Fernando Mendoza
9637984

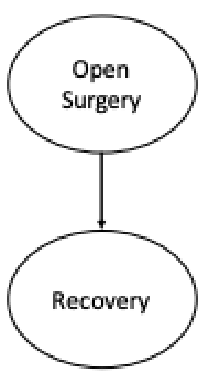
Worked together with Blake Johnson on the homework but only helped each other with concepts.

Problem 1

a. No it is not true that P(X|Y,W) = P(X|W) because X and Y are not conditionally independent given W. Therefore P(X|Y,W) != P(Y|W).
b. P(X|Y) = P(X)
c. Yes, because a node is conditionally independent of its non-descendants given its parent. In this case X has no parents so it is conditionally independent of V and Z. W on the other hand shares parent Y with V and is therefore independent of V and Z given Y.
d. No because W and V have common parent Y.
e. No because W and Z share common ancestor Y.
f. W is conditionally independent of all other nodes in the network given X,U,Y. Y is conditionally independent of all other nodes in the network given W and V.
g. P(U = 1, V = 1, W = 1, X = 0, Y = 0, Z = 1)
= P(X=0) * P(Y=0) * P(W=1| X=0,Y=0) * P(V=1|Y=0) * P(U=1|V=1,W=1) * P(Z=1|V=1)

Problem 2

a.

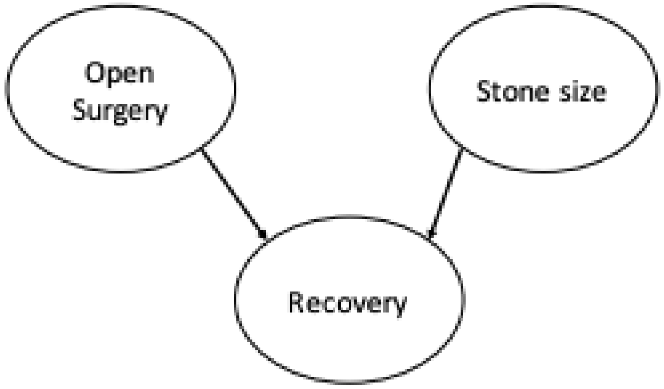


P(OS)

P(OS=true)	P(OS=false)
0.5	0.5

P(R|OS)

OS	P(R=true)	P(R=false)
T	0.78	0.22
F	0.83	0.17



P(SS)

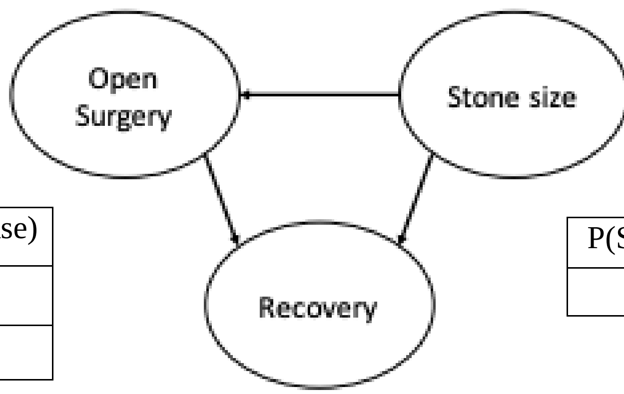
P(OS)

P(OS)	P(¬OS)
0.5	0.5

P(SS=small)	P(SS=large)
0.51	0.49

P(R|OS,SS)

OS	SS	P(R=true)	P(R=false)
true	small	0.93	0.07
true	large	0.73	0.27
false	small	0.87	0.13
false	large	0.69	0.31



P(OS|SS)

S	P(OS=true)	P(OS=false)
small	0.12	0.39
large	0.38	0.11

P(SS)

P(S=small)	P(S=large)
0.51	0.49

P(R|OS,SS)

OS	SS	P(R=true)	P(R=false)
true	small	0.93	0.07
true	large	0.73	0.27
false	small	0.87	0.13
false	large	0.69	0.31

b.

a. P(OS | R)
$= P(R, OS) / P(R)$
$= [P(R|OS)* P(OS)] / [\sum_i P(R|OS_i) * P(OS_i)]$
$= 0.49$

b. P(OS | R)
$= P(R,OS)/P(R)$
$= \sum_i P(R,OS,SS_i)/\sum_i \sum_j P(R,OS_i,SS_j)$
$= [\sum_i P(OS) * P(SS_i) * P(R|OS,SS_i)] / [\sum_i \sum_j P(OS_i) * P(SS_j) * P(R|OS_i,SS_j)]$
$= 0.52$

c. P(OS|R)
$= P(R,OS)/P(R)$
$=\sum_i P(R,OS,SS_i)/\sum_i \sum_j P(R,OS_i,SS_j)$
$=[\sum_i P(SS_i)* P(OS|SS_i)* P(R|OS,SS_i)] / [\sum_i \sum_j P(SS_j)* P(OS_i|SS_j)* P(R|OS_i,SS_j)]$
$= 0.48$

The answers are not consistent throughout models because they have different dependencies and therefore their CPT are different.

c. Model B: Open Surgery is marginally independent of Stone Size.
Model C: No marginal or conditional independence

d. Model C is the one that best fits the data because the size of the stone influences the recovery. Additionally, out of all open surgeries, most of the patients had large stones. Which helps defend the statement that the size of the stone influences the type of operation that happens.

