Hint of the solution of Assignment I. There may be some minor mistake. Plese correct it accordingly.

Thank you sec 7 for the answers.

1. Given a set of training data, probabilistic generative models learn the joint probability distribution that represents the training data & uses this underlying distribution to generate new data similar to the training data.

Applications of probabilistic generative models are :-

i. Data (image, audio, text) generation & data augmentation
ii. Anomaly detection & fraud detection
iii. Autonomous vehicles & Virtual reality gaming

2.

Probabilistic generative models	Deterministic models
i. Given a set of training data, it learns the joint probability distribution that represents the training data & uses this underlying distribution to generate new data similar to the training data.	i. Given a set of training data, it constructs a deterministic function & uses the function to generate new output for the input.
ii. It can generate different output for the same input.	ii. It can produce the same output for the same input.
iii. Ex- Naive Bayes classification	iii. Linear & logistic regression, Decision tree

3. 3 cards are drawn in succession without replacement from a deck of 52 cards. A_1 is the event that the 1^{st} card is a red ace, A_2 is the event that the 2^{nd} card is a 10 or a jack & A_3 is the event that the 3^{rd} card is greater than 3 but less than 7.

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\,P(A_2 | A_1)\,P(A_3 | A_1 \cap A_2)$$

$$= \frac{2}{52} \times \frac{8}{51} \times \frac{12}{50} = \frac{24}{16575} = 0.00145$$

4. Let A_1, A_2, A_3 & A_4 be the events of selecting a good quart of milk in succession in each draw without replacement.

So, probability of randomly selecting 4 good quarts of milk out of 20 quarts, of which 5 have spoiled

$= P(A_1 \cap A_2 \cap A_3 \cap A_4)$

$= P(A_1)\, P(A_2 | A_1)\, P(A_3 | A_1 \cap A_2)\, P(A_4 | A_1 \cap A_2 \cap A_3)$

$= \frac{15}{20} \times \frac{14}{19} \times \frac{13}{18} \times \frac{12}{17} = 0.282$

5. $f(x) = \begin{cases} \frac{x}{2}, & 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$, $y(x) = 1 - \frac{\sqrt{4-x^2}}{2}$

So, $\frac{dy}{dx} = \left(-\frac{1}{2}\right) \times \frac{1}{2\sqrt{4-x^2}} \times (-2x) = \frac{x}{2\sqrt{4-x^2}}$

$\Rightarrow \frac{dx}{dy} = \frac{2\sqrt{4-x^2}}{x}$

Now, $y = 1 - \frac{\sqrt{4-x^2}}{2}$

$\Rightarrow \frac{\sqrt{4-x^2}}{2} = 1 - y \Rightarrow \sqrt{4-x^2} = 2 - 2y \Rightarrow 4 - x^2 = (2-2y)^2$

$\Rightarrow 4 - x^2 = 4 + 4y^2 - 8y \Rightarrow x^2 = 8y - 4y^2 \Rightarrow x^2 = 4y(2-y)$

$\Rightarrow x = 2\sqrt{y(2-y)}$ when $x = 0, y = 0$ & $x = 2, y = 1$.

Now, $\frac{dx}{dy} = \frac{2\sqrt{4-x^2}}{x} = \frac{2(2-2y)}{2\sqrt{y(2-y)}} = \frac{2(1-y)}{\sqrt{y(2-y)}}$

