COMP307 - Assignment 3
greenthom - 300536064

Part 1: Naïve Bayes Method

1. **The conditional probabilities P(Xi = xi | Y = y) for each feature Xi (e.g. age), it's possible value xi (e.g. 10-19), and each class label Y = y (y can be no-recurrence-events or recurrence-events).**

Saved in: nb_train_output.txt

P(age = 30-39 | no-recurrence-events) = 0.11282051282051282
P(age = 40-49 | no-recurrence-events) = 0.31794871794871793
P(age = 60-69 | no-recurrence-events) = 0.19487179487179487
P(age = 50-59 | no-recurrence-events) = 0.3333333333333333
P(age = 70-79 | no-recurrence-events) = 0.03076923076923077
P(age = 20-29 | no-recurrence-events) = 0.010256410256410256
P(menopause = premeno | no-recurrence-events) = 0.5104166666666666
P(menopause = ge40 | no-recurrence-events) = 0.4583333333333333
P(menopause = lt40 | no-recurrence-events) = 0.03125
P(tumor-size = 30-34 | no-recurrence-events) = 0.17
P(tumor-size = 20-24 | no-recurrence-events) = 0.175
P(tumor-size = 15-19 | no-recurrence-events) = 0.115
P(tumor-size = 0-4 | no-recurrence-events) = 0.04
P(tumor-size = 25-29 | no-recurrence-events) = 0.16
P(tumor-size = 50-54 | no-recurrence-events) = 0.025
P(tumor-size = 10-14 | no-recurrence-events) = 0.13
P(tumor-size = 40-44 | no-recurrence-events) = 0.085
P(tumor-size = 35-39 | no-recurrence-events) = 0.06
P(tumor-size = 5-9 | no-recurrence-events) = 0.025
P(tumor-size = 45-49 | no-recurrence-events) = 0.015
P(inv-nodes = 0-2 | no-recurrence-events) = 0.8214285714285714
P(inv-nodes = 6-8 | no-recurrence-events) = 0.04081632653061224
P(inv-nodes = 9-11 | no-recurrence-events) = 0.015306122448979591
P(inv-nodes = 3-5 | no-recurrence-events) = 0.08673469387755102
P(inv-nodes = 15-17 | no-recurrence-events) = 0.02040816326530612
P(inv-nodes = 12-14 | no-recurrence-events) = 0.01020408163265306
P(inv-nodes = 24-26 | no-recurrence-events) = 0.00510204081632653
P(node-caps = no | no-recurrence-events) = 0.8743455497382199
P(node-caps = yes | no-recurrence-events) = 0.1256544502617801
P(deg-malig = 3 | no-recurrence-events) = 0.19791666666666666
P(deg-malig = 2 | no-recurrence-events) = 0.5104166666666666
P(deg-malig = 1 | no-recurrence-events) = 0.2916666666666667
P(breast = left | no-recurrence-events) = 0.5078534031413613

P(breast = right | no-recurrence-events) = 0.49214659685863876
P(breast-quad = left_low | no-recurrence-events) = 0.36597938144329895
P(breast-quad = right_up | no-recurrence-events) = 0.10824742268041238
P(breast-quad = left_up | no-recurrence-events) = 0.34536082474226804
P(breast-quad = right_low | no-recurrence-events) = 0.09278350515463918
P(breast-quad = central | no-recurrence-events) = 0.08762886597938144
P(irradiat = no | no-recurrence-events) = 0.8429319371727748
P(irradiat = yes | no-recurrence-events) = 0.15706806282722513
P(age = 30-39 | recurrence-events) = 0.19047619047619047
P(age = 40-49 | recurrence-events) = 0.32142857142857145
P(age = 60-69 | recurrence-events) = 0.20238095238095238
P(age = 50-59 | recurrence-events) = 0.2619047619047619
P(age = 70-79 | recurrence-events) = 0.011904761904761904
P(age = 20-29 | recurrence-events) = 0.011904761904761904
P(menopause = premeno | recurrence-events) = 0.6049382716049383
P(menopause = ge40 | recurrence-events) = 0.38271604938271603
P(menopause = lt40 | recurrence-events) = 0.012345679012345678
P(tumor-size = 30-34 | recurrence-events) = 0.25842696629213485
P(tumor-size = 20-24 | recurrence-events) = 0.15730337078651685
P(tumor-size = 15-19 | recurrence-events) = 0.07865168539325842
P(tumor-size = 0-4 | recurrence-events) = 0.02247191011235955
P(tumor-size = 25-29 | recurrence-events) = 0.21348314606741572
P(tumor-size = 50-54 | recurrence-events) = 0.0449438202247191
P(tumor-size = 10-14 | recurrence-events) = 0.02247191011235955
P(tumor-size = 40-44 | recurrence-events) = 0.07865168539325842
P(tumor-size = 35-39 | recurrence-events) = 0.0898876404494382
P(tumor-size = 5-9 | recurrence-events) = 0.011235955056179775
P(tumor-size = 45-49 | recurrence-events) = 0.02247191011235955
P(inv-nodes = 0-2 | recurrence-events) = 0.5058823529411764
P(inv-nodes = 6-8 | recurrence-events) = 0.12941176470588237
P(inv-nodes = 9-11 | recurrence-events) = 0.07058823529411765
P(inv-nodes = 3-5 | recurrence-events) = 0.18823529411764706
P(inv-nodes = 15-17 | recurrence-events) = 0.047058823529411764
P(inv-nodes = 12-14 | recurrence-events) = 0.03529411764705882
P(inv-nodes = 24-26 | recurrence-events) = 0.023529411764705882
P(node-caps = no | recurrence-events) = 0.6
P(node-caps = yes | recurrence-events) = 0.4
P(deg-malig = 3 | recurrence-events) = 0.5308641975308642
P(deg-malig = 2 | recurrence-events) = 0.35802469135802467
P(deg-malig = 1 | recurrence-events) = 0.1111111111111111
P(breast = left | recurrence-events) = 0.55
P(breast = right | recurrence-events) = 0.45
P(breast-quad = left_low | recurrence-events) = 0.3855421686746988
P(breast-quad = right_up | recurrence-events) = 0.1686746987951807

