Cam Olssen
300492582
cam.olssen@gmail.com

1. Compute the table of the joint distribution *P(X, Y, Z)*

X	Y	Z	P(X, Y, Z)
0	0	0	0.0245
0	0	1	0.0105
0	1	0	0.063
0	1	1	0.252
1	0	0	0.273
1	0	1	0.117
1	1	0	0.052
1	1	1	0.208

Chain rule: P(X, Y, Z) = P(X | Y, Z) * P(Y | Z) * P(Z)
or, restructuring: P(Z,Y,X) = P(Z|Y,X) * P(Y|X) * P(X)
As Z and X are conditionally independent given Y, P(X|Y, Z) = P(X|Y) and P(Z|X,Y) = P(Z|Y)
Therefore P(X,Y,Z) = P(Z|Y)*P(Y|X)*P(X)
Therefore P(X=0, Y=0, Z=0) = 0.70*0.10*0.35 = 0.0245
P(X=0,Y=0,Z=1) = 0.35*0.10*0.30 = 0.0105
P(X=0,Y=1,Z=0 = 0.35*0.90*0.70 = 0.063
P(X=0,Y=1,Z=1)=0.35*0.90*0.80 = 0.252
P(X=1,Y=0,Z=0)=0.65*0.60*0.70) = 0.273
P(X=1,Y=0,Z=1)=0.65*0.60*0.30)=0.117
P(X=1,Y=1,Z=0)=0.65*0.40*0.20=0.052
P(X=1,Y=1,Z=1)=0.65*0.40*0.80=0.208
Normalisation rule: All possibilities will sum up to 1
0.208+0.052+0.117+0.273+0.252+0.063+0.0105+0.0245 = 1

2. Create the full joint probability table of X and Y

	X=0	X=1	Sum
Y=0	0.035	0.39	0.425
Y=1	0.315	0.26	0.575
Sum	0.35	0.65	1

Chain rule again: P(X,Y) = P(X|Y)*P(Y)
Restructuring this, P(Y, X) = P(Y|X)*P(X)
Therefore P(X=0,Y=0)=0.35*0.10=0.035
P(X=0,Y=1)=0.35*0.90 = 0.315
P(X=1,Y=0)=0.65*0.60=0.39
P(X=1,Y=1)=0.65*0.40=0.26
Sum rule: Probability of an event is equal to the sum of all joint probabilities with another event. We already have P(X) and can see this lines up
Normalisation rule shows that all possibilities add up to 1

3. From the above joint probability table of X, Y and Z, calculate the following probabilities.
   a. P(Z=0). By sum rule we know that P(Z=0) is the sum of all joint probabilities where Z=0
   Therefore P(Z=0) = 0.0245+0.063+0.273+0.052 = 0.4125

Cam Olssen
300492582
cam.olssen@gmail.com

b. P(X=0, Z=0). Sum rule applies again here for P(X=0,Z=0) = 0.0245+0.063 = 0.0875
c. P(X=1, Y=0 | Z=1) = P(X=1,Y=0,Z=1)/P(Z=1). Due to normalisation rule we know that P(Z=1) = 0.5875 (1-0.4125), and that P(X=1,Y=0,z=1) = 0.117 from the joint probability table.
Therefore P(X=1, Y=0 | Z=1) = 0.117/0.5875 = 0.199148936.
d. P(X=0 | Y=0, Z=0) = P(X=0,Y=0,Z=0)/P(Y=0,Z=0)
P(Y=0, Z=0) = P(Y=0)*P(Z=0) = 0.425*0.4125 = 0.1753125
P(X=0,Y=0,Z=0) = 0.0245 from the joint probability table
Therefore, P(X=0| Y=0, Z=0) = 0.0245/0.1753125 = 0.139750446

4. Consider 3 Boolean variables A, B, and C (can take t or f). We have the following probabilities.
P(B=t) = 0.7
P(C=t) = 0.4
P(A=t|B=t) = 0.3
P(A=t|C=t) = 0.5
P(B=t|C=t) = 0.2