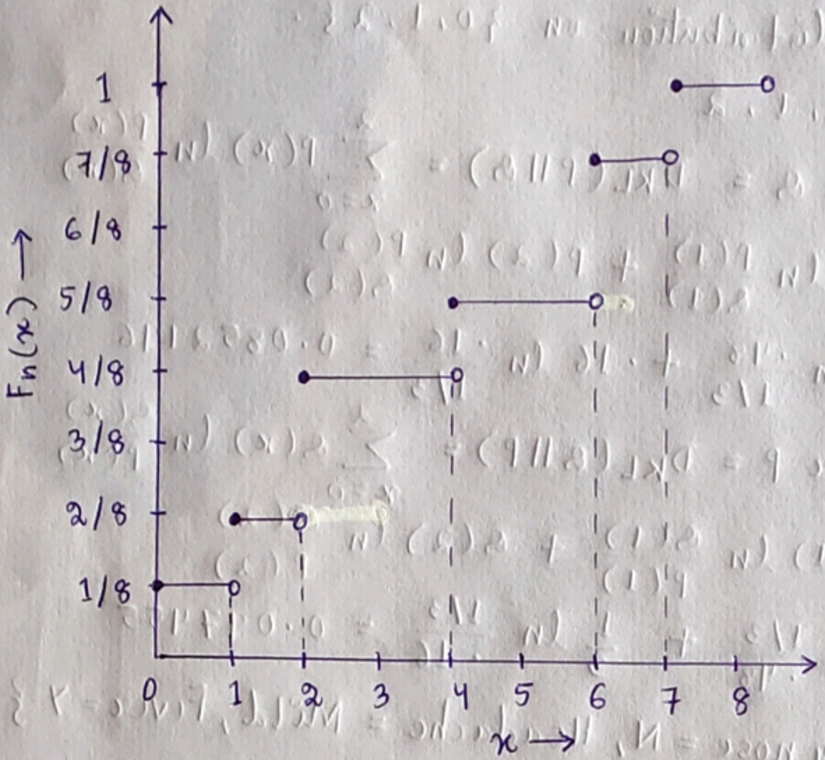
So, $g(y) = f(x(y)) \left| \frac{dx(y)}{dy} \right| = f\left(2\sqrt{y(2-y)}\right) \left| \frac{2(1-y)}{\sqrt{y(2-y)}} \right|$

$= \frac{2\sqrt{y(2-y)}}{2} \times \frac{2(1-y)}{\sqrt{y(2-y)}} = 2(1-y)$

So, $g(y) = \begin{cases} 2(1-y), & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$

7. data = [0, 1, 2, 2, 4, 6, 6, 7] , $n = 8$

Empirical distribution $F_n(x) = \begin{cases} 0, & x < 0 \\ 1/8, & 0 \le x < 1 \\ 2/8, & 1 \le x < 2 \\ 4/8, & 2 \le x < 4 \\ 5/8, & 4 \le x < 6 \\ 7/8, & 6 \le x < 7 \\ 1, & x \ge 7 \end{cases}$



Graph of $F_n(x)$

8. $f(x;\theta) = \begin{cases} \dfrac{\theta}{x^{\theta+1}}, & x > 1, \ \theta > 0 \\ 0, & \text{otherwise} \end{cases}$

$$L(\theta) = f(x_1)\, f(x_2) \cdots f(x_n) = \frac{\theta^n}{\left(\prod_{i=1}^{n} x_i\right)^{\theta+1}} = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)}$$

$$\ln L(\theta) = \ln\left(\theta^n \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)}\right) = \ln \theta^n + \ln\left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)}$$

$$= n \ln\theta - (\theta+1) \ln\left(\prod_{i=1}^{n} x_i\right) = n\ln\theta - (\theta+1) \sum_{i=1}^{n} \ln(x_i)$$

Now, $\dfrac{d \ln L(\theta)}{d\theta} = \dfrac{n}{\theta} - \sum_{i=1}^{n} \ln(x_i) = 0 \Rightarrow \dfrac{n}{\theta} = \sum_{i=1}^{n} \ln(x_i)$

$$\Rightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln(x_i)}$$

Here, $n = 6$. $\theta = \dfrac{6}{\ln(12) + \ln(11.2) + \ln(13.5) + \ln(12.3) + \ln(13.8) + \ln(11.9)} = 0.397$

9. Random variable X takes values $\{0, 1, 2\}$.

P follows binomial distribution with $n = 2$, $p = 0.4$.

$$P = {}^2C_x (0.4)^x (0.6)^{2-x}, \quad x = 0, 1, 2$$

$$P(0) = {}^2C_0 (0.4)^0 (0.6)^2 = 0.36$$
$$P(1) = {}^2C_1 (0.4)^1 (0.6)^1 = 0.48$$
$$P(2) = {}^2C_2 (0.4)^2 (0.6)^0 = 0.16$$

Q follows uniform distribution on $\{0, 1, 2\}$.