Problem 3

a. $L(\text{Ø}: x[y]_i) =$
$= \prod_{i=1}^{n} f(x[y]_i :\mu_j,ø_j)$
$= \prod_{i=1}^{n} (ø_j^2\ 2\pi\ )^\wedge(-1/2)\ \ *\ e^\wedge(\ (-1/2) * (\ (x[y]_i - \mu_j) / ø_j)^\wedge 2\ )$ //carry out pi notation
$= (ø_j^2\ 2\pi\ )^\wedge(-n/2)\ \ *\ e^\wedge(-1/(2ø_j^2) * \sum_{i=1}^{n} (x[y]_i - \mu_j)^\wedge 2)$ //turns to sigma notation for exponents

b. $\ln[L(Ø{:}x_i)] =$

$= \ln[(ø_j^2\ 2\pi\ )\hat{}(-n/2)\ \ *\ e\hat{}(-1/(2ø_j^2)\ *\ \sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2)]$ //take log

$= \ln[(ø_j^2\ 2\pi\ )\hat{}(-n/2)] + \ln[e\hat{}(-1/(2ø_j^2)\ *\ \sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2)]$ //e^() cancels with ln

$= (-n/2)*\ln(ø_j^2\ 2\pi) - (1/(2ø_j^2)\ *\ \sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2))$

c. $\partial\mu/\mu\ \ [-(1/(2ø_j^2))\ *\ \sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2))]$

$= -\ (1/(2ø_j^2))\ *\ (-2)\ *\ \sum_{i=1}^{n} (x[y]_i - \mu_j)$ //take derivative with respect to μ

$= \sum_{i=1}^{n} (x[y]_i - \mu_j)\ /\ ø_j^2$ //simplify

$\sum_{i=1}^{n} (x[y]_i - \mu_j)\ /\ ø_j^2 = 0$ //equal equation to 0

$\sum_{i=1}^{n} (x[y]_i - \mu_j) = 0$ //multiply equation by $ø_j^2$

$\sum_{i=1}^{n} x[y]_i = n*\mu_j$ //add μ_j to opposite side and multiply by n

$\hat{u}_j = \sum_{i=1}^{n} x[y]_i\ /\ n$ //divide equation by n

$\partial ø/ø\ [\ (-n/2)*\ln(ø_j^2\ 2\pi)] + \partial ø/ø[\ -\sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2\ /\ (2ø_j^2)]$

$= [(-n/2)*((\ ø_j\ 4\pi)/(\ ø_j^2\ 2\pi))] + [\sum_{i=1}^{n} ((x[y]_i - \mu_j)\hat{}2)*2/(2ø_j^3)]]$ //take derivative of function

$= [(-n/ø_j)\ ] + [\sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2\ /\ ø_j^3]$ // simplify

$[(-n/ø_j)\ ] + [\sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2\ /\ ø_j^3] = 0$ // multiply $ø_j$ to both sides

$-n + \sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2/\ ø_j^2 = 0$ //simplify

$\sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2/\ ø_j^2 = n$ //add n to 0

$ø_j^2 = \sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2\ /\ n$ // multiply $ø_j^2$ and then divide by n

$\hat{o}_j = \mathrm{sqrt}(\sum_{i=1}^{n} (x[y]_i - \mu_j)\hat{}2\ /\ n)$ //take sqrt of both sides

$\partial\ \lambda/\lambda\ [\ \prod_{i=1}^{k}\ \lambda_i]$?

Problem 4

Text	Label
"A great game"	Sports
"The election was over"	Non-Sports
"Very clean match"	Sports
"A clean but forgettable game"	Sports
"It was a close election"	Non-Sports

a.

Vocabulary

class	a	great	game	very	clean	match	but	forgettable
S-1	1	1	1	0	0	0	0	0

NS-1	0	0	0	0	0	0	0	0
S-2	0	0	0	1	1	1	0	0
S-3	1	0	1	0	1	1	1	1
NS-2	1	0	0	0	0	0	0	0

Vocabulary Continued

class	the	election	was	over	it	close
S-1	0	0	0	0	0	0
NS-1	1	1	1	1	0	0
S-2	0	0	0	0	0	0
S-3	0	0	0	0	0	0
NS-2	0	1	1	0	1	1

Problem 5

Data Points = [(1,3), (2,3), (1,4), (1,5), (1,6), (2,6), (3,3), (4,1), (4,2), (5,1), (5,2)]
Distance = $(x_{xi}-c_{xj})^2 + (x_{yi} - c_{yj})^2$
Clusters: C1 = (2,4) C2 = (5,2)

Data Point	**C1 Distance**	**C2 Distance**	**Cluster**	**Data Point**	**C1 Distance**	**C2 Distance**	**Cluster**
(1,3)	2	17	C1	(3,3)	2	5	C1
(2,3)	1	10	C1	(4,1)	13	2	C2
(1,4)	1	20	C1	(4,2)	8	1	C2
(1,5)	2	25	C1	(5,1)	18	1	C2
(1,6)	5	32	C1	(5,2)	13	0	C2
(2,6)	4	25	C1				

New Clusters after 1st iteration:

C1 = x: (1+2+1+1+1+2+3) / 7 = 1.57 y: (3+3+4+5+6+6+3)/7 = 4.29
C1 = (1.57,4.29)
C2 = x: (4+4+5+5)/4 = 4.5 y: (1+2+1+2)/4 = 1.5
C2 = (4.5,1.5)

Data Point	**C1 Distance**	**C2 Distance**	**Cluster**	**Data Point**	**C1 Distance**	**C2 Distance**	**Cluster**
(1,3)	1.99	14.5	C1	(3,3)	3.70	4.5	C1
(2,3)	1.85	8.5	C1	(4,1)	16.73	0.50	C2
(1,4)	0.41	18.5	C1	(4,2)	11.15	0.50	C2
(1,5)	0.83	24.5	C1	(5,1)	22.59	0.50	C2
(1,6)	3.25	32.5	C1	(5,2)	17.00	0.50	C2
(2,6)	3.11	26.5	C1				

After the 2nd iteration the partitions between the classes does not change and therefore neither do the cluster means of the data