.7*0.4+0.06*0.4*0.3=0.7602
Answer: 0.7602

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

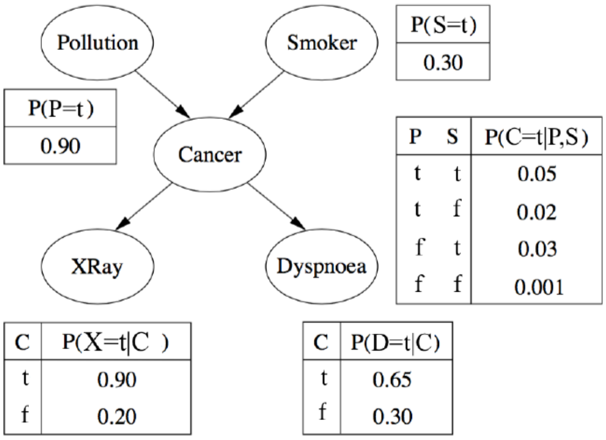
P(⌐Li , Lo |O) - ?
P(⌐Li , Lo |O) = P(⌐Li |O)* P(Lo |O) = 0.5*0.8= 0.4

### 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 Rachel's light is on ?*

P(Li|Lo)-?

As we know Lo⊥Li
It means that according to independence rule P(Li|Lo)=P(Li)
So it will not give any effect of P(Li) and on students believe.

# Part 4: Inference in Bayesian Networks (Did not finish)




***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***

P(P|x) = α*P(P|x)*1/P(x)
P(P|x) = P(P,S,C,D,x)=P(P)*P(S)*P(C|P,S)*P(x|C)*P(D|C)

X-evidence
P – query nodes
C,S ,D– hidden nodes

***(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***

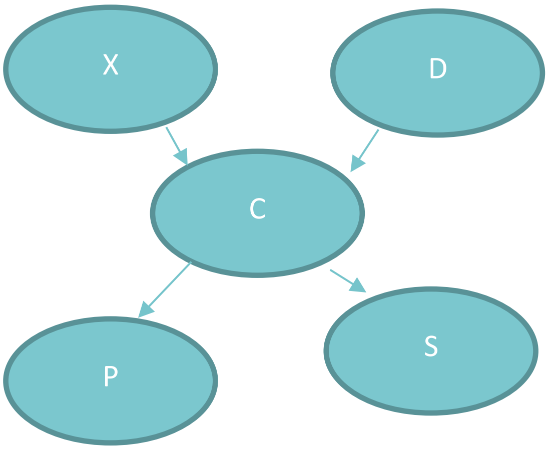
P(P)*P(S)*P(C|P,S)*P(x|C)*P(D|C)
f1 f2 f3 f4 f5
f1(P)*f2(S)*f3(C|P,S)*f4(C)*f5(D|C)
f6 = P(S)*P(C|P,S) (f2 and f3)
f1(P)*f6(C,P)*F4(C)*F5(C)
f7 = f6(C,P)*F4(C)*f5(D|C)
f1(P)*f7(P,D)

***(iii) report the probability,***
***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.***

X⊥D,
P⊥S
X⊥P,S|C
D⊥P,S|C
X⊥D|C
D⊥X|C

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.



# Part 5: Bayesian Network: Applications

### *(i) Clearly de_ne the random variables and their domains.*

Variables: Logic Problems, Wrong formula, Typo, Correct result , B result
Domains: True and False
Variable: Points for assignment
Domain: A,B,C,D

### *(ii) Clearly describe their relationships (using plain language).*

Student has written the script for assignment.
There is probability in typos in his script 10%, logic errors 5%, wrong formulas 20%
The script will give correct result with only typos in 90% , with only wrong logic –30 % and only wrong formulas – 20%.
If there will be both typos and logic – 5%, typos and formulas -3%, formulas and logic -1%.
All together – 0%
The assignment will be graded A if something wrong with script in 5%, B - 20%, C-50% and D – 60%
If student has problem with formulas on this topic he could receive more than B with probability 60%
If he is not very accurate having typos more than B will be with probability 80% , with both 50% , and without any 90%