$$Q = 1/3, \quad x = 0, 1, 2$$

KL-divergence of P over Q $= D_{KL}(P \| Q) = \sum_{x=0}^{2} P(x) \ln \frac{P(x)}{Q(x)}$

$$= P(0) \ln \frac{P(0)}{Q(0)} + P(1) \ln \frac{P(1)}{Q(1)} + P(2) \ln \frac{P(2)}{Q(2)}$$

$$= .36 \ln \frac{.36}{1/3} + .48 \ln \frac{.48}{1/3} + .16 \ln \frac{.16}{1/3} = 0.0852996$$

KL-divergence of Q over P $= D_{KL}(Q \| P) = \sum_{x=0}^{2} Q(x) \ln \frac{Q(x)}{P(x)}$

$$= Q(0) \ln \frac{Q(0)}{P(0)} + Q(1) \ln \frac{Q(1)}{P(1)} + Q(2) \ln \frac{Q(2)}{P(2)}$$

$$= \frac{1}{3} \ln \frac{1/3}{.36} + \frac{1}{3} \ln \frac{1/3}{.48} + \frac{1}{3} \ln \frac{1/3}{.16} = 0.097455$$

10. $X = \{$ Chills = Y, Runny nose = N, Headache = Mild, Fever = Y $\}$

$$P(\text{Flu} = \text{Yes} \mid X) = \frac{P(X \mid \text{Flu} = \text{Yes})\, P(\text{Flu} = \text{Yes})}{P(X)}$$

$$= \frac{P(\text{Chills} = Y \mid \text{Flu} = \text{Yes})\, P(\text{Runny nose} = N \mid \text{Flu} = \text{Yes})\, P(\text{Headache} = \text{Mild} \mid \text{Flu} = \text{Yes})\, P(\text{Fever} = Y \mid \text{Flu} = \text{Yes})\, P(\text{Flu} = \text{Yes})}{P(X)}$$

$$= \frac{(3/5)(1/5)(2/5)(4/5)(5/8)}{P(X)} = \frac{0.024}{P(X)}$$

$$P(\text{Flu} = \text{No} \mid X) = \frac{P(X \mid \text{Flu} = \text{No})\, P(\text{Flu} = \text{No})}{P(X)}$$

$$= \frac{P(\text{Chills} = Y \mid \text{Flu} = \text{No})\, P(\text{Runny nose} = N \mid \text{Flu} = \text{No})\, P(\text{Headache} = \text{Mild} \mid \text{Flu} = \text{No})\, P(\text{Fever} = Y \mid \text{Flu} = \text{No})\, P(\text{Flu} = \text{No})}{P(X)}$$

$$= \frac{(1/3)(2/3)(1/3)(1/3)(3/8)}{P(X)} = \frac{0.009}{P(X)}$$

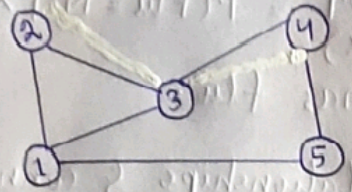
$\because P(Flu = Yes \mid X) = \dfrac{0.024}{P(X)} > \dfrac{0.009}{P(X)} = P(Flu = No \mid X)$

$\Rightarrow$ patient with symptoms X has flu.

11. In AI, graphs represent relationships & connections between entities & are well suited to model any problem where the goal is to understand a discrete collection of objects through the relationships among them.

Node features - We can assign to each $node_i$ a list of d features bundled together in a vector $\overrightarrow{features_{node_i}}$. We can then bundle all the feature vectors of all the n nodes of the graph in a matrix $Features_{Nodes}$ of size $d \times n$.

Edge features - We can assign to each $edge_{ij}$ a list of c features bundled together in a vector $\overrightarrow{features_{edge_{ij}}}$. We can then bundle all the feature vectors of all the m edges of the graph in a matrix $Features_{Edges}$ of size $c \times m$.

12. 

Different types of graph representations are adjacency matrix & adjacency list.