P(breast-quad = left_up | recurrence-events) = 0.30120481927710846
P(breast-quad = right_low | recurrence-events) = 0.08433734939759036
P(breast-quad = central | recurrence-events) = 0.060240963855421686
P(irradiat = no | recurrence-events) = 0.6125
P(irradiat = yes | recurrence-events) = 0.3875

2. **The class probabilities P(Y = y) for each class label Y = y**

From output file: nb_train_output.txt

P(no-recurrence-events) = 0.7063197026022305
P(recurrence-events) = 0.2936802973977695

3. **For each test instance, given the input vector X = [X1 = x1,...,X9 = x9], give the calculated: score(Y = no-recurrence-events, X), score(Y = recurrence-events, X), predicted class of the input vector.**

Saved in output file: nb_test_output.txt

Test Instance = 1
score(no-recurrence-events = 4.225449295343082e-06)
score(recurrence-events = 7.95733952513876e-06)

Test Instance = 2
score(no-recurrence-events = 0.00035090896541482136)
score(recurrence-events = 2.917691159217544e-05)

Test Instance = 3
score(no-recurrence-events = 4.950751020419909e-05)
score(recurrence-events = 1.0893302834424374e-06)

Test Instance = 4
score(no-recurrence-events = 0.00015947302312979033)
score(recurrence-events = 1.145856891619981e-05)

Test Instance = 5
score(no-recurrence-events = 4.569885534557127e-06)
score(recurrence-events = 2.153533950851044e-06)

Test Instance = 6
score(no-recurrence-events = 0.0006310330377608908)
score(recurrence-events = 4.351027225465415e-05)

Test Instance = 7
score(no-recurrence-events = 0.00021783648604782526)
score(recurrence-events = 8.314945855647813e-05)

Test Instance = 8
score(no-recurrence-events = 0.00033133672335758455)
score(recurrence-events = 1.0018390240701734e-05)

Test Instance = 9
score(no-recurrence-events = 4.114665927685032e-05)
score(recurrence-events = 7.270264415795743e-05)