We also know that A and B are conditionally independent given C. Calculate the following probabilities.
Chain rule means that P(A,B,C) = P(A|B,C)*P(B|C)*P(C)
a. P(B=t, C=t) = 0.28
P(B=t)*P(C=5) = 0.7*0.4 = 0.28

b. P(A=f|B=t) = 0.4
P(A=t|B=t) = 0.3 and P(B=t) = 0.7
Sum rule means that P(A=t|B=t)+P(A=f|B=t)=P(B=t).
Therefore P(A=f|B=t) = P(B=t)-P(A=t|B=t) = 0.4

c. P(A=t, B=t | C=t) = P(A=t,B=t,C=t)/P(C=t) = 0.1.
As A and B are conditionally independent given C, P(A|B,C) = P(A|C) and B(C|A,B) = P(C|A)
Therefore by chain rule P(A=t,B=t,C=t) = P(A=t|C=t)*P(B=t|C=t)*P(C=t)
or 0.5*0.2*0.4 = 0.04
Therefore, P(A=t, B=t | C=t) = 0.04/0.4 = 0.1

d. P(A=t|B=t,C=t) = 0.5
As A and B are conditionally independent given C, P(A=t|B=t,C=t) = P(A=t|C=t) which we already know to be 0.5

e. P(A=t,B=t,C=t) = P(A|B,C)*P(B|C)*P(C) = 0.5*0.2*0.4 = 0.04 by chain rule.

Part 2 – Naïve Bayes Method:

1: Conditional probabilities for each feature X

inv-nodes=18-20|class=recurrence-events=0.01098901098901099
breast-quad=central|class=recurrence-events=0.060240963855421686

Cam Olssen
300492582
cam.olssen@gmail.com

breast=left|class=recurrence-events=0.55
breast-quad=left_up|class=recurrence-events=0.30120481927710846
menopause=ge40|class=no-recurrence-events=0.4583333333333333
menopause=lt40|class=recurrence-events=0.012345679012345678
tumor-size=25-29|class=recurrence-events=0.2111111111111111
age=90-99|class=no-recurrence-events=0.005050505050505051
age=10-19|class=no-recurrence-events=0.005050505050505051
tumor-size=45-49|class=recurrence-events=0.022222222222222223
tumor-size=25-29|class=no-recurrence-events=0.15920398009950248
node-caps=yes|class=no-recurrence-events=0.1256544502617801
inv-nodes=27-29|class=no-recurrence-events=0.0049504950495049506
deg-malig=1|class=recurrence-events=0.1111111111111111
tumor-size=20-24|class=recurrence-events=0.15555555555555556
age=90-99|class=recurrence-events=0.011494252873563218
age=20-29|class=recurrence-events=0.011494252873563218
breast-quad=central|class=no-recurrence-events=0.08762886597938144
age=80-89|class=no-recurrence-events=0.005050505050505051
breast-quad=left_low|class=no-recurrence-events=0.36597938144329895
tumor-size=15-19|class=no-recurrence-events=0.11442786069651742
age=40-49|class=recurrence-events=0.3103448275862069
inv-nodes=33-35|class=recurrence-events=0.01098901098901099
tumor-size=40-44|class=recurrence-events=0.07777777777777778
inv-nodes=21-23|class=recurrence-events=0.01098901098901099
inv-nodes=15-17|class=no-recurrence-events=0.019801980198019802
inv-nodes=15-17|class=recurrence-events=0.04395604395604396
age=60-69|class=recurrence-events=0.19540229885057472
tumor-size=5-9|class=recurrence-events=0.011111111111111112
breast-quad=right_low|class=recurrence-events=0.08433734939759036
breast=right|class=no-recurrence-events=0.49214659685863876
age=50-59|class=recurrence-events=0.25287356321839083
tumor-size=10-14|class=recurrence-events=0.022222222222222223
age=30-39|class=no-recurrence-events=0.1111111111111111
tumor-size=40-44|class=no-recurrence-events=0.0845771144278607
tumor-size=45-49|class=no-recurrence-events=0.014925373134328358
deg-malig=1|class=no-recurrence-events=0.2916666666666667
breast-quad=right_up|class=recurrence-events=0.1686746987951807
inv-nodes=33-35|class=no-recurrence-events=0.0049504950495049506
breast-quad=left_low|class=recurrence-events=0.3855421686746988
irradiat=no|class=no-recurrence-events=0.8429319371727748
inv-nodes=36-39|class=recurrence-events=0.01098901098901099
node-caps=no|class=no-recurrence-events=0.8743455497382199
tumor-size=55-59|class=recurrence-events=0.011111111111111112
inv-nodes=9-11|class=recurrence-events=0.06593406593406594
age=20-29|class=no-recurrence-events=0.010101010101010102
tumor-size=50-54|class=no-recurrence-events=0.024875621890547265
tumor-size=35-39|class=no-recurrence-events=0.05970149253731343
tumor-size=50-54|class=recurrence-events=0.044444444444444446
inv-nodes=0-2|class=recurrence-events=0.4725274725274725
node-caps=no|class=recurrence-events=0.6
inv-nodes=3-5|class=no-recurrence-events=0.08415841584158416