Adjacency matrix

	1	2	3	4	5
1	0	1	1	0	1
2	1	0	1	0	0
3	1	1	0	1	0
4	0	0	1	0	1
5	1	0	0	1	0

Adjacency list

1 → 2 → 3 → 5 → null

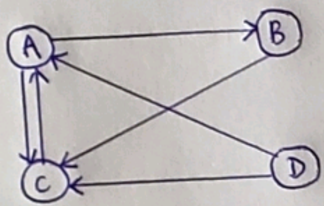
2 → 1 → 3 → null

3 → 1 → 2 → 4 → null

4 → 3 → 5 → null

5 → 1 → 4 → null

13. PageRank algorithm is an algorithm used by Google search to rank webpages based on their importance (number & quality of incoming links) in a network. It was developed by Larry Page.

14. a. G: 

Transition probability (linking) matrix –:

$$\begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

b. Given, $C(A) = 2$, $C(B) = 1$, $C(C) = 1$, $C(D) = 2$

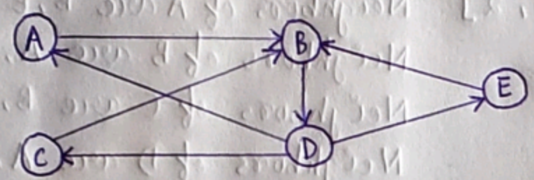
Page	0th iter	1st iter	2nd iter	3rd iter
$PR(A) = \frac{PR(C)}{C(C)} + \frac{PR(D)}{C(D)}$	0.25	$\frac{.25}{1} + \frac{.25}{2} = 0.375$	$\frac{.5}{1} + \frac{.25}{2} = 0.625$	$\frac{.4375}{1} + \frac{.25}{2} = 0.5625$
$PR(B) = \frac{PR(A)}{C(A)}$	0.25	$\frac{.25}{2} = 0.125$	$\frac{.375}{2} = 0.1875$	$\frac{.625}{2} = 0.3125$
$PR(C) = \frac{PR(A)}{C(A)} + \frac{PR(B)}{C(B)} + \frac{PR(D)}{C(D)}$	0.25	$\frac{.25}{2} + \frac{.25}{1} + \frac{.25}{2} = 0.5$	$\frac{.375}{2} + \frac{.125}{1} + \frac{.25}{2} = 0.4375$	$\frac{.625}{2} + \frac{.1875}{1} + \frac{.25}{2} = 0.625$
PR(D)	0.25	0.25	0.25	0.25

c. When the damping factor $d = 0.85$, $PR(A) = (1-d) + d \sum_{i=1}^{n} \frac{PR(T_i)}{C(T_i)}$

$$= (1 - .85) + .85 \times \sum_{i=1}^{n} \frac{PR(T_i)}{C(T_i)} = .15 + .85 \times \sum_{i=1}^{n} \frac{PR(T_i)}{C(T_i)}$$

Page	0th iter	1st iter	2nd iter	3rd iter
$PR(A) = \frac{PR(C)}{C(C)} + \frac{PR(D)}{C(D)}$	0.25	$.15 + .85 \times .375 = 0.46875$	$.15 + .85 \times .65 = 0.7025$	$.15 + .85 \times .7057 = 0.7498$
$PR(B) = \frac{PR(A)}{C(A)}$	0.25	$.15 + .85 \times .125 = 0.25625$	$.15 + .85 \times .234375 = 0.3492$	$.15 + .85 \times .35125 = 0.4485$
$PR(C) = \frac{PR(A)}{C(A)} + \frac{PR(B)}{C(B)} + \frac{PR(D)}{C(D)}$	0.25	$.15 + .85 \times .5 = 0.575$	$.15 + .85 \times .565625 = 0.6307$	$.15 + .85 \times .77545 = 0.8091$
$PR(D)$	0.25	$.15 + .85 \times 0 = 0.15$	$.15 + .85 \times 0 = 0.15$	$.15 + .85 \times 0 = 0.15$

15.



a. Transition probability (linking) matrix -:

$$\begin{bmatrix} 0 & 0 & 0 & 1/3 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \end{bmatrix} = P$$