(i) Draw the Bayesian network that can reflect the described relationship.

| Typos | P(T) |
|---|---|
| T | 0.1 |
| ⌐T | 0.9 |

| Wrong formula | P(Wf) |
|---|---|
| Wf | 0.2 |
| ⌐Wf | 0.8 |

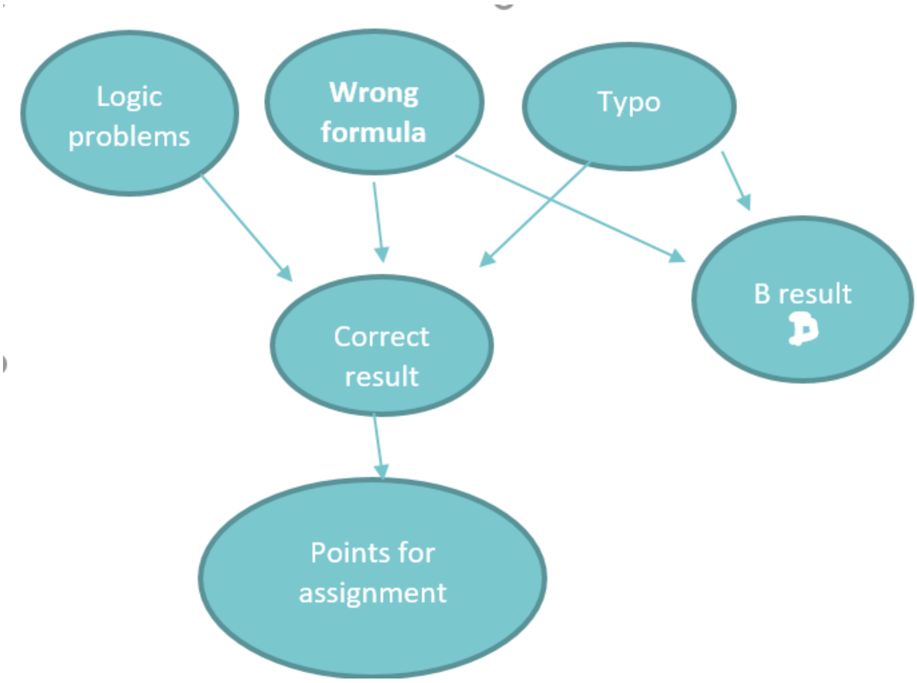
| Lp | P(Lp) |
|---|---|
| Lp | 0.05 |
| ⌐Lp | 0.95 |

| Wf | T | Lp | P(C\|Wf,T,Lp) |
|---|---|---|---|
| Wf | T | Lp | 1 |
| Wf | T | ⌐Lp | 0.3 |
| Wf | ⌐T | Lp | 0.9 |
| Wf | ⌐T | ⌐Lp | 0.05 |
| ⌐Wf | T | Lp | 0.2 |
| ⌐Wf | T | ⌐Lp | 0.01 |
| ⌐Wf | ⌐T | Lp | 0.03 |
| ⌐Wf | ⌐T | ⌐Lp | 0 |

| G | P(G\|C) |
|---|---|
| A | 0.05 |
| B | 0.2 |
| C | 0.5 |
| D | 0.6 |

| Wf | Typo | P(R\|Wf,T) |
|---|---|---|
| Wf | T | 0.9 |

| | | |
|---|---|---|
| Wf | ⌐T | 0.6 |
| ⌐Wf | T | 0.8 |
| ⌐Wf | ⌐T | 0.5 |




(ii) Write the factorisation of the Bayesian network.

P(Lp,Wf,T,C,P,R) = P(Lp)*P(Wf)*P(T)*P(C|Lp,Wf,T)*P(P|C)*P(R|T,Wf)

Lp⊥Wf⊥T

P⊥Lp,Wf,T|C

R⊥C|Lp,Wf,T