Cam Olssen
300492582
cam.olssen@gmail.com

breast=left|class=no-recurrence-events=0.5078534031413613
inv-nodes=30-32|class=no-recurrence-events=0.0049504950495049506
age=10-19|class=recurrence-events=0.011494252873563218
age=50-59|class=no-recurrence-events=0.3282828282828283
inv-nodes=36-39|class=no-recurrence-events=0.0049504950495049506
tumor-size=20-24|class=no-recurrence-events=0.17412935323383086
tumor-size=0-4|class=recurrence-events=0.022222222222222223
tumor-size=35-39|class=recurrence-events=0.08888888888888889
breast-quad=right_up|class=no-recurrence-events=0.10824742268041238
breast-quad=left_up|class=no-recurrence-events=0.34536082474226804
inv-nodes=12-14|class=recurrence-events=0.03296703296703297
menopause=ge40|class=recurrence-events=0.38271604938271603
age=40-49|class=no-recurrence-events=0.31313131313131315
tumor-size=5-9|class=no-recurrence-events=0.024875621890547265
inv-nodes=3-5|class=recurrence-events=0.17582417582417584
tumor-size=30-34|class=no-recurrence-events=0.1691542288557214
tumor-size=55-59|class=no-recurrence-events=0.004975124378109453
inv-nodes=18-20|class=no-recurrence-events=0.0049504950495049506
tumor-size=15-19|class=recurrence-events=0.07777777777777778
breast-quad=right_low|class=no-recurrence-events=0.09278350515463918
inv-nodes=21-23|class=no-recurrence-events=0.0049504950495049506
inv-nodes=12-14|class=no-recurrence-events=0.009900990099009901
deg-malig=3|class=recurrence-events=0.5308641975308642
menopause=premeno|class=no-recurrence-events=0.5104166666666666
inv-nodes=24-26|class=no-recurrence-events=0.0049504950495049506
deg-malig=2|class=no-recurrence-events=0.5104166666666666
inv-nodes=6-8|class=no-recurrence-events=0.039603960396039604
node-caps=yes|class=recurrence-events=0.4
age=70-79|class=no-recurrence-events=0.030303030303030304
irradiat=yes|class=recurrence-events=0.3875
age=80-89|class=recurrence-events=0.011494252873563218
tumor-size=0-4|class=no-recurrence-events=0.03980099502487562
age=70-79|class=recurrence-events=0.011494252873563218
inv-nodes=30-32|class=recurrence-events=0.01098901098901099
inv-nodes=24-26|class=recurrence-events=0.02197802197802198
menopause=lt40|class=no-recurrence-events=0.03125
age=30-39|class=recurrence-events=0.1839080459770115
tumor-size=30-34|class=recurrence-events=0.25555555555555554
breast=right|class=recurrence-events=0.45
irradiat=no|class=recurrence-events=0.6125
inv-nodes=6-8|class=recurrence-events=0.12087912087912088
age=60-69|class=no-recurrence-events=0.1919191919191919
tumor-size=10-14|class=no-recurrence-events=0.12935323383084577
inv-nodes=0-2|class=no-recurrence-events=0.7970297029702971
inv-nodes=9-11|class=no-recurrence-events=0.01485148514851485
irradiat=yes|class=no-recurrence-events=0.15706806282722513
menopause=premeno|class=recurrence-events=0.6049382716049383
deg-malig=2|class=recurrence-events=0.35802469135802467
inv-nodes=27-29|class=recurrence-events=0.01098901098901099
deg-malig=3|class=no-recurrence-events=0.19791666666666666

Cam Olssen
300492582
cam.olssen@gmail.com

2: Class probabilities (P(Y=y)) for each class label Y=y
recurrence-events=0.2936802973977695
no-recurrence-events=0.7063197026022305

3: Instance 1 recurrence-events score:7.096642912782161E-6
Instance 1 no-recurrence-events score:4.017731924138001E-6
Final label prediction: recurrence-events | Correct label: no-recurrence-events

Instance 2 recurrence-events score:2.6021024013534593E-5
Instance 2 no-recurrence-events score:3.3365875538186835E-4
Final label prediction: no-recurrence-events | Correct label: no-recurrence-events

Instance 3 recurrence-events score:9.715041077797865E-7
Instance 3 no-recurrence-events score:4.707378797592419E-5
Final label prediction: no-recurrence-events | Correct label: no-recurrence-events

Instance 4 recurrence-events score:1.0219165794406317E-5
Instance 4 no-recurrence-events score:1.5163354504655905E-4
Final label prediction: no-recurrence-events | Correct label: no-recurrence-events

Instance 5 recurrence-events score:1.9205993914751727E-6
Instance 5 no-recurrence-events score:4.34523614378287E-6
Final label prediction: no-recurrence-events | Correct label: no-recurrence-events