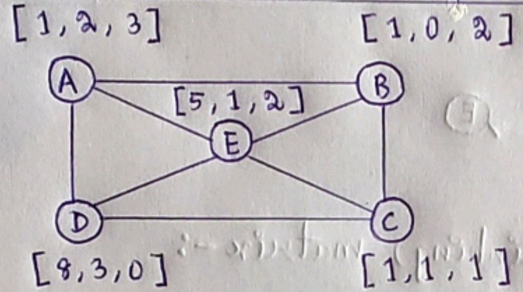
b. $$rank^{(1)} = P \cdot rank^{(0)} = \begin{bmatrix} 0 & 0 & 0 & 1/3 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \end{bmatrix} \begin{bmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 0.067 \\ 0.6 \\ 0.067 \\ 0.2 \\ 0.067 \end{bmatrix}$$

$$rank^{(2)} = P \cdot rank^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 1/3 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \end{bmatrix} \begin{bmatrix} 0.067 \\ 0.6 \\ 0.067 \\ 0.2 \\ 0.067 \end{bmatrix} = \begin{bmatrix} 0.0223 \\ 1.8 \\ 0.0223 \\ 0.2 \\ 0.0223 \end{bmatrix}$$

c. Given, $C(A) = C(B) = C(C) = C(E) = 1$, $C(D) = 3$

Page	0th iter	1st iter	2nd iter
$PR(A) = \frac{PR(D)}{C(D)}$	0.2	$\frac{0.2}{3} = 0.067$	$\frac{0.2}{3} = 0.067$
$PR(B) = \frac{PR(A)}{C(A)} + \frac{PR(C)}{C(C)} + \frac{PR(E)}{C(E)}$	0.2	$\frac{0.2}{1} + \frac{0.2}{1} + \frac{0.2}{1} = 0.6$	$\frac{.067}{1} + \frac{.067}{1} + \frac{.067}{1} = 0.201$
$PR(C) = \frac{PR(D)}{C(D)}$	0.2	$\frac{0.2}{3} = 0.067$	$\frac{0.2}{3} = 0.067$
$PR(D) = \frac{PR(B)}{C(B)}$	0.2	$\frac{0.2}{1} = 0.2$	$\frac{0.6}{1} = 0.6$
$PR(E) = \frac{PR(D)}{C(D)}$	0.2	$\frac{0.2}{3} = 0.067$	$\frac{0.2}{3} = 0.067$

16.



Neighbors of A are B, E & D.
Neighbors of B are A, E & C.
Neighbors of C are B, E & D.
Neighbors of D are A, E & C.
Neighbors of E are A, B, C & D.

a. Aggregated Message (Node) = Average of node values

$$A' = \left[\frac{1+1+5+8}{4}, \frac{2+0+1+3}{4}, \frac{3+2+2+0}{4}\right] = [3.75, 1.5, 1.75]$$

$$B' = \left[\frac{1+1+5+1}{4}, \frac{2+0+1+1}{4}, \frac{3+2+2+1}{4}\right] = [2, 1, 2]$$

$$C' = \left[\frac{1+5+8+1}{4}, \frac{0+1+3+1}{4}, \frac{2+2+0+1}{4}\right] = [3.75, 1.25, 1.25]$$

$$D' = \left[\frac{1+5+8+1}{4}, \frac{3+2+1+1}{4}, \frac{3+2+0+1}{4}\right] = [3.75, 1.75, 1.5]$$

$$E' = \left[\frac{1+1+1+5+8}{5}, \frac{2+0+1+3+1}{5}, \frac{3+2+2+0+1}{5}\right] = [3.2, 1.4, 1.6]$$

b. Aggregated Message (Node) = $\frac{\text{product of node values}}{\text{sum of node values}}$

$$A' = \left[\frac{1\times1\times5\times8}{1+1+5+8}, \frac{2\times0\times1\times3}{2+0+1+3}, \frac{3\times2\times2\times0}{3+2+2+0}\right] = [2.67, 0, 0]$$

$$B' = \left[\frac{1\times1\times5\times1}{1+1+5+1}, \frac{2\times0\times1\times1}{2+0+1+1}, \frac{3\times2\times2\times1}{3+2+2+1}\right] = [0.625, 0, 1.5]$$

$$C' = \left[\frac{1\times5\times8\times1}{1+5+8+1}, \frac{0\times1\times3\times1}{0+1+3+1}, \frac{2\times2\times0\times1}{2+2+0+1}\right] = [2.67, 0, 0]$$

$$D' = \left[\frac{1\times5\times8\times1}{1+5+8+1}, \frac{2\times1\times3\times1}{2+1+3+1}, \frac{3\times2\times0\times1}{3+2+0+1}\right] = [2.67, 0.857, 0]$$

$$E' = \left[\frac{1\times1\times1\times5\times8}{1+1+1+5+8}, \frac{2\times0\times1\times3\times1}{2+0+1+3+1}, \frac{3\times2\times2\times0\times1}{3+2+2+0+1}\right] = [2.5, 0, 0]$$