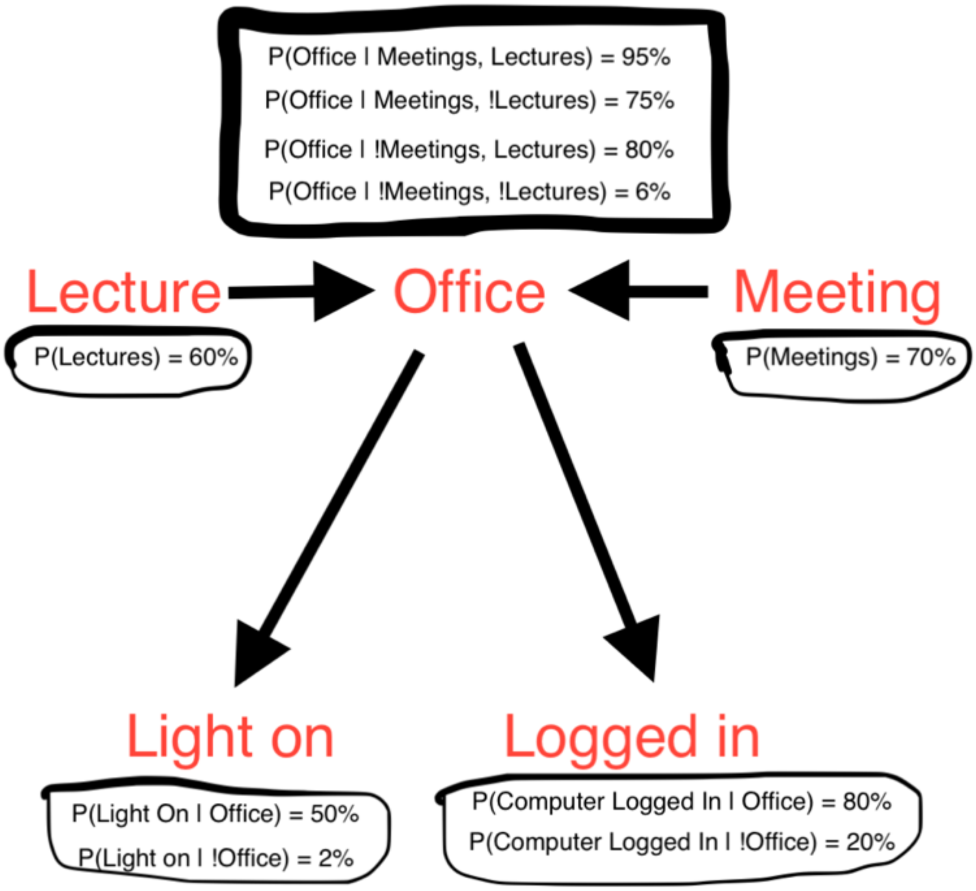
Test Instance = 10
score(no-recurrence-events = 4.313762666121404e-05)
score(recurrence-events = 5.923919153611345e-05)

Predicted Test Class 1 = recurrence-events, Actual Test Class = no-recurrence-events
Predicted Test Class 2 = no-recurrence-events, Actual Test Class = no-recurrence-events
Predicted Test Class 3 = no-recurrence-events, Actual Test Class = no-recurrence-events
Predicted Test Class 4 = no-recurrence-events, Actual Test Class = no-recurrence-events
Predicted Test Class 5 = no-recurrence-events, Actual Test Class = no-recurrence-events
Predicted Test Class 6 = no-recurrence-events, Actual Test Class = no-recurrence-events
Predicted Test Class 7 = no-recurrence-events, Actual Test Class = no-recurrence-events
Predicted Test Class 8 = no-recurrence-events, Actual Test Class = recurrence-events
Predicted Test Class 9 = recurrence-events, Actual Test Class = recurrence-events
Predicted Test Class 10 = recurrence-events, Actual Test Class = recurrence-events

ACCURACY = 80.0%

Part 2: Building Bayesian Network

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



2. **Calculate the number of free parameters in your Bayesian network.**

2 ①

Lecture

L	P(L)
T	0.6

~~Lecture~~ Meeting

M	P(M)
T	0.7

Office

L	M	P(O,L,M)
F	F	P(O|!L,!M)=0.6
F	T	P(O|!L,M)=0.75
T	F	P(O|L,!M)=0.8
T	T	P(O|L,M)=0.95

Light On

T	O	P(T,O)
T	T	P(T|O)=0.5
T	F	P(T|!O)=0.02

Logged In

C	O	P(C,O)
T	T	P(C|O)=0.8
T	F	P(C|!O)=0.2

Total free = 1(Lecture) + 1(Meeting) + 4(Office) + 2(Logged In) + 2(Light On)

=> 10 free parameters

3. **What is the joint probability that Eve has lectures, has no meetings, she is in her office and logged on her computer but with lights off.**

(3) Eve → has lectures & no meetings &
in office & logged on & no lights

→ P(L, !M, O, !T, C)

→ P(L) * P(!M) * P(O | !M, L) * P(O|C) *
P(!T|O)

→ 0.6 * 1-0.7 * 0.8 * 0.8 * 1-0.5

→ 0.6 * 0.3 * 0.8 * 0.8 * 0.5

= 0.0576

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

A Eve → in office

P(O) → P(O|M,L) * P(M) * P(L) + P(O|M,!L) *
P(M) * P(!L) + P(O|!M,L) * P(!M) * P(L) +
P(O|!M,!L) * P(!M) * P(!L)

→ (0.95 * 0.7 * 0.6) + (0.75 * 0.7 * 1-0.6) + (0.8 *
1-0.7 * 0.6) + (0.06 * 1-0.7 * 1-0.6)
→ (0.95 * 0.7 * 0.6) + (0.75 * 0.7 * 0.4) + (0.8
* 0.3 * 0.6) + (0.06 * 0.3 * 0.4)
→ 0.399 + 0.21 + 0.144 + 0.0072
= 0.7602

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

(5) Eve → in office & logged on & light off

$P(C, !T | O)$

$\rightarrow P(C|O) * P(!T|O)$

$\rightarrow 0.8 * 1 - 0.5$

$\rightarrow 0.8 * 0.5$

$= 0.4$