Instance 6 recurrence-events score:3.880403296274196E-5
Instance 6 no-recurrence-events score:6.000123072810078E-4
Final label prediction: no-recurrence-events | Correct label: no-recurrence-events

Instance 7 recurrence-events score:7.415569159796725E-5
Instance 7 no-recurrence-events score:2.071279390811688E-4
Final label prediction: no-recurrence-events | Correct label: no-recurrence-events

Instance 8 recurrence-events score:8.934762413310772E-6
Instance 8 no-recurrence-events score:3.1504865826699285E-4
Final label prediction: no-recurrence-events | Correct label: recurrence-events

Instance 9 recurrence-events score:6.483884504037111E-5
Instance 9 no-recurrence-events score:3.9123945169671666E-5
Final label prediction: recurrence-events | Correct label: recurrence-events

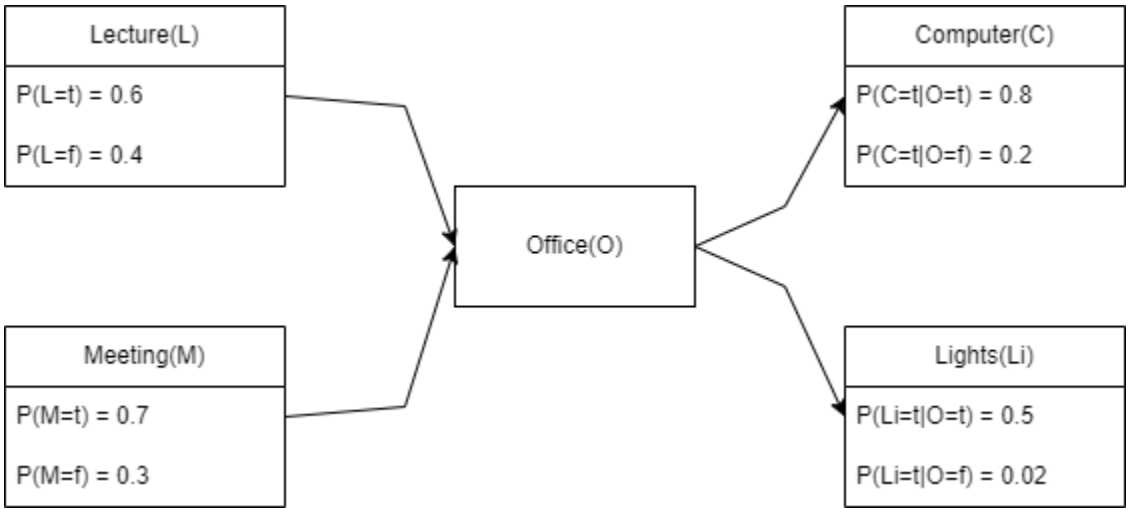
Instance 10 recurrence-events score:5.283165151437645E-5
Instance 10 no-recurrence-events score:4.1017039290784815E-5
Final label prediction: recurrence-events | Correct label: recurrence-events

Part 3: Constructing Bayesian Network

Meeting(M)	**Lecture(L)**	**P(O=t|M,L)**
t	**t**	0.95
t	**f**	0.75
f	**t**	0.8

Cam Olssen
300492582
cam.olssen@gmail.com

f	**f**	0.06



2. There are 10 free parameters in this Bayesian network.

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(L=t,M=f,O=t,C=t,Li=f) = P(L=t)*P(M=f)*P(O=t | P(M=f,L=t)*P(C=t|O=t)+P(Li=f|O=t) = 0.6*0.3*0.8*0.8*0.5 = 0.0576

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

P(O=t). By the sum rule, this is equal to the sum of all joint probabilities where O=t.
P(O=t) = P(L=t,M=t,O=t)+P(L=f,M=t,O=t)+P(L=t,M=f,O=t)+P(L=f,M=f,O=t)
Because L and M are independent
P(O=t,L=t,M=t) = P(O=t|L=t,M=t)*P(M=t*P(L=t)=0.95*0.6*0.7=0.399
P(O=t,L=t,M=f) = P(O=t|L=t,M=f)*P(L=t)*P(M=f)=0.8*0.6*0.3=0.144
P(O=t,L=f,M=t) = P(O=t|L=f,M=t)*P(L=f)*P(M=t)=0.75*0.4*0.7=0.21
P(O=t,L=f,M=f) = P(O=f|L=f,M=f)*P(L=f)*P(M=f)=0.06*0.4*0.3=0.0072

0.399+0.144+0.21+0.0072 = 0.7602

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

P(C=t, Li=f | O=t) = P(C=t | O=t) * P(Li=f | O=t) = 0.8*0.5 = 0.4