C. Aggregated Message (Node) = $\sqrt{\text{sum of node values}}$

$$A' = [\sqrt{1+1+5+8}, \sqrt{2+0+1+3}, \sqrt{3+2+2+0}]$$
$$= [3.872, 2.449, 2.645]$$

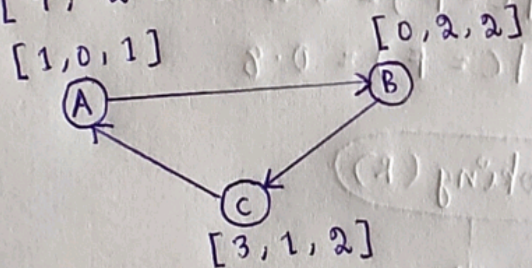
$$B' = [\sqrt{1+1+5+1}, \sqrt{2+0+1+1}, \sqrt{3+2+2+1}]$$
$$= [2.828, 2, 2.828]$$

$$C' = [\sqrt{1+5+8+1}, \sqrt{0+1+3+1}, \sqrt{2+2+0+1}]$$
$$= [3.872, 2.236, 2.236]$$

$$D' = [\sqrt{1+5+8+1}, \sqrt{2+1+3+1}, \sqrt{3+2+0+1}]$$
$$= [3.872, 2.645, 2.449]$$

$$E' = [\sqrt{1+1+1+5+8}, \sqrt{2+0+1+3+1}, \sqrt{3+2+2+0+1}]$$
$$= [4, 2.645, 2.828]$$

17.



Neighbor of A is B.
Neighbor of B is C.
Neighbor of C is A.

Aggregated Message (Node) = Cross product of node values

$$A' = A \times B = \langle 1,0,1\rangle \times \langle 0,2,2\rangle = [-2,-2,2]$$
$$B' = B \times C = \langle 0,2,2\rangle \times \langle 3,1,2\rangle = [2,6,-6]$$
$$C' = C \times A = \langle 3,1,2\rangle \times \langle 1,0,1\rangle = [1,-1,-1]$$

Note- If $A = [a,b,c]$ & $B = [d,e,f]$, then $A \times B$ is given by $[(bf-ce), (cd-af), (ae-bd)]$

18. Bayesian network is a graphical model that represents probabilistic relationships among a set of variables. It is useful for reasoning under uncertainity, as it allows to model complex domains with conditional dependencies.

Components of a Bayesian network -:

i. Nodes - Represent random variables, either discrete or continuous.

ii. Edges - Directed edges bet^n nodes represent conditional dependencies bet^n the variables.

19. Given, $P(R=T) = 0.3$, $P(R=F) = 0.7$

$P(C=T \mid R=T) = 0.4$ $\quad P(S1=T \mid C=T) = 0.8$

$P(C=F \mid R=T) = 0.6$ $\quad P(S1=F \mid C=T) = 0.2$

$P(C=T \mid R=F) = 0.05$ $\quad P(S1=T \mid C=F) = 0.1$

$P(C=F \mid R=F) = 0.95$ $\quad P(S1=F \mid C=F) = 0.9$

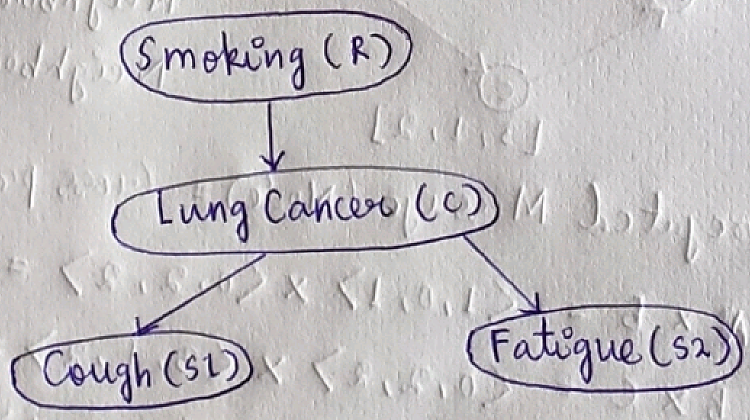
$P(S2=T \mid C=T) = 0.7$

$P(S2=F \mid C=T) = 0.3$

$P(S2=T \mid C=F) = 0.2$

$P(S2=F \mid C=F) = 0.8$

a. DAG -:



c. We need to calculate $P(C=T \mid S1=T, S2=T)$

$$= \frac{P(S1=T, S2=T \mid C=T)\ P(C=T)}{P(S1=T, S2=T)}$$

Now, $P(S1=T, S2=T | C=T) = P(S1=T|C=T)\,P(S2=T|C=T) = .8 \times .7$
$= .56$

$P(S1=T, S2=T|C=F) = P(S1=T|C=F)\,P(S2=T|C=F) = .1 \times .2 = .02$

$P(C=T) = P(C=T|R=T)\,P(R=T) + P(C=T|R=F)\,P(R=F)$

$= .4 \times .3 + .05 \times .7 = .12 + .035 = .155$

$P(C=F) = 1 - P(C=T) = 1 - .155 = .845$

$P(S1=T, S2=T) = P(S1=T, S2=T | C=T)\,P(C=T)$
$+ P(S1=T, S2=T | C=F)\,P(C=F)$

$= .56 \times .155 + .02 \times .845 = .0868 + .0169 = .1037$

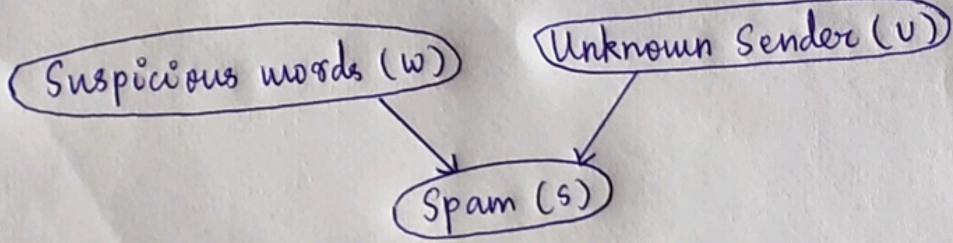
So, $P(C=T|S1=T, S2=T) = \dfrac{.56 \times .155}{.1037} = 0.837$

$\Rightarrow$ 83.7% of patients with cough & fatigue have lung cancer.

20. Given, $P(W=1) = 0.4$, $P(W=0) = 0.6$
$P(U=1) = 0.3$, $P(U=0) = 0.7$

$P(S=1|W=1, U=1) = 0.95$
$P(S=1|W=1, U=0) = 0.8$
$P(S=1|W=0, U=1) = 0.7$
$P(S=1|W=0, U=0) = 0.1$

a. DAG -:



c. We need to calculate $P(W=1|S=1) = \dfrac{P(S=1|W=1)\,P(W=1)}{P(S=1)}$

Now, $P(S=1) = P(S=1|W=1, U=1)\,P(W=1)\,P(U=1)$
$+ P(S=1|W=1, U=0)\,P(W=1)\,P(U=0)$
$+ P(S=1|W=0, U=1)\,P(W=0)\,P(U=1)$
$+ P(S=1|W=0, U=0)\,P(W=0)\,P(U=0)$

$= .95 \times .4 \times .3 + .8 \times .4 \times .7 + .7 \times .6 \times .3 + .1 \times .6 \times .7$

$= .114 + .224 + .126 + .042 = .506$

$$P(S=1 \mid W=1) = P(S=1 \mid W=1, U=1)\,P(U=1) + P(S=1 \mid W=1, U=0)\,P(U=0)$$

$$= .95 \times .3 + .8 \times .7 = .285 + .56 = .845$$

So, $P(W=1 \mid S=1) = \dfrac{.845 \times .4}{.506} = 0.668$

$\Rightarrow$ 66.8% of emails that contain suspicious words are marked as